1989

Rater Accuracy in a Job Simulation Oral Examination for Entry Level Supervisors

Robert L. Holmgren
Loyola University Chicago

Rater Accuracy in a Job Simulation Oral Examination For Entry Level Supervisors

By

Robert L. Holmgren

A Dissertation Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

November

1989

ACKNOWLEDGEMENTS

I wish to acknowledge the support of my advisor, Dr. John Edwards, and the other members of my committee, Dr R. Scott Tindale, and Dr. Fred Bryant. The persistence and efforts of these men made this dissertation possible.

My real debt of gratitude is to Mr. Sidney Teske, Personnel Supervisor, for supporting me throughout the entire test development and data collection process and to the nine actors and actresses whose fame will undoubtedly blossom after performing in the videotapes used in this research: Barb, Denise, Juan, Matt, Nancy, Rick, Rachel, Sid, and Steve.

This research is an attempt to conduct a true experiment using job incumbents and their supervisors in as close to a "real world" job simulation examination setting as possible. It would not have been possible without the volunteers and staff participants of the Hennepin County, Minneapolis, Minnesota, Department of Personnel and Department of Economic Assistance. If this research assists personnel selection in only a small degree, then the work will have been worthwhile.

Finally, I wish to thank my father and mother. More than any other people, they encouraged and prayed for me throughout this long process. This dissertation is dedicated to them.

VITA

Robert Lloyd Holmgren is the son of Lloyd Raymond Holmgren and Elsie (Bergstrom) Holmgren. He was born on July 24, 1953 in Milwaukee, Wisconsin.

Bob attended 95th Street Elementary School, Wilbur Wright Junior High School, and John Marshall Senior High School in Milwaukee, Wisconsin. He received his High School diploma in 1971.

In August, 1971, Bob entered The University of Wisconsin - Stevens Point. He was graduated cum laude with a Bachelors of Arts in both Psychology and Philosophy in June of 1976. During the 1976-1977 school year, Bob attended the University of Minnesota as a special student.

In the fall of 1977, Bob entered the Applied Social Psychology graduate program at Loyola University of Chicago. As part of the program, Bob obtained an internship with the City of Chicago Department of Personnel. In 1980 Bob was promoted to the permanent, full time position of Examination Research Specialist. Bob was awarded his Masters of Arts degree in 1984.

In October 1983, after completing his course work, Bob accepted a position as a Personnel Psychologist for Hennepin County (Minneapolis), Minnesota. Bob was promoted in 1986 to Senior Personnel Representative.

In November 1987, Bob accepted a position as Associate Research Scientist for the American Institutes for Research, in Palo Alto, California.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

TABLE OF APPENDICES

THE CONCEPT OF RATER ACCURACY

Do Accurate Ratings Require Experts?

Early research on rating processes used by expert clinical psychologists centered on a traits. Ratings by expert clinicians were more than just valid, they were empathetic. A good rater had the quality of "understanding" or "accurate judgement" about the actions or feelings of another person. Rater accuracy was assessed by having research participants predict other person's responses to rating scales (e.g., personal values, clinical assessment scales, etc.). The measure of rater accuracy was the difference between the predicted and the actual score produced by the client.

Wiggins (1973, chapter 4) describes the history of clinical prediction by expert psychologists. He outlines and evaluates the various hypotheses about why clinicians were thought to be more accurate than others in predicting people's performance. Clinicians were first thought to be more accurate because of their specialized training. The evidence does not support this conclusion.

Kelly and Fiske (1954) found that advanced psychology graduate students were no more accurate than beginning graduate students on predicting the responses of psychiatric patients on personality tests questions. In Kremers (1960), graduate students viewing a ten minute speech were no more accurate than undergraduates on predicting how the speech maker would react in a variety of situations.

The second hypothesis was that clinicians are more accurate than others because of professional experience gained on the job. Goldberg (1959) found that Ph.D. clinicians were no more accurate at predicting the responses of organic and nonorganic brain damaged WWII veterans on the Bender-Gestalt than were psychology trainees or the secretaries of the clinicians given a brief training course. Wiggins (1973) cites 13 additional studies that support the results that clinical predictions are no better for professional clinical psychologists and psychiatrists than for nonclinicians given a brief training course.

If clinicians are not more accurate raters based on their experience or training, it may be possible that they are more accurate in the utilization of the available data. Wiggins (1973) reviews three studies suggesting that this possibility is to a large extent not true: Kostlan (1954), Sines (1959) and Golden (1964).

Kostlan (1954), for example, had 20 experienced clinical psychologists judge the diagnostic category of five World War II veterans. The conclusion was that any generalization about the use of clinical information must be weighted with factors such as the individual user of the information, the particular test used, and the people who are to be categorized.

These studies are important because they show that expert rater accuracy (at least for clinical psychology ratings) does not improve linearly with the addition of more test information. Accuracy varies considerably across experts. Exactly why accuracy does not predictably improve with increased rater expertise, experience or amount of information available is not clear. There does seem to be some minimal

benefit to using ratings of another's performance, but the research paradigm does not lend itself to increasing the utility of the rating process. The clinical expert research only evaluates the quality of the ratings. An experimental method for measuring improvements in rating quality was needed.

Paul Meehl in his 1954 book titled, "Clinical versus statistical prediction: A theoretical analysis and a review of the evidence," presents strong evidence that statistical predictions of psychological variables are consistently more accurate than clinical expert ratings. Since that book, much research has been focused on attempting to improve the accuracy of clinical ratings.

Accuracy Is A Poorly Understood Outcome Measure, Not A Trait.

The analysis of accuracy difference scores.

A year after Meehl's analysis, Gage and Cronbach (1955) described some important methodological and conceptual problems with the interviewer accuracy and interpersonal perception literature. Accuracy was typically seen as a trait or characteristic of a rater. A rater was either good, mediocre or bad at accurately judging other people. Accuracy was seen as a personal characteristic. The research results, on the other hand, had found that a rater's accuracy was not consistent across rating situations or sets of responses. These results indicated to Gage and Cronbach (1955) that accuracy is not a personality trait, instead it is a poorly understood outcome measure used in research studies.

Gage and Cronbach (1955) stated that rater judgements have a variety of objectives. The most broad perspective in which to consider rater accuracy is in predicting how people in general will behave on a particular rating instrument or on a particular rating dimension. This conceptualization is closest to what was being measured in the typical rater accuracy study. It masks a variety of important distinctions.

Difference score accuracy may be divided among four different pieces: (1) how most people generally behave, (2) how categories of people deviate from one another (that is, how different stereotyped categories of people differ from one another), (3) how an individual differs from the stereotyped category in which they are placed, and (4) how an individual differs in their own behavior across different situations.

Accuracy translated into mathematical terms.

Cronbach (1955) translated these conceptual distinctions among the components of accuracy into mathematical terms. For the typical selection interview experiment, a rater would make ratings on a group of applicants on a set of different items or dimensions. The accuracy score would have been the difference between the predicted ratings and the observed or "true" ratings (cf. Cronbach, 1955, p. 192).

Cronbach (1955) proceeds to partition this accuracy variance into the four different components conceptualized above. The formulae for these components are presented in Table 1. The first partition corresponds to the difference between the grand mean of the predicted scores and the grand mean of the "true" scores. This accuracy component is

referred to as elevation (EL). Elevation describes the degree to which the rater's overall mean rating predicts the overall mean "true" score of all applicants rated over all items. The elevation rating is increased (and EL accuracy is decreased) by any difference between a rater's average way of rating and the average of the true scores. EL is important only as a gross measure of a rater's ability to predict the average true score over all rated items and ratees. This component of accuracy is reminiscent of the grand mean in a general linear model (e.g., a multiple regression equation).

Table 1: Formulae for the Cronbach (1955) accuracy measures		
Variable	Symbol	Formula
Elevation	EL	$E^2 = (\overline{X}_{..} - \overline{T}_{..})^2$
Differential Elevation	DE	$DE^2 = \left[\frac{1}{n}\right] \sum_i [(\overline{X}_{i.} - \overline{X}_{..}) - (T_{i.} - \overline{T}_{..})]^2$
Differential Elevation Correlation	DEr	$DEr = r_{\overline{X}_{i.}\overline{T}_{i.}}$
Stereotype Accuracy	SA	$SA^2 = \left[\frac{1}{k}\right] \sum_j [(\overline{X}_{.j} - \overline{X}_{..}) - (\overline{T}_{.j} - \overline{T}_{..})]^2$
Stereotype Accuracy Correlation	SAr	$SAr = r_{\overline{X}_{i.}\overline{T}_{.j}}$
Differential Accuracy	DA	$DA^2 = \left[\frac{1}{kn}\right] \sum_i \sum_j [(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..}) - (T_{ij} - \overline{T}_{i.} - \overline{T}_{.j} + \overline{T}_{..})]^2$
Differential Accuracy Correlation	DAr	$DAr = r_{\overline{X}'_{ij}\overline{T}'_{ij}}$
Note:	X'_{ij}	$(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})$
	T'_{ij}	$(T_{ij} - \overline{T}_{i.} - \overline{T}_{.j} + \overline{T}_{..})$
	X_i	Raw score for ratee "i"
	T_j	True score for dimension "j"

The second factor of the typical accuracy score is called differential elevation (DE). This measure is an indicator of the elevation accuracy of a rater's ratings for each individual ratee, rather than over all ratees, as in EL. Differential elevation describes a rater's ability to predict the degree to which the ratees deviate from the true scores over all rating dimensions. Differential Elevation is similar in appearance to a main effect for ratees in a two factor (i.e., ratee by dimension) fully crossed factoral design analysis of variance (ANOVA).

The differential elevation accuracy measure can further be broken down into a correlation component. Cronbach (1955) labels the correlation component "DEr," the differential elevation correlation. This correlation represents a rater's ability to accurately rank order the ratees on their overall score. This factor is, of course, not independent of the DE accuracy measure.

The third component in the accuracy measure, stereotype accuracy (SA), looks very similar to an ANOVA main effect for items or dimensions. SA reflects the ability of the rater to predict the "norm" or average rating on each of the dimensions measured on an exam, summed over all ratees. SA measures the degree to which a rater's scores reflect the "true" dimensions being measured.

Stereotype accuracy can also be analyzed into a variance and a correlational component. Stereotype accuracy measures the rater's ability to predict the overall shape and scatter of the profile of item

responses averaged over individuals. The stereotype accuracy correlation measures the accuracy of the rater in judging the average profile shape or the rank order of item difficulties (summed over ratees).

The final component into which Cronbach (1955) partitions accuracy is referred to as differential accuracy (DA). This is the component which most people think of as rater accuracy. Statistically, it appears very similar to the interaction between the ratee and the dimension (or item) "main effects." DA is a measure of the sensitivity of the rater to individual differences between ratees in the examination. It is the essential component of accuracy. It measures the rater's ability to accurately predict each ratee's true score on each item.

The differential accuracy correlation (DAr) pairs the individual rating on each dimension for each candidate with the true score for that rating. This correlation measures the rater's ability to accurately assess the rank order of ratee's scores on each item. One DAr can be calculated for each item.

The practical utility of the accuracy measures.

Crow (1954; cited in Cline, 1964) emphasizes the importance of using both the difference score and the correlational techniques for measuring the components of rater accuracy. The difference score method defines accuracy as the rater's ability to accurately predict the exact rating situation. Using this methodology a rater will get lower accuracy scores for systematic under or over estimations of the magnitude of the true scores. In the correlational techniques, rater

accuracy is defined as the rater's ability to have ratings change as the true scores change. Thus, while the correlational accuracy measures will not penalize a rater for errors in absolute magnitude of ratings, they will cause lower accuracy scores for a rater whose predictions do not vary concomitantly with the true scores.

Becker and Cardy (1986) state, and Cronbach (1955) implies that the correlation subcomponents of the accuracy components, are cleaner measures of the common meaning attached to the pieces of the accuracy construct. Practically, however, the correlational accuracy measures provide information similar in kind to criterion validity correlations.

The measures that will be of greater value in accuracy research are the mean / variance accuracy scores. Mean / variance accuracy is measured around a specific score and provides information about how close a rating is to its true score. These measures provide pinpoint accuracy measures, information that is lacking in correlations. Correlation accuracy is related to rank ordering, not to accuracy around a specific score. Differential accuracy is the measure used most often in research. Differential accuracy eliminates the effects of the overall score, the rating scale stereotypes and the ratee differences, and is left with the unique effect of the ratee on the particular rating dimension.

The Move To A Simulation Methodology

The Cronbach (1955) article spurred a great deal of interpersonal perception research that focused on measures that he suggested (cf. Cline, 1964, for a review). One line of research attempted to answer

the question of who the accurate raters were, if they are not clinical psychologists. In other words, what type of personality profile did accurate raters have? The conclusion is surprising. An accuracy trait could be measured, but it included "traits" such as an accurate understanding of the rating dimensions. The traits were defined in terms of cognitive process outcomes.

Taft (1955), in an early review of the trait literature, noted a number of methodological and conceptual difficulties in the rater accuracy studies reviewed, but was still able to note some general factors that contributed to accurate judgement. Generally, as children got older, they became more accurate judges. Adults of higher intelligence, esthetic interests in art and drama, or who had higher degrees of self insight, emotional adjustment, and social skill tended to be more accurate raters of other people.

Cline and Richards (1960, 1961, 1962, 1963; Richards and Cline, 1963; Richards, 1963) conducted a series of experiments investigating the factors underlying rater accuracy. The main difference between these and previous studies, in addition to the use of the Cronbach (1955) accuracy measures, was the use of motion pictures as stimuli on which to base ratings.

The conclusion drawn from these studies was that a generalized ability to rate others could be measured. That general trait was comprised primarily of stereotype accuracy (SA), but there was also a significant amount of differential accuracy (DA). Cline (1964) states that a rater may be an accurate judge of others provided that he or she

has either an accurate stereotype of the ratees and/or because the rater is able to predict differences between ratees on specific dimension.

Gage and Cronbach (1955) indicate that a rater with high stereotype accuracy is a rater who can predict the "pooled responses of a given category of persons" whereas a rater with high degrees of differential accuracy is a person who is able to "differentiate among individuals within a rating category" (page 417).

Borman, Hough, and Dunnette (1978) present a study that utilizes a modified version of Cronbach's (1955) differential accuracy correlation (DAr) as a measure of rater accuracy to assess the underlying characteristics of accurate raters. The study is an attempt to reassess the variables that Taft (1955) found to be important, generalizable characteristics of raters who have the ability to be accurate. The study utilizes a method for generating true scores and a method for creating videotaped stimuli that has become the standard in rater accuracy research. For this reason the methodology will be outlined in some detail.

Two different jobs were selected to be the basis of the study: a college recruiting interviewer and a manager. For each of these jobs behaviorally anchored rating scales (BARS) were developed with behavioral anchors at three levels of performance (effective, adequate, and ineffective) for each dimension. The interrater agreement ranged from .86 to 1.00.

The "true scores" were generated by a group of subject matter experts. The judges (all familiar with correlations) estimated the correlations between each of the pairs of rating dimensions. The reliability estimates on these judgements were .81 and .82 for the manager and recruiter jobs respectively. The true scores were created by averaging the correlation rating scale estimates.

The performances of the job candidates (recruiter and manager) were developed based on the true scores on the rating dimensions. Sixteen separate scripts were written, eight for each job. The scripts were rewritten until they adequately displayed the performance dimensions to the satisfaction of the three independent raters.

The written scripts were performed and videotaped using nonprofessional actors. The tapes were reviewed for realism and, if necessary, retaped. The actors were allowed to deviate from the scripts in an effort to make them realistic.

The final true scores were developed by having a panel of experts review the videotapes and rate the behaviors on the rating scales. The experts had access to the written scripts, the rating scales, and the videotapes. The reliabilities for the fourteen raters ranged from .95 to .98 for the recruiter job and ranged from .91 to .98 for the manager job. The true scores ratings showed considerable convergent and discriminant validity and relatively little halo error.

Finally, the preset and the expert assigned true scores were correlated. This is a measure of the degree to which the original true scores and the true scores set by evaluating the videotapes were

actually measuring the same thing. The mean correlations for all raters ranged from .42 to .94, with a median of .91. Borman et al. (1978) note that professional and graduate student raters were not noticeably different from each other on expertise of rating.

In the major part of the study, a group of 256 students completed a set of personality scales. The 146 students with the largest overall difference from the group's mean were selected to rate both the recruiter and the manager tapes.

Borman et al. (1978) conclude by noting that accuracy, as an ability, was present and somewhat reliable across rating situations, but that it was higher within a particular job. Accurate raters were found to: (1) be free from self doubt, (2) tend to not worry or become stressed, (3) be intelligent, (4) have high grades, (5) have investigative interests, and (6) tend to be detail oriented in their approach to tasks. Accuracy was sufficiently generalizable over situations for Borman et al. to conclude that further study was warranted.

Two additional studies were completed using the same data. Both were Monte Carlo type studies. In the first study, Borman et al. (1978) found that the differential accuracy correlation improved when raters pooled their ratings. The largest increase in DAr came in adding a second rater. The next largest increase came when adding the third rater. Although the accuracy increased as more raters were added, little increase was found after the fourth rater.

The second study assessed the degree to which interrater reliability predicted rater accuracy. The correlations ranged from .27 to .53

for the recruiter job and from .53 to .66 for the manager job. Thus, to a certain extent raters high in interrater reliability had a stronger tendency to be accurate.

Finally, Borman et al. (1978) strongly recommended the videotape approach to researching the factors that affect rater accuracy. Although the development process is time consuming and costly, the methodology provides a sound way of measuring factors that can be used to improve the accuracy and decrease the errors in rating the behaviors of others.

The study described in this paper uses a videotaped job simulation as the basis for measuring whether different rating dimensions influence rater accuracy. Using videotapes of job simulation is a bit different from the typical style of videotape stimuli in performance appraisal based research on rater accuracy, but the difference leads to an improvement in experimental design.

Job simulation examinations are like performance appraisal ratings made under ideal conditions. They are attempts to place candidates in situations that are as much like the key portions of the to-be-gained job as are possible in brief (up to an hour) controlled situations. Raters observe the candidates in the simulated job situations and evaluate their performance. The better performing candidates are selected to do the actual job. There are four main reasons why a job simulation format of rater cognitive process research is an improvement over performance appraisal research.

Unlike most performance appraisal situations, the raters in simulation exam settings watch the candidates perform the key portions of the job. The raters in performance appraisal settings may need to rely on rumors or work products as the basis of their judgements. The key portions of the job are defined in advance on the simulation exams.

A second difference between performance ratings in simulation exams and performance appraisals lies in the immediate rating of performance in simulations. Performance is rated either as soon as the simulation is completed, based on notes taken during the action, or is done after several simulations are completed, each one of which is designed to simulate another important aspect of the job. Performance appraisals, on the other hand, can be completed a year after the actual job performance situation was noted.

A third difference between a simulation and a performance appraisal rating is the environment in which the ratings are done. Simulations are done in a selection context. Candidates are thinking carefully about what they will do and are emotionally charged to do their best. Some candidates may be frightened and anxious in the simulation. This anxiety is certainly a problem in the use of simulations. Most attempt to overcome the problem by making the simulation as realistic as possible. This allows the nervous candidate to acclimate to the situation. Ratings are made on the total performance, and not just the nervous floundering in the first few minutes of the simulation.

Finally, situation for the rater is much more controlled in a simulation examination. Ratings will be less subject to manipulation from situational pressures since the ratings are done quickly and without much consultation.

Except for these differences, the performance appraisal situation and the job simulation test are very similar, the simulation environment is simply much cleaner. The raters are both typically supervisors using rating scales for the first time. The processes used in completing the ratings are similar except that the simulation ratings are made under more idealized conditions. In a research situation, one in which the "candidate" ratees are actually carefully developed and videotaped in advance, the problem of nervousness can be controlled. The simulation setting is then an ideal situation in which to investigate the processes of rater accuracy.

More important than the use of a videotaped job simulation exam, however, is the focus of the present research on rater cognitive processes. According to Landy and Farr (1980) research on rating scales and performance appraisal variables has proved to be a collection of unconnected findings. The practical value of the studies has been limited because they were not connected by any underlying model or theory. The research has not increased the overall quality of performance ratings.

Landy and Farr indicate that a more fruitful area of research is the cognitive processes of the rater. Past research has shown areas where cognitive processes have been important in improving the quality of ratings. Specifically, raters who have a clear understanding of the

rating scales are higher on discriminant validity than others; convergent validity is about the same for all raters given a brief rater training session. It is to the cognitive model of rater processes that we now turn.

FELDMAN'S COGNITIVE PROCESS MODEL OF A PERFORMANCE RATER

Ilgen, Barnes-Farrell, and McKillin (1987), in their summary of the rater accuracy literature since the Landy and Farr (1980) article, present an outline of a general cognitive model of rater processes. This rater process model contains three steps for performance appraisal ratings. First, the rater must acquire information about the ratee. This is typically done through observation of ratee behaviors. Second, the rater must organize and retain that information in memory. Third, the rater must retrieve the information stored in memory, integrate it into the conceptual framework required by the rating scales, and then assign the ratings. Raters can vary on any of the three steps: observation, storage, and recall / evaluation. These general steps are described in more detail in the model by Feldman (1980, Ilgen and Feldman, 1983).

Feldman begins by emphasizing that any model of rater processes must be understood in the context of the rating environment. Performance appraisal is one of many supervisory duties. Information that is incorporated into supervisory ratings is often fragmentary and is rarely gained from direct personal contact. The limited personal contact is usually restricted to staff meetings or crisis situations. The exact nature of a subordinate's job is rarely understood completely.

In this environment a supervisor must make performance ratings. And, as outlined by Ilgen et al. (1987) above, several internal, cognitive processes must occur before any accurate rating can be made. The Feldman model will be discussed around these four component processes:

(1) Attending to and recognizing relevant information, (2) organizing and storing information, (3) seeking and recalling information, and (4) making performance rating judgements. Within each of these four sections the relevant research will be presented. Following the discussion of the components of the model, the various sources of accuracy lowering bias will be reviewed. The section will conclude by stating the hypotheses explored in the current research.

The Feldman model is centered around the concept of a cognitive "category." A category is defined as a "fuzzy" conceptual set of correlated factors that collectively provide a single framework within which to integrate the observed performance of a ratee. Each category is defined by a "prototype." A category prototype is an "abstract analog or image summarizing the family resemblances [but not a common set of attributes] among the category members" (Feldman, 1981, p.130, explanation added). If the prototype is a verbal or propositional representation of a situation that describes a complex pattern of behaviors in a manner similar to a movie script, the prototype is referred to as a schema (cf. Markus, 1977).

Categories used during observation are single concepts that are used to classify a variety of information. Their hypothesized purpose is to ease a person's burden of remembering large amounts of factual detail. Cognitive categories provide conceptual nets into which observed information about similar situations or behaviors or individuals can be placed. To classify multiple observations of a ratee, however, either the same category can be expanded or multiple categories can be used. Presumably, any one observation will be placed into a

single category with a single prototype defining the family resemblances of the category. Raters would have a tendency to start with fewer rather than more categories to store information about a person.

The importance of the Feldman approach is that it attempts to describe both why rater errors occur and the conditions under which ratings should be more accurate. The approach is an integration of the information processing model (cf., Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977) and the attribution theory of social psychology (cf, Kelley, 1971a, 1971b).

Attending To and Recognizing Relevant Information

The model makes an important distinction between evaluations and judgements. Evaluations of the observed behavior are made immediately upon observation and are stored separately from the factual information. Judgements and decisions, however, are based on the information stored in the category. Evaluations can be recalled without the rater remembering any of the stored factual information. The reverse is not true. Recalling information stored in the category will also cause the evaluative component to be recalled as well.

Murphy, Garcia, Kerkar, Martin, and Balzer (1982) provide research evidence that observation and evaluation rating accuracy are separable. Forty-four undergraduates viewed each of four videotaped lectures and rated the lecturers on frequency of occurrence of a set of critical behaviors and on eight different dimensions of teacher performance. These are the tapes used in other studies by the same lead author reviewed below.

Murphy, Garcia et al. (1982) present information on the reliability of the "true scores." Discriminant validities were .57 for the frequency ratings and .70 for the performance evaluation ratings. Convergent validities for the frequency ratings and the performance evaluation ratings were .21 and .47. None of these validities are particularly high. This fact should be kept in mind in the interpretation of their accuracy findings. It is questionable whether or not the "true" scores are in fact true scores.

Murphy, Garcia et al. (1982) present correlations between the four accuracy component measures for the observation frequency estimates and the performance evaluations. A correlation of .70 between the two EL measures shows that even with the questionable "true" scores, the raters generally agreed on the overall accuracy of observation and evaluation ratings. The correlation between the two differential elevation measures is .38 ($p < .01$). Although significant, the raters were not as consistently accurate for the individual candidate scores and frequency measures. This may reflect the limitation in the convergent validity of the ratings. The stereotype accuracy correlation of .10 does not reach statistical significance. This may be an indication of the lack of discriminant validity of the "true" score measures. It could also indicate a lack of understanding of the dimensions of performance by the raters. In either case, it shows that the raters did not agree on the ratings assigned to the dimensions as a whole. The differential accuracy correlation between the frequency and the performance evaluation measures is .43 ($p < .01$). Despite the problems

above, the raters were still able to consistently rate individual teachers on the frequency of critical behaviors displayed in the videotapes and to evaluate the performance of these same behaviors.

The major conclusion is that the accuracy subcomponent measures are sensitive to differences between observation and evaluation accuracy. Observational accuracy and evaluation accuracy have some accuracy component measures in common, but they also have some differences. Observation and evaluation are different tasks with different degrees of accuracy of rating. Understanding of the rating dimensions was low (the stereotype accuracy correlation), yet the statistically significant differential accuracy correlation indicated that raters were able to make, on the average, common accurate judgements about the quality of individual teachers on individual rating dimensions and on the behaviors performed.

Organizing and Storing Information: The Categorization Process

The Feldman model continues by discussing the processes under which observed information is organized and stored in memory. When a rater is observing a ratee, the Feldman model proposes that a cognitive category prototype is brought into the rater's cognitive work area and is compared with the features of the observation. This observation, organization,, and storage process may be automatic (and done out of the person's conscious awareness) or controlled (and done completely within the rater's conscious awareness).

Most observations are automatically mapped into memory via category prototypes. Observations are not "stored" in long term memory with perfect detail, rather observations are quickly and unconsciously compared in working memory with a temporarily salient category prototype. If the overlap between the components of the observation and the category prototype is sufficient, the observation is automatically "stored" in memory as a member of the category. The true information is blended into the information that describes the prototype used to store the original observation.

When a rater observes someone's performance, the cognitive process utilized will depend on whether or not the rater has a stereotyped understanding of the dimensions on which the ratings will be made. If a stereotype is available, then the rater will automatically invoke it, seeking out information to verify that expectation. The example used by Feldman (1981) is that of a salesman. If one of a supervisor's subordinates is a salesman, a typical stereotype for this subordinate would be one of talkativeness. This is a preexisting category into which the person's actions would be placed. These categories are built by experience and by the processes described in attribution theory (cf Jones and Davis, 1965).

If the category prototype and the observation do not overlap sufficiently, then conscious attributional processes are invoked to establish another category into which the observation can be stored. Controlled processes occur only under conditions of "variable" rather than "constant" mapping, by using attributional processes when the automatic categories do not work. By variable mapping, Feldman means

the cognitive process where a stimulus cannot be readily translated (i.e., mapped) onto the perceiving mechanism, and as a result the receiving mechanism must be altered.

Controlled observation categorization processes are explained by attribution theory. Briefly, attribution theory indicates that people will think like a "naive scientist" when they interpret the actions of other people. When a controlled attribution process is invoked, the first step is to form a causal attribution. The decision is whether the ratee or the situation caused the observed event. The observer will form a trait, judgement, or dispositional attribution based on the observer's own implicit personality theory and stereotypes (cf. Hastorf, Schneider, and Polefka, 1970; Kelly, 1955). These attributions will be tempered by the situation in which the judgements occur.

Attribution theory proposes that these decisions will be driven by two main principles. First, actions have their causes attributed to be actions that co-vary with their occurrence. More specifically, causes for actions will be from one of three possible categories of factors: people, situations, or things (cf. Kelley, 1971a). Second, any other causal explanations will be discounted, and not used, unless they become more plausible than the original explanation (cf. Kruglanski, 1970). The result of both the controlled and automatic processes is assigning the ratee to a category based on a prototype-matching process.

Feldman (1981) presents six different conditions under which incorrect causal attributions can be predicted (cf. Ross, 1977). The first attribution error is the tendency to overestimate the importance

of within-person causes. Feldman (1981) indicates that people have a tendency to attribute the cause to the person making the action more often than would be true.

Second, the person making the action is biased toward attributing situational causes, while the actor is biased toward making internal or person causes (cf., Jones and Nisbitt, 1971).

Covariance is the third bias mentioned by Feldman (1981). Raters have a tendency to attribute causes to anything that covaries with an action's occurrence.

Fourth, if an action is seen to have some emotional (i.e., "hedonic") relevance to the actor, then the action will typically be attributed to the actor, rather than the situation.

Fifth, raters tend not to utilize the causal information available in base rate information (cf., Tversky and Kahneman, 1971). They generally fail to utilize the information available on the average frequency of occurrence of the behaviors they observe. This unused base rate information would suggest that the observed behavior probably occurred by chance. The rater would not need a causal explanation. Overlooking this information, the rater instead attributes the cause for the behavior to the first plausible explanation. The rater tends to attribute the cause of the observed action to behavioral reinforcers. Positive reinforcers are attributed as causes more frequently than negative reinforcers. A person will be seen as a cause for a rewarded action more often than as a cause for an action that leads to the avoidance of some loss or punishment.

Finally, attributions tend to follow a person's affective relations. A liked person doing a good act will receive an internal attribution. A disliked person doing a good act will receive an external attribution.

Research on the existence of the categorization process

Several research studies present evidence on the existence of the categorization process in performance appraisal settings. The evidence is not completely positive, but does tend to support the notion that raters will categorize information gathered from observing ratees.

Observation versus evaluation accuracy. In Murphy, Martin, and Garcia (1982), a group of undergraduates rated four randomly chosen videotapes of college lecturers. Ratings were made on a set of eight graphic rating scales and the two behavioral observation (BOS) scales. Two of the tapes were rated immediately after viewing and the other two were rated the following day.

The correlations between the two BOS scales and the eight graphic rating scales increased from the immediate to the delayed rating condition. The range of correlation increases was from as little as .01 to .25. The authors conclude that raters do not simply recall behaviors when they rate on a BOS scale after a time delay, rather they forget the exact detail of the behavior and recall a category or stereotyped response for that applicant. This general impression is used to make predictions on the frequency of actual behaviors. As more and more memory is relied on in a rating task the observation task switches from

pure observation to one of inference.

Memory effects on categorization accuracy. The Nathan and Lord (1983) study contradicts the findings of the Murphy, Martin, and Garcia (1982). Instead, these authors find no effect of time delay in rating accuracy. Nathan and Lord propose that the Borman model better fits the data than the Feldman model. Cronbach accuracy measures were not used.

As discussed by Nathan and Lord (1983), the Borman model suggests a three step cognitive process for ratings of performance. The process is quite direct. First, a rater must observe the relevant to-be-rated behaviors. Second, these behaviors are evaluated for effectiveness. Finally, these evaluations are weighted to arrive at the numerical rating on the dimension. There is no mention of a memory process.

To contrast the two models, Nathan and Lord (1983) developed two videotapes, each about 25 minutes in length, of the same college lecturer. Each of the tapes contained a set of carefully scripted behavioral incidents that were examples of either good or bad lecturing technique. One tape contained mostly examples of good techniques plus some poor examples. The "unfavorable" videotape contained mostly poor incidents with a few good examples. The raters assigned ratings to the lecturer either immediately after viewing the videotape or after a two day delay.

Nathan and Lord (1983) indicate that different results would be expected for the two different theories. The Borman (1978) model would

predict that ratings would be accurate provided that the rater had sufficient opportunity to observe the relevant behaviors, the biasing effects of short and long term memory are minimized, and the behaviors seen by the raters are interpreted as being examples of the rating dimensions.

Nathan and Lord say the Feldman (1981) model would predict that raters would use more controlled processes the longer the length of time between the performance and the rating. This would require the rater to depend on attribution style stereotypes generated during the videotape watching on which to base the performance ratings. These ratings would be closer to the stereotyped performances rather than the actual performances. The amount of halo error in the ratings would rise in the delayed rating over that in the immediate condition.

The specific predictions made by Nathan and Lord (1983) are dependent on the number versus the ratio of good and poor behavioral examples for each of the rated dimensions. The Borman model would predict that the numerical ratings for each dimension would depend on the ratio of good to poor incidents for each dimension separately. The Feldman (1981) model would indicate that as more controlled cognitive processes are involved (increasing with rating delay, assuming that they were used to start with), the greater would a ratee's ratings depend on the overall proportion (i.e., the number) of good to poor performance across the dimensions. The ratings in the delayed condition would be controlled by a general impression (e.g., halo) of the lecturer. The ratings for each dimension would not depend on the actual ratio of good to poor behavioral incidents in each dimension.

There was no effect for the two day delay of the rating. Mean score of the various dimensions did not have an overall multivariate significant difference from immediate to delayed rating conditions, nor was there a significant interaction between time of rating and favorability. There was a significant multivariate effect for favorability of the tape.

Based on their analyses, Nathan and Lord state that the Borman (1978) model can better explain the data. The also find support for the Feldman (1981) model. The first conclusion presented is that even under these optimal rating conditions the raters do not effectively utilize the five different dimensions of performance. Second, they indicate that the temporal delay is ineffective at altering the ratings. The dimension specific differences, however, were not measured. Finally, the temporal processes did have an effect on halo error. The ratees did show different types of recognition errors in the incidents that they recalled.

The authors conclude their presentation with an about turn statement, saying that they believe that the Feldman information processing model "will ultimately prove to be a more manageable and more profitable approach to improving ratings" (p 113).

The data in this study, however, were not analyzed using the Cronbach (1955) accuracy measures. Counts of good and poor behaviors were used to assess accuracy. This is clearly an overall accuracy measure. It does not control for the effects of the item (differential elevation) and the raters (stereotype accuracy) as does differential accu-

racy. This difference in outcome measures make the Nathan and Lord results a difficult to place into the body of research evidence related to the Feldman (1981) model.

The importance of this study is that it shows that raters are not perfectly accurate in their ratings even under these quite optimal conditions of rating only two ratees. The delay of the rating process, which should have invoked controlled categorization and attribution theory style stereotyped rating reactions did not decrease the accuracy of the ratings. These findings contradict the results of Murphy et al. (1982).

Easy versus Hard to Classify Observed Behavior. Favaro and Ilgen (1983, cited in Ilgen et al., 1987) present a study that supports the hypothesis that most observations are done automatically, without the rater's attention being actively placed on the ratee's behaviors. The study suggests that under conditions where it is easy for the observer to classify the ratee into a specific category, the rater will pay little attention. As a result, the ratings made on these "automatic" observations are less accurate than when the observers do not have a clear cut, easy to use category into which the ratee's behaviors can be classified.

Raters were asked to act as a nursing supervisor and observe and rate the performance of a simulated subordinate nurse. Some of the raters were told that the nurse was a "social activist" and given some examples of the nurse's behavioral style. Other raters were provided with several traits describing the nurse, but the traits did not fit a

specific and easily identifiable category. Each of the raters was then allowed to perform various supervisory activities including observing the subordinate nurse conducting day to day activities (on videotapes) and rating the performance of the nurse.

Raters told that the nurse fit the social activist trait label spent significantly less time observing the nurse than did raters who observed the same videotapes without such an expectation. Favaro and Ilgen (1983) concluded that raters would spend less time observing someone when the rater's task was limited to confirmation of a category. This finding supports the Feldman (1981) model of rater cognitive processes since the observation of actions which confirm a cognitive category of the rater would lead the rater to stop seeking additional information.

Rater accuracy was also linked to the ease of categorization, but not in a positive way. Raters were found to be more accurate the longer the amount of time they spent observing the ratee. Since the easily categorized ratees resulted in raters observing only until the category would be confirmed, the ratees were observed less frequently and thus were given less accurate ratings. Rater accuracy was lower for the social activist ratee on Cronbach's (1955) measures of elevation and stereotype accuracy as well as overall (i.e., difference score) accuracy. This study supports Feldman's suggestion that raters with easily accessible stereotype categories will use them to store information about the ratee rather than formulate and use more accurate categories.

Immediate versus delayed ratings. Murphy and Balzer (1986) presented more data in support of the Feldman model's style of categorization of observations by studying the effects on accuracy of immediate versus memory based ratings of videotaped lectures. Undergraduate research participants viewed a random sample of four of eight possible videotapes, two on each of two topics. Respondents rated two tapes (one on each topic) immediately after viewing. The other two tapes, again, one on each of the two topics, were rated one day later.

Stereotype accuracy and differential accuracy were significantly better in the delayed rating condition than in the immediate rating condition on both an objective (behavioral) and a subjective (performance appraisal) measure. Respondents showed less accurate differential elevation accuracy for the behavior scales and more accurate differential elevation for the performance evaluation scales. Halo error, measured as the average intercorrelation between the rating dimensions for a rater, rose significantly from the immediate to the delayed condition. Thus, raters not only improved their accuracy in the delayed rating condition, but also increase the amount of halo error displayed.

Murphy and Balzer (1986) conclude by saying that raters may tend to remember general impressions of the ratees, rather than the factual detail. It is these general impressions that are recalled when ratings must be made in delayed rating conditions. These general impressions may contain more valid information than the immediately recalled information.

Summary of Automatic and Controlled Process Research. These four studies all address whether raters use categories to store information about observations and do not store the exact details themselves. Favaro and Ilgen, and both of the Murphy studies support the hypothesis. The Favaro and Ilgen study does so without requiring delayed ratings. The Nathan and Lord study does not support the hypothesis, and instead finds no effect for delay of ratings on rater accuracy. This latter study, however, can be criticized on its accuracy measures. The Nathan and Lord study concluded by supporting the Feldman model. The contradictory experimental support should not, however, be taken lightly. The existence of memory categories (rather than actual detail) in memory is crucial to the Feldman model. No explanation is presented here to try to deal with these contradictory findings.

Research on category selection

If the categorization process exists, the next question the model must answer is how a specific category is chosen. Categorization would be easiest when every aspect of the observed behavior fits in exactly with a preexisting category in the observer's working memory. Such would be the case with Feldman's example of the talkative salesman. Typical observations, however, are made in much more complex situations. Both the situation and the individual could influence the category chosen by the rater.

Citing interpersonal perception research as support, Feldman suggests that situational factors make certain categories more likely to be used. Anything that makes one aspect of a situation more salient

will tend to cause that unique factor to be the basis for the categorization. In addition, the context in which the judgement is to be made will also have a controlling influence on the category chosen. The context would tend to make some factors more salient to the rater.

An example of a salience manipulation is being the only female executive in a room full of men in gray flannel suits. The woman will tend to be categorized according to her unique feminine status rather than by other available categories. Any novel component in a situation will tend to control the categorization process (cf. Taylor and Fiske, 1978). Feldman postulates that the temporarily salient aspect of the situation controls the categorization process by directing attention to aspects of the situation that are examples of the observational categories and prototypes. Attention is directed away from examples that do not fit the salient category.

The categorization process has been the focus of much of the research on rater accuracy. One study, by Barnes-Farrell and Courture (1984), focuses on the salience of the rating task, observational thoroughness, and rater accuracy.

Observation and the timing of rating scale presentation. In a study briefly discussed in the Ilgen et al. (1987) article, Barnes-Farrell and Courture (1984) explored the importance of making rating dimensions salient to the raters. Raters were presented the rating scales and the rating dimensions at different times in relationship to behavioral observation. Raters were also asked to recall and evaluate their observations either immediately or after a time delay.

When raters were required to make delayed evaluations of previously observed behaviors, the evaluations were more accurate when the rating task was made salient prior to observing the performance. The salience of the rating task had no effect when the ratings were made immediately after the observation. This study supports the hypothesis (in a performance appraisal context) that the salience of a particular aspect of a rating situation, in this case the rating dimensions themselves, will cause that aspect to be attended to more closely.

Judgement context in category selection. Another important influence on category selection is the judgement context (Tversky, 1977). Certain contexts, such as ratings made for different purposes, make some categories more salient for some persons than for others. The context and a person's difference from others in the context makes the difference more salient.

Four studies will be discussed that investigate variables that focus an observer's attention on different aspects of a situation. The ease or difficulty of classifying a ratee, the purpose of the rating, the rating situation, and the rater's motivations all have significant effects on the accuracy of the ratings. Each is proposed to achieve this effect by changing the category selected for use in the situation.

The technical report by Favaro and Ilgen (1983) discussed above is relevant here. Raters who rated a person who was a "social activist" recalled these traits during rating. Other raters given trait expectations not fitting into a specific category did not have their category selection manipulated. In turn, these latter observers needed to pay

closer attention to the situation. Raters spent more time observing and were more accurate when they did not have a predefined, easy to use category for classification of a ratee.

The classic study indicating that raters are differentially affected by the purpose of or the context in which ratings are made is that of Zedeck and Cascio (1982). The ratees in the study were checkers and baggers of groceries in a supermarket. Raters were given written descriptions of how the ratees were performing. Accuracy was measured by predicting objective measures of performance, rather than the more typical subjective ratings of performance. Raters who were told to make ratings for merit pay increases were less accurate than raters who made ratings to provide developmental feedback or to remember as much as possible about the performance.

Williams, DeNisi, Blencoe, and Cafferty (1985) conducted a study similar in intent to the Zedeck and Casio study. Raters were told that the appraisal being conducted was either for salary increases, for promotional recommendations or for remedial training purposes. In each of the three conditions raters were allowed to seek information about how the ratee was performing his or her job.

Although all raters tended to seek some similar types of information, raters also sought different types of information depending on the purpose of the rating. Williams et al. (1985) conclude that raters are sensitive to the demands of the rating situation. Rater may be differentially motivated to rate accurately depending on the purpose of the rating, but they will also seek different information depending on the purpose the rating is to achieve.

Ilgen et al. (1987) review an article by Bernardin, Cardy, and Abbot (1982) investigating the effects on performance accuracy of both the rater's knowledge of the ratee's job, and exposure to the rating dimensions and scales prior to observing the ratees. In addition, Bernardin et al. (1982) had respondents describe what they believed to be the "important dimensions" of the job of an interviewer. These written descriptions were used to rank order the respondents on the degree to which their personal understanding agreed with expert's judgements about the job. The videotapes used were the Borman (1978) tapes. None of the manipulated variables (job knowledge information, prior exposure to the rating scales, and compatibility of rater and rating scale dimension categories) had any effect on the accuracy of the ratings.

In a follow up study, Bernardin et al. (1982) manipulated the motivation to rate accurately by providing half of the undergraduate raters with a monetary bonus for rating accurately. The original variables were also added to the motivation manipulation. This time, differential accuracy was higher for more highly motivated raters than for the raters not given the incentive to rate accurately. In addition, highly motivated raters whose prior beliefs about the job were similar to the actual categories used in the rating scales also rated more accurately than those motivated raters whose prior beliefs were different.

These studies points out the importance of situational and contextual variables. They may mask, eliminate, or increase the effects of other, more subtle variables. Before subtle cognitive variables can be assessed, the situational factors must be controlled. If the rater is

not motivated to rate accurately, either due to lack of reward or due to social constraints, then, in all likelihood, accurate ratings will not be provided.

Individual Differences in Category Selection. In addition to situational influences on the choice of a category into which an observation will be stored, raters will have individual and cultural differences in their choices of categories. Feldman notes that people differ in the nature and number of categories they possess (i.e., their stereotypes and implicit personality theories). Categories are formed by the rater observing covariations in the environment. These covariations can be unique to the individual. They can also be common to the culture. As an example of the latter, Feldman notes that one culture's friendly banter may be another culture's disrespectful familiarity.

One research example of the effects of culture common variance in category selection is presented by Schmitt and Lappin (1982). This research topic was, prior to the Landy and Farr (1980) article, one of the major areas of performance appraisal research. It concerned the effects of the race of the rater and the race and sex of the ratee on rater errors and mean scores assigned to the ratees. These articles were largely focused on the differential validity debate raging in the journals (cf. Hunter, Schmidt, and Hunter, 1979). Schmitt and Lappin (1980) continued in this tradition and added a correlational measure of accuracy as well. Their measure of accuracy was the correlation between ratee true scores and observed scores.

Sixty different two minute videotapes were constructed of white or black college students shelving different numbers of books at a library. Ratings were made on performance quality and confidence in the performance rating. Significant effects occurred for race of rater and sex of ratee. White raters were significantly more accurate than black raters. Male ratees were rated significantly more accurately than female ratees. Black raters were more accurate at rating black males than black females and white raters were more accurate at rating black females than black males. Ratings of black ratees by black raters and of white ratees by white raters were more accurate than cross race rater-ratee combinations. White male ratees were rated most accurately of all ratee race-sex combinations and white females were least accurately rated. Black females were rated more accurately than black males.

Schmitt and Lappin (1980) conclude their study by comparing the amount of variance in ratees' scores due to the true scores and to other factors. Approximately 12% of the variance in true score could be accounted for by the rater and ratee race and sex variables. This is a statistically and practically significant amount. However, roughly 70% of the variance in true scores could be accounted for by rated performances. Ilgen, Barnes-Farrell, and Mckillin (1987) point out that the finding of accuracy being higher for raters rating members of their own race is not inconsistent with previous research findings that raters tend to rate members of their own race more leniently. The

reason for the absence of a contradiction between these two findings is because Schmitt and Lappin (1980) use a correlational measure of accuracy. Correlational measures are not sensitive to mean differences.

Research on Factors that Affect Categorization

A number of research studies have shown other factors that can have significant effects on rater cognitive processes. These factors serve to further restrict rater differential accuracy.

Rater Factors that Affect Controlled or Automatic Categorization. Kozlowski, Kirsch, and Chao (1986) conducted a study that assessed the effect of a rater's preexisting cognitive schema on the accuracy and halo of performance ratings. The study tested the implications of Cooper's (1981a, 1981b) semantic conceptual similarity cognitive process. Cooper proposed a systematic distortion hypothesis which states that a rater will observe, store, and recall information relating to a rating task by systematically biasing the retained information in the direction of the rater's "implicit covariance schemata." More specifically, Kozlowski et al. hypothesize that a rater's personal cognitive schemata will be more likely to lower rating accuracy and increase halo when the rater lacks job knowledge, lacks knowledge of the person being rated or must recall information from memory.

Raters who knew a great deal about baseball (i.e., had high job knowledge) were more accurate (as measured by average intercorrelations between ratings and true scores on each of the seven rating dimensions) than were low job knowledge raters. The high job knowledge raters

relied on actual data when they knew the ratee well, but relied on internal, conceptual similarity schemata when the ratee was not well known to them. The low job knowledge raters were more sensitive to their own conceptual similarity schemata, even when the ratee was well known. Kozlowski et al. (1986) concluded that the low job knowledge raters relied on their conceptual schemata, rather than the actual data, even when they had relatively more knowledge about the ratee.

The importance of this study is that job knowledge (albeit about the "job" of baseball batters) was found to be an important consideration in rater accuracy and halo. High job knowledge raters (i.e., experts) rely on data in well understood situations and on conceptual similarity data (stereotypes) in low knowledge situations. The low knowledge raters (i.e., nonexperts) always rely on the conceptual similarity data (i.e., stereotypes). The accuracy measure used in this study, however, was a correlational one, over only seven data points which were objective measures of performance. The Cronbach (1955) accuracy measures were not used.

Kozlowski and Kirsch (1987) published a second study utilizing baseball batters as the target ratees on whom accuracy could be checked. In this second study, Kozlowski et al. (1987) were able to analyze the accuracy of the individual raters with respect to Cooper's (1981) systematic distortion hypothesis.

Expert raters, those with more (baseball) job knowledge, provided ratings that were more accurate on stereotype and differential accuracy (plus difference score and overall correlational accuracy measures and less halo) than raters with less job knowledge. In addition, high job

knowledge raters showed significantly smaller correlations between their conceptual schemata and their ratings and significantly larger correlations between their ratings and the true scores than did less knowledgeable raters.

Valenzi and Andrews (1973) present a study in which raters using more "clinical judgement" (as opposed to actuarial or statistical style) job applicant rating techniques will make reliable ratings, but may be inaccurate compared to their intended methods of utilizing the information available. In this study, four placement interviewers rated 243 different applicants (twice to assess reliability) based on written descriptions of five dimensions critical to the job of secretary. After rating the "applicants," the judges independently decided on the relative importance of the five different dimensions on which each of the applicants had been rated.

One of the four raters used a statistical judgement technique and was dropped from the analyses. The other three raters each used "clinical" methods of rating the candidates. The intercorrelations between the latter three raters ranged from .72 to .79. The rate-rerate reliabilities for these three judges were all above .80, with two of the raters' reliabilities above .90.

The three judges agreed quite closely on the intended or "true" dimension weights. More importantly, there were large discrepancies between the intended dimension weights assigned by each of the judges and the actual weights observed in the ratings. All of the judges had moderate to large rank differences between the actual and the intended cue weights.

Although the judges all agreed on the weights to be used in rating the dimensions, and although the judges were each internally consistent, and although the judges modestly agreed with each other in the ratings, the judges did not utilize the cues consistently with their perceived relative importance. The judges intended to do one thing and ended up consistently (i.e., reliably) doing something else. One factor that may differentiate between expert and nonexpert rater accuracy may be that expert raters are able to make judgements that are consistent with their intended (or the trained) dimension definitions.

Cardy and Kehoe (1984) present another rater factor that proved to have an effect on accuracy. Research participants were first given the hidden figures test (Jackson, Messick, and Myers, 1964) and categorized as either high or low on field dependence. These categorized participants then rated the performance of written vignettes of four hypothetical college instructors. Raters high in selective attention were more differentially accurate (i.e., high on Cronbach's DA) than those low in selective attention. When raters have a tendency to focus their attention on the specific features of an observation situation, they will provide more differentially accurate ratings than raters who skip over the details and focus on the global situation.

Cardy and Dobbins (1986) showed that a rater's liking of a ratee can bias differential accuracy of job performance ratings, and do so outside of the rater's awareness. When liking was varied independently of true job performance (as opposed to being held constant across a number of ratees) liking interfered with the cognitive processing of

observed ratee behavior and lowered subsequent rating accuracy regardless of whether or not the rater was aware of the process. Ratees who were liked were given higher ratings than they deserved and ratees who were disliked given lower ratings than they deserved.

Ratee Performance Consistency. Yountz and Ilgen (1986, cited in Ilgen et al., 1987) developed a set of videotapes of clerical employees. These employees performed either well or poorly and either consistently or inconsistently over time. True scores were developed utilizing the signal detection theory model presented in Lord (1985). This model permitted the assessment of both behavioral (i.e., observational) and classification accuracy. Behavioral accuracy was defined as the proportion of behaviors that the rater was able to correctly identify as being performed by the ratee. Classification accuracy was defined as the proportion of behaviors taken from a list of possible behaviors each at the correct level of performance that the rater correctly identified as being performed by the ratee.

Classification accuracy was found to be higher for the consistently performing clerical employees than for the inconsistently performing ones. For the consistently performing clerical employees, however, raters were more likely to report that they had observed not-performed behaviors that were consistent with the appropriate general classification level (i.e., prototype) of performance than they were for inconsistently performed or not-performed behaviors. Finally, raters in general were more accurate at rating good performers than poor performers.

Ilgen et al. (1987) hypothesize that the reason for this was that the good performing clerical employees were more typical in real situations, and since raters were more likely to have practice in dealing with the consistent performers than the inconsistent ones, the raters would be in a better position to judge such behaviors. In other words, some levels of performance are easier to rate accurately than other levels, perhaps since the raters have more experience observing these levels of performance.

Mount and Thompson (1987) also studied the cognitive processes that affect rater accuracy in performance appraisal. A sample of 255 middle level managers were each rated by their supervisors and their (minimum of four) subordinates. Managers were rated on their perceived role congruence.

Manager effectiveness "true scores" were the average of the manager's supervisor, the manager's own, and average of the manager's subordinates ratings on 14 behavioral statements. Accuracy was defined as the absolute mean difference between a randomly selected subordinate's rating and the calculated true score. This accuracy measure is a combination of the Cronbach (1955) components of accuracy, but has been utilized in other published research (e.g., Bernardin and Pence, 1980).

Ratings on the three performance dimensions were more accurate for managers that were perceived as performing their jobs congruently with the rater's expectations of a good performer than for managers perceived as incongruent. Second, raters who perceived their managers as performing congruent to their expectations of a top quality manager, also rated with more halo than those with less congruent managers.

Finally, raters were more lenient when rating managers who were perceived to be congruent, than raters with incongruent managers. Thus, accuracy, halo, and leniency are all higher when the ratee is seen to be performing within expectations about how a good manager should be performing.

One explanation for these findings has been that ratings are more accurate when the behaviors exhibited are more salient (i.e., within the rater's cognitive conceptual category) and as a result are noticed and recalled more easily. Thus, raters who are capable of perceiving their manager's role within a category congruent with that of a well performing manager, will rate their managers more accurately (and also have more halo and leniency error) than raters whose managers perform their work inconsistently with the rater's expectations. These results, state Mount and Thompson (1987), are consistent with the Bernardin et al. (1982) results that training a rater on the performance rating dimensions increases rater accuracy. The authors do point out, however, that the overall effects of congruency were relatively weak, with omega squares in the range of .03 to .06.

Rating scale factors. DeNisi and Summers (1986) investigated the effect of the timing of the presentation of the rating scale and training procedure with respect to the observation of behavior. Raters were given the rating scales either prior to watching videotapes, after watching the videotapes but prior to being asked to recall each ratee's behaviors, or after watching the videotapes and after being asked to complete the recall task.

Raters given the rating scales prior to observing ratee behaviors were more accurate at recalling ratee behavior and more accurate at rating the behavior. The raters who evaluated ratee performance on task based (rather than trait based) rating scales were also more accurate. Thus, raters given the rating scales prior to observing performance and asked to evaluate that performance on task based rating scales were the most accurate of the six conditions in the study.

Raters using task based rating scales and raters given the rating scales in advance of observation were more likely to have identifiable classification schemas. The raters who used these more consistent, easily identifiable classification schemas to organize ratee behavior were more accurate in recalling the behaviors and in rating performance.

Raters were also classified as using person oriented, task oriented, or trait oriented information storage patterns. Raters with the person oriented classification schemas were highest at recall accuracy. The task based storage patterns had the second highest recall accuracy and the trait based patterns was third. The raters with the person oriented memory patterns were also the most accurate at evaluating the performance of the ratees, followed by the task (performance) and then the trait oriented memory patterns.

Raters given the rating scale prior to observation were both more accurate at recalling behaviors and more accurate at rating the behavior. Task based scales were more accurate than trait based scales.

Person oriented observation classification schemas were highest at recall accuracy. Task based was second and trait based was third. Person oriented schema raters were also highest at evaluation accuracy.

Pulakos (1986) presents a study that investigates the relationship between the rating scale and the type of rater training used. The study investigates the effects of three different training programs (observation or evaluation accuracy training, plus a no training control group) on evaluation and observation oriented rating scales.

Rater accuracy on all four Cronbach (1955) accuracy measures (EL, DE, SA, DA), was higher when the rating tasks and the training condition were congruent. That is, raters receiving observation training were more accurate than the other raters on observation frequency judging (BOS) scales and on elevation, differential elevation, and differential accuracy, but were not higher on stereotype accuracy. Raters receiving observation training and using the evaluation rating scales were higher than the control conditions only for differential accuracy measures. Raters given evaluation accuracy training were more accurate on evaluation (BARS) rating scales on all four accuracy measures than for incongruent matches of training condition and rating scale style. Raters given evaluation training and completing the observation (BOS) rating scales were not significantly different from the no training control condition on any of the accuracy measures. There was no difference between accuracy measures on the two rating scales in the control condition.

Pulakos (1986) concludes her study by suggesting that raters should be trained in accordance with the demands of the rating scale to be used. Utilizing a sophisticated rating scale format does not ensure accurate rating unless raters are also correctly trained.

Ostroff and Ilgen (1985a) provide additional support for the importance of a connection between the demands of the rating task and rater training. In this study, raters were given training that was either functional or dysfunctional. Functionality of training was defined as making the cognitive categories discussed in training consistent with the cognitive categories used in the rating scales. One training program provided feedback to raters on the convergence between rater categories and rating scale categories. The other training program was specifically focused on improving rater accuracy.

Both training programs improved rater accuracy. Both training programs also led to increases in the convergence between raters' use of cognitive categories and the categories used in the rating scales. The training specifically focused on providing feedback to raters on convergence of the cognitive categories tended to provide larger improvements in both cognitive category convergence and rater accuracy than did the general accuracy training. The effect sizes were, however, quite small.

In summary, the Favaro and Ilgen (1983) study presents perhaps the strongest evidence for the existence of a categorization process for storing observations. When ratees were given a quick, easy to use category into which observational information could be placed, they

used it. Little other information was stored. When it came time to utilize the information in making ratings, only the information related to the observation category was recalled.

The categorization process also proved to be subject to situational constraints. Raters would have lower accuracy under conditions where they would not be motivated to rate accurately (Bernardin, Cardy, and Abbot, 1982). For example, raters would tend to be less accurate, probably using less extreme low scores and more high scores than justified, when they were making ratings that would affect an employee's salary.

Categorization was also shown to be subject to manipulation by cognitive processes outside of the rater's conscious control. Liking a ratee would tend to unconsciously lead to higher scores (Cardy and Dobbins, 1986). Raters were even found to deviate from their own intended use for information gathered in observations (Valenzi and Andrews, 1973).

The individual rater's personal skills also came into play in classification accuracy. Raters with higher job knowledge proved to be more accurate than those who were not as knowledgeable (Kozlowski, Kirsch, and Chao, 1986, Kozlowski and Kirsch, 1987).

More importantly, classification accuracy would improve when the rater viewed the ratee as performing more congruently with the rater's expectations of a good performer (Mount and Thompson, 1987). When the rater had a clear picture of a solidly performing manager and the rater

viewed the rated manager as performing congruently with that expectation, then the ratings were more accurate then when the manager was incongruent with the expectation (Yountz and Ilgen, 1986).

The presentation of the rating scales before the observation process took place proved to be an effective way of increasing observation and rating accuracy (DeNisi and Summers, 1986). Changes in the typical rater training program also was found to improve rater observational and rating accuracy. Pulakos (1986) found that when the training and the rating scales were congruent--either both observationally or evaluationally oriented--raters were more accurate.

The research on the integration and storage of observations shows that raters are not simply passive video cameras onto which all information observed is recorded. The process of integrating information in memory is complex. The research cited above supports the Feldman model's use of a categorization process. The next step in the rating process would be to recall the information and make the performance rating.

Recalling and Integrating Information

In the Feldman model, when a rater is presented with a rating scale in a performance appraisal situation, the rater will not recall specific details about the ratee. Instead, the rater will recall the category in which the information about the ratee was stored. This recalled category will contain not only the stored details about the

ratee that fell within the scope of the category, but also the family resemblances of the category described in the prototype, even if these prototype details were not part of the actual observation.

By incorporating the theory of Tversky and Kahneman (1974), Feldman proposes that two recall processes explain how a category is brought to mind in an observation or a rating situation. First, when the observed behavior resembles another behavior or trait, the behavior is associated with the trait. This is the "representativeness heuristic." Raters will assign a greater probability to a behavior being caused by some trait or associated behavior the more the two resemble each other. Second, when viewing a behavior, if raters can easily call to mind a trait or behavior, then that easily recalled trait will be associated with the observed behavior. This is the "availability heuristic."

If, for example, the rating scale contains behavioral anchors, these anchors may cause the rater to recall "representative" or "available" categories and observations classified within them. The rating scales may inadvertently cause the rater to recall confirmatory evidence. The same would be true for the process of storing observations in memory. If the rating scale makes "available" or "representative" certain categories, then these categories may be used to store information. This set of processes is another key point in the Feldman model. If memory categories exist, then the next question becomes how they are invoked.

Considerable research has focused on the issue of how training affects the information that a person can recall about an observation. This research supports Feldman's hypothesis that instructions given raters during training and presentation of the rating dimension themselves prior to observation can strongly influence the type and accuracy of information stored and recalled by raters.

Research on training information storing techniques

Foti and Lord (1987) presented raters with one of three types of rater training. After the training, raters viewed a fifteen minute videotape of a mock board meeting. In the memory training condition, raters were told to try to remember as much information as possible about the behavioral events. In the impression formation condition, raters were directed to view the videotape and form an impression of the chairman of the meeting. In the control training condition raters were told that they would see a videotape of a board meeting created using professional actors and actresses. The job of the rater in the control condition was to evaluate the realism of the roles and behaviors in the videotape.

Half of the participants in each of the three memory training conditions were given information about the goal (or purpose) of the meeting (the "goal knowledge" condition). The other half of the respondents were not given any information about the purpose of the meeting.

Foti and Lord (1987) considered the respondents in the memory condition to be placed in a position were they would be forming a "script" based memory of the events. In other words, the memory condition subjects would be forming more temporally sequenced memories of the videotape than other conditions. Participants in the impression formation condition would be lead to form impressions of the leader. These leader impressions would be centered around the prototype category of leadership.

Both recall and recognition memory are studied. Participants told to try to recall as much information as possible (i.e., the memory condition) remembered a greater proportion of temporally sequenced information ("scripts") than did respondents in the impression formation condition. Likewise, participants in the impression formation condition, who were told to try and form an impression about the group leader, recalled a greater proportion of behaviorally descriptive statements about the board chairman than did the respondents in the memory condition.

Participants told of the goal of the to-be-viewed videotape and directed to remember as much as possible (i.e., the goal and memory condition) recalled a greater proportion of temporally sequenced (script) information than subjects not told the purpose of the group. Interestingly, recall was better for prototype and antiprototype (i.e., items obviously the opposite of the prototype) information than for prototype neutral information. In the impression formation condition, on the other hand, the participants not told the goal of the group

recalled a greater proportion of prototype information than participants told the goal of the group. Recall memory in this case was better for prototype and prototype-neutral information than for antiprototype information.

The subjects told to form an impression and told the goal of the group tended to recall the same relative amount of information as the no-goal-knowledge condition and less than the script-purpose and no-purpose conditions.

Participants told of the goal of the group, no matter which memory training condition they were in, tended to recall a greater proportion of the script items in the correct temporal order than participants without knowledge of the group goals. The observational purpose training manipulations did not have any significant effect on the temporal order of recall of information.

In the recognition memory portion of the study, there was no significant effect of the memory training manipulation. There was also no effect for the manipulation of the knowledge of group goals on memory for script or prototype recognition items. There was, however, greater ability to recognize prototypically neutral than for prototypical events and behaviors. For script events, as the prototypicality of a recognition memory test item increased (i.e., as the item became more and more closely associated with the occurrence of a board meeting), the number of false positive recognition errors increased in the recognition of prototypical items. In other words, many absent, but prototypical script items were falsely recognized. Participants were able to accurately distinguish between the presence or absence of behavior

(i.e., better recognition accuracy) for neutral and antiprototypical items than for prototypical events and behaviors. For prototypical items, participants were better able to accurately recognize present prototypical items than accurately eliminate non-present prototypical items.

For leader (i.e., prototype, rather than serial script recognition) items participants were more accurate at recognizing the present prototypical items than the absent ones. The respondents were less accurate at recognizing the present antiprototypical items than the absent antiprototypical items.

Foti and Lord (1987) conclude this section of their analysis by addressing the question of which type of information remembering scheme was the most accurate. The participants who were directed to utilize the script schemas (rather than the leader impression formation prototypes) were more accurate in recognizing script items and were equally accurate in recognizing the prototypical and neutral leader prototype items than were the respondents using the person focused schemas. For the antiprototypical items the script schema respondents were more accurate than those using the person schemas. In general then, the persons using the script schema were more accurate than those using the person focused schemas.

Foti and Lord (1987) also measured the respondents reaction time for responding to the various items and the respondents confidence in the ratings. The two major factors in the design, observational purpose and goal knowledge, had no main effect either on reaction time or rater confidence. However, raters in the script memory condition had

slow reaction times and low confidence ratings for absent prototypical items. Raters in the person oriented memory impression formation condition had the lowest confidence ratings and the slowest reaction times for neutral (especially absent neutral) items. These raters were fastest and most confident in rating the presence or absence of anti-prototypical items.

In general the confidence ratings were negatively correlated with rater accuracy. Quick responses, however, were accurate for present prototypical items for both script and leader items. The slow responses were accurate for the absent leader items that were neutral or antiprototypical. Foti and Lord (1987) conclude that the goal of a rater's observations will affect the processing of the information. In addition, accuracy was negatively correlated with prototypicality for both script and leader impression formation items. The memory instruction pattern that focused respondents on remembering script based rather than impression formation information tended to result in greater behavioral accuracy.

In sum, the participants who were told of the goal of the group or who were told to remember what they were observing in the group were found to be using a "script" based memory of the events. The participants who were told to form an impression of the leader of the group or who were in the control condition were found to be using schema category prototypes for classifying their observations. When asked to recognize which events had actually occurred during the board meeting, participants in general were less accurate, took longer to respond, and had less confidence in their ratings for prototypical behaviors than

for prototype neutral actions, and took longer for antiprototype behaviors than for the prototype neutral behaviors. When asked to recognize leader behaviors, participants generally were less accurate, less confident, and took longer to respond to the absent prototypical behaviors than to the present prototypical behaviors. In addition, these raters were also less accurate, but more confident in recognizing present, antiprototypical items than prototypical or neutral items.

Research on information recall and search processes

Two previously mentioned studies provide evidence that categorization of observations causes raters to seek confirmatory evidence and prevents them from seeking or recognizing contradictory evidence.

The Williams et al. (1985) study assessed the effects of the purpose of a performance appraisal rating on the type of information sought by the raters. Raters sought different types of information depending on the purpose of the rating. Williams et al. conclude, despite some evidence that raters do seek some common types of information, that raters are sensitive to the demands of the rating situation and seek information relative to those demands.

In rating the performance of subordinate nurses, raters in the Favaro and Ilgen (1983) study spent less time observing a nurse labeled as a "social activist" because observation tasks were limited to confirming the category prototype (i.e., stereotype). Raters were more accurate the longer the amount of time spent observing the ratee.

Accuracy dropped when rating easy to classify people because less time was spent observing their performance. Ratings were made based on the stereotype.

Integration and judgement processes

The final step in a performance rating process is to integrate the recalled observation categories and the rating dimensions. The result is a rating assigned to the ratee. Feldman notes that there are two types of integration processes: cognitive and evaluative. As mentioned previously, Feldman proposes that information is stored separately from evaluations. A rater can form a positive or negative evaluation almost immediately upon seeing an action by the ratee. A judgement or decision will wait and be made upon information stored in the category.

Attitudinal or evaluational integration yields a trait or feeling statement. Feldman borrows the information integration model of Anderson (1974). Anderson's weighted average model suggests that new information is averaged into the rater's current impression about a ratee. A rater may initially change rapidly in impression of a ratee, but as more data is gathered, the impression will stabilize toward a mean.

Cognitive integration is hypothesized to be a result of a consciously controlled attribution by the rater. A belief statement or prediction of future behavior of the ratee is formed. This controlled process is proposed to occur only if the automatic process proves to be unable to categorize the observation. The less easily categorized the

behaviors, the more likely it is that the categorization processes will be biased by the various misattributions that can plague a rater. The more biased the attributions, the lower will be the rater's accuracy.

Feldman's model begins to take on the appearance of asserting that information once categorized never changes. Common sense dictates that this is not the case. Feldman invokes attribution theory to explain how and why observations once categorized may switch categories. Attribution theory was used previously in the model to explain how a rater forms a category for the storage of a ratee's behavior that does not fit into the category automatically brought to mind during observation. Feldman postulates that some threshold or trigger must be reached in order for an observation to be placed in an attribution created category, however, he does not speculate more about that trigger. When this threshold is reached, the two processes of attribution theory (i.e., covariation and discounting) are applied.

Osburn, Timmreck,, and Bigby (1981) present a study that shows that raters will make accurate distinctions between the quality of applicants on job relevant dimensions and less accurate distinctions on more generic rating scales. As is true of most selection interviews, this study utilized unstructured interviews. The accuracy methodology was chosen to assess the effect of directly relevant job descriptions on the accuracy of interview ratings. Participants were 52 professional interviewers.

Job descriptions were written for two different clerical employees: an administrative secretary and a generic office clerk.

Two different, adjective anchored graphic rating scales were developed for each job, plus an overall qualifications rating scale and an acceptability for hire rating scale.

Two hypothetical applicants were scripted and videotaped. The first tape was scripted to contain high levels of administrative secretary skills and average and low levels of the other skills. The second applicant was scripted to be performing at high levels on the general clerical skills with lower skill levels on the administrative secretary skills.

The data showed that raters could make strong distinctions between applicants on the relevant dimension scale and no distinctions in the generic rating scale condition. The Osburn et al. (1981) study indicates that under conditions of two widely different jobs, accurate distinctions can be made between two differentially qualified applicants. In these extreme conditions, raters make more accurate judgements when they are given job information that accurately describes a candidate's performance than when given information that does not closely relate to the candidate's actions.

Making Performance Rating Judgements

Evaluations of the observed behavior are made immediately upon observation and are stored separately from the factual information. Judgements and decisions, however, are based on the information stored in the category. Evaluations can be recalled without the rater remem-

bering any of the categorized factual information stored in memory. The reverse is not true. Recalling information stored in the category will bring with it the evaluative component as well.

When a formal judgment on a performance rating is required, the rater must recall the stored information about the observations. Recalling an observation is actually recalling a category prototype into which the ratee's actions were integrated plus as much of the real situation as was connected with the prototype used to map the observation.

One possible inference from the Feldman model is that ratings will be accurate to the degree the category prototype used in observing, storing, and recalling the behavior agree with the category prototypes used in the rating scale completed by the rater (and possibly the category prototype followed by the ratee when performing the behavior).

The key to rater differential accuracy would then be the degree of overlap between the category prototype used by the rater in observing the ratee and the category prototype written into the rating scale used by the rater to assess the ratee. If the two prototypes overlapped completely (and situational influences were held to a minimum) then rater differential accuracy would be higher than if they were different. Furthermore, if the category prototypes used in the different rating scales were independent of each other (i.e., had no category prototype family resemblances in common with each other), then differential accuracy would be higher than if the prototypes were not independent.

When a judgment on a performance rating is required, the rater must recall the stored information about the observations. Recalling an observation is actually recalling a category prototype into which the ratee's actions were integrated plus as much of the real situation as was connected with the prototype used to map the observation.

These categorized observations, and not the actual details, are mapped onto the performance rating scales. Performance appraisals and other rating scales, however, are typically multidimensional. They differ from the categories used in observation by using a variety of prototypes within a single rating dimension instrument. Behaviorally anchored rating scales, for example have prototype definitions at each scale value. Each of these multiple prototypes would presumably have their own category. The separate rating dimensions are each categories with their own prototype. These rating dimensions are seldom independent of each other. If they are not independent, then rating one dimension may interfere with the rating of the other dimensions. Within a rating dimension, the scale anchors are usually ordinally scaled on the same named rating dimension. This may not be sufficient for rater differential accuracy. A single unidimensional rating dimension may contain several distinct prototypes. In addition, the various unidimensional rating scales may be defined by prototypes with elements in common with other rating scales.

Expected outcomes for various scale formats in formal ratings

One of the more interesting portions of the Feldman (1981) article is his discussion of the implications of the rater cognitive process

theory for different types of rating scales. Landy and Farr (1980) had noted that the research emphasis on developing new types of rating scale formats was generally not profitable. This caveat was, to a large extent, the cause of the cognitive models of rater processes developed. Feldman discussed why each of several popular types of rating scale formats were not effective.

The most obvious problem, in light of the Feldman theory, would be with global ratings of effectiveness. Since evaluations are made directly from observations and are stored apart from categorized information, global effectiveness ratings need not be based on any facts. Trait ratings would be based on the categorized information, but would be based on the category prototype and not the actual category member. Forced choice ratings cause a rater to make a rating decision based on category membership. Raters will tend to make decisions based on the category prototype, and not the actual data. On behaviorally anchored rating scales (BARS), the rater, according to Feldman, may review the behavioral anchors and simply match up an evaluation with a scale anchor. In this case, the BARS would be little more than a global effectiveness rating. The problem with behavior observation scales (BOS) should be obvious, they would not be based on observations at all, but rather projections about behavior frequency based on category membership.

Training procedures could be used to establish the correct observation categories into which observations should be placed. These observations have proved to be an impetus for renewed research interest in rater processes and rating mechanics.

Research on rater accuracy and the rating process

Rating scale factors. The performance appraisal rating process, which is the main focus of the Feldman model, is a complex process for a rater. Observations over a long period of time must be integrated into ratings on a single rating form. The model can also be examined under more manageable circumstances. But even optimal conditions for observation, the rater may be forced to use a rating scale category different from the prototypes used during observation. Two studies are relevant to this possibility, Borman (1979b) and Ostroff and Ilgen (1985b).

Borman (1979b) studied the reliability and convergent and discriminant validity of five rating scale formats. Stimuli were the Borman (1975, 1978, 1979a) videotapes of recruiters and managers. The summated scale had lower halo than BARS, numerical, behavioral summary, and trait scales. Training reduced halo for the numerical scale. Convergent validities were constant and in the low .70s. Discriminant validities ranged .20 to .36. Borman's DAr ranged from .68 to .93. Numerical scales were best on accuracy. The sophisticated and costly scales like the BARS did no better than the simple numerical rating scale.

Borman notes that some dimensions were more difficult than others to rate accurately, no matter which rating scale was used. One possibility suggested by Borman (1979b) is that accuracy ratings were highest on those dimensions that were understood clearly not only by the raters, by also by the actors in the tapes as well. The raters who

have a clear operational understanding of the rating dimensions will be the raters who will tend to make the most accurate ratings. These raters will also tend to make more accurate dimension ratings when the ratees also have the same clear understanding of the operational definition of the rating scales.

One of Borman's most interesting conclusions concerns the effects of rater training. He concludes that interrater agreement may be increased when three conditions are met. First, the observation of behavior should be standardized in the training. Second, a common frame of reference or performance standard should be used by all of the raters. Third, the relative importance of various behaviors on the different dimensions to be rated should be agreed upon by the raters. The raters should understand which behaviors performed by the ratee are relevant to the dimensions being rated.

The importance of this study is that Borman (1979b) shows that rater accuracy can be manipulated by varying both the rating dimensions and the scale format. Borman suggests that accuracy may be highest for those dimensions most clearly understood by both the raters and the performers of the dimensions.

Another study discussed by Ilgen et al. (1987) is the study by Ostroff and Ilgen (1985b). Although the details of the study are not discussed, the study focused on the different cognitive categorization styles that raters could use. Overall difference score accuracy was found to be improved to the extent that cognitive categories used by the raters and designed into the rating scales were consistent. In

addition, previous experience in rating and the rater's ability to differentiate among the rating dimensions both lead to improvements in accuracy. Two other variables studied, job experience and job position, did not relate directly to rater accuracy, but they did have an influencing effect on the cognitive categories used by the raters. Finally, Ostroff and Ilgen (1985b) found that a match between the trait or behavior orientation of the rater and the rating scales did not affect rater accuracy.

Rater Factors. The Bernardin, Cardy, and Abbot (1982) study, mentioned earlier, is also relevant in the context of rater factors that have an effect on rating accuracy. Bernardin et al. (1982) found that differential accuracy was higher for motivated than for not motivated raters. When motivation was added, the rater's with similar prior belief and rating scale categories had better differential accuracy than raters whose prior belief categories were different from those in the rating scale.

In the Borman (1979) study he once again uses the two videotaped job classes (recruiters and managers). The 14 raters in the study had convergent validity correlations of .64 for the rated recruiters and .69 for the rated managers. The ratings had discriminant validities of .57 for recruiter ratees and .85 for managers. This strong support for the accuracy of the scale is reinforced by the rater by ratee ("halo") interaction correlations of .12 for recruiters and .16 for managers.

Becker and Cardy (1986) analyzed Borman's (1975, 1978) validity / accuracy correlations and pointed out that accuracy correlations mea-

sure a rater's ability to rank order the ratees within the dimensions. Becker and Cardy (1986) conclude that the Borman correlational accuracy measures are a combination of Cronbach's (1955) differential accuracy (DA) and differential elevation (DE) correlations. High correlations in the Borman studies utilizing the measure may mean either that the rater was accurate in ranking ratees within the separate dimensions or they were accurate in ranking the ratees on total scores averaged across the dimensions.

Borman (1978) presented videotapes to a group of 146 college students and calculated convergent and discriminant validity statistics. The students did as well as the expert raters on the convergent validity correlations (.53 and .69 for recruiters and managers, respectively). The discriminant validity correlations were only half as large as those of the expert raters (.28 and .25 for the recruiters and managers, respectively). This trend is reflected in a doubling of the halo correlations.

Expert and nonexpert raters were equal in convergent validity, but the nonexperts had only half the discriminant validity of the experts. The expert's greater familiarity with the rating dimensions is mostly measurable in terms of their ability to differentiate between the rating dimensions. They know not only the general content of the rating scales, but also understand more clearly to what the rating scales <u>do not</u> refer. Additional analyses indicate that the raters disagreed substantially on the ratings assigned for some of the dimensions.

Rater training. Borman (1975) measured the effects of a brief rater error reduction training session on the reliability and convergent and discriminant validity of the ratings. Participants in the study were 90 low and middle level managers from an insurance company. The study used six videotaped vignettes.

Pretest to posttest reliability correlations were high and halo variance dropped from pretest to posttest. Intraclass correlations were very high. Raters had more variance (and as a result less halo) in the dimension ratings for the applicants in the posttest. Individual dimension accuracy correlations ranged from .60 to .91. Correlation validity estimates were in the low .90's for both pretest and posttest and were not affected by the training as a whole.

Bernardin and Pence (1980) assessed the effects of two different types of rater training on the psychometric qualities of the ratings. Raters were trained either to minimize rater errors (called the RET group) or to maximize rater accuracy (the RAT group). There was also a no training control group.

The three groups rated videotapes containing two different professors. True score were developed for 13 dimensions using a relatively weak criterion in a retranslation and scaling methodology (Smith and Kendall, 1963). The RET error training group showed less lenient ratings than the RAT accuracy group and the control group. The RET group had lower mean dimension intercorrelations (the halo measure) than the other conditions. A post hoc contrast showed that the RAT and control groups both had significantly higher mean deviation accuracy than did the RET group.

Bernardin and Pence (1980) suggest that raters should not be given error training in which raters are shown actual examples of numbers assigned to candidates. The results may suggest some additional conclusions. Despite the problems in this study, error training and accuracy training showed relatively independent effects. Raters may learn one of the skills and not the other. Expert raters may be better than nonexperts at both avoiding errors and accurately rating the performance of ratees. Nonexperts, alternately, may learn one or the other of these skills, but not both.

Pulakos (1984) conducted another study assessing the differential effects of rater error training (RET) and rater accuracy training (RAT). One group of raters received error training (cf. Latham et al., 1975), a second accuracy training, the third a combination of accuracy and error training (RET/RAT), and the fourth no training.

Upon completion of the training participants watched the six Borman (1977) videotapes of managers dealing with a problem subordinate. Each tape was rated on five behaviorally anchored rating scales. In an interesting aside, Pulakos notes that originally the raters were required to make ratings on seven different rating scales, but pilot research indicated that the accurate raters could not assimilate that much information in the rater accuracy training session. Two dimensions from the original group of seven were randomly deleted.

Raters receiving only accuracy training (i.e., the RAT condition) showed significantly higher differential accuracy than any other type of training on five of six dimensions. The RET and RET/RAT training groups were not different from each other on accuracy, but were both

better than the no training control condition. Pulakos concludes that raters who receive error training will have fewer rater errors than raters who do not receive error training. Likewise, raters who receive accuracy training will display more accurate ratings than raters who do not. Raters who receive both forms of training will show some signs of both types of training, but will also show some deterioration of accuracy and error reduction, perhaps due to the shorter amount of time focused on the respective topics.

Specifically, raters can be trained to be more accurate by focusing on the "dimension structure" of the rating scales. In addition, by focusing on the particular effective, average, and ineffective behaviors that relate to the various dimensions, the raters are given cues that are easily detectable.

More important for the present discussion was the fact that accurate raters differed in their accuracy depending on which rating scales were being rated. Pulakos' post hoc analyses of the data suggest that RAT raters were more accurate on the dimensions that were objectively defined in terms of the particular effective, average, and ineffective behaviors. The scales that were rated most accurate by the RAT raters were scales that were less ambiguous.

Conclusions Based on the Feldman (1980) Model Research

Feldman (1980) lists three sources of bias that can lead to lower accuracy. He proposes that the time dependent nature of performance ratings is the major source of bias. Since ratings are completed some time after the observations, ratings must be based on information

stored in long term memory. The rater will be forced to store observations in memory categories defined by prototypes. These categories can be manipulated by situational and contextual factors as well as by training. Recall of categorized information can also be biased. The categories used as the bases for storing the observations can be subjected to stereotypic over-interpretations of information and to the rater's own subjectively developed implicit personality theory. Any difference between the category used in observation and the category used in the rating scales will lead to a lower rater accuracy.

Several research studies investigating Feldman's hypotheses have shown factors that interfere with accurate rating. First, the recalled observation category will contain details from the category prototype not found in the original observation. Second, the recalled observation category may not have a one-to-one link with the rating dimension category used by the rater. Since category usage in observation is an automatic, rather than a controlled process, the observation categories may be different from those intended for use by the rating scale developers. Third, just as the biasing effect of liking is not relevant to the ratings in the Cardy and Dobbins (1986) study, other factors may unconsciously bias the integration and processing of the observations and the mapping of the categories onto the rating dimensions. Even though the rating scale categories may be statistically independent (and many are not) the ratings on these scales may be based on an observational prototype that contains information that interferes with the intended interpretation of the rating scale dimension.

Even a rating based on a single observation, such as in a job simulation examination, may try to force the rater to use a number of categories to store the observed information, while in reality the rater may use only one or two. The stored category prototype(s) may be different from the prototypes used in the rating scales. The observation prototypes would then need to be translated into the rating scale dimension prototypes. The possibility of these several observational and rating scale prototypes overlapping, even with training, is less than perfect. Given the hypothesis that observations are stored imperfectly in categories, and the added requirement of translating the less than perfect observation based category members into the rating scale dictated categories, judgements made based on the original observations will be far short of perfection.

A single observation category may be used to generate ratings on two or more rating dimensions. Each rating dimension may contain several different category prototypes. Each of these intended categories may be rated based on a single categorization of the observed experience. The single prototype holding together the observation may cause the ratings, which are intended to be independent from one another, to actually interfere or interact with each other.

These categorized observations, and not the actual details, are mapped onto the performance rating scales. Performance appraisals and other rating scales, however, are typically multidimensional. They differ from the categories used in observation by using a variety of prototypes within a single rating dimension instrument.

Behaviorally anchored rating scales, for example, have prototype definitions at each scale value. Each of these multiple prototypes would presumably have their own category. The separate rating dimensions are each categories with their own prototype. These rating dimensions are seldom independent of each other. If they are not independent, then rating one dimension may interfere with the rating of the other dimensions.

Within a rating dimension, the scale anchors are often ordinally scaled on the same named rating dimension. This may not be sufficient for rater differential accuracy. A single unidimensional rating dimension may contain several distinct prototypes. In addition, the various unidimensional rating scales may be defined by prototypes with elements in common with other rating scales.

In addition to this problem, a number of research studies have shown that other factors can have significant effects on the cognitive processes involved in the rating process that serve to further restrict rater differential accuracy.

For example, Cardy and Dobbins (1986) have shown that a rater's liking of a ratee can bias differential accuracy of job performance ratings, and do so outside of the rater's awareness. When liking was varied independently of true job performance (as opposed to being held constant across a number of ratees) liking interfered with the cognitive processing of observed ratee behavior and lowered subsequent rating accuracy regardless of whether or not the rater was aware of the

process. Ratees who were liked were given higher ratings than they deserved and ratees who were disliked given lower ratings than they deserved.

Four important conclusions are drawn from the research on the Feldman (1980) cognitive model of rater processes. First, raters may have problems translating recalled observation categories into rating scale categories. Second, many rating scales may force a split into a rater's single observation category. Third, many of the categories used in performance rating scales may not be independent, causing information on one "independent" rating to affect the ratings on another supposedly independent rating. Fourth, factors such as liking may unconsciously interfere with categorization.

HYPOTHESES BASED ON THE FELDMAN RATER COGNITIVE PROCESS MODEL

The Feldman model proposes that ratings will be accurate to the degree the category prototype used in observing, storing, and recalling the behavior agree with the category prototypes used in the rating scale completed by the rater (and possibly the category prototype followed by the ratee when performing the behavior). Based on the research relevant to the model cited above, several conclusions can be drawn about where additional research may be productive.

First, accurate ratings may require the rater's categories and the rating scale's categories to overlap. The key to rater accuracy is the degree of overlap between the category prototype used by the rater in observing the ratee and the category prototype written into the rating scale used by the rater to assess the ratee. If the two prototypes overlap completely (and situational influences are held to a minimum) then rater accuracy may be higher than if they were different. Furthermore, if the category prototypes used in the rating scales were independent of each other, accuracy would be higher than if the prototypes were not independent.

Perfect overlap of rating dimensions and rater cognitive categories would be difficult to achieve. Rater categorization is subject to a host of influences. One of these influencing factors would be the potential problem of the rating on the various rating dimensions interacting with each other. The ratings on the scale dimensions may be based on a single observation category. The ratings would have the information in the category prototype in common.

In the early research on rating accuracy, researchers sought to show that one way of improving accuracy was to use experts. The experts, however, usually proved to be just as inaccurate as nonexperts. The Feldman model may suggest why this is the case and where experts may outperform nonexperts. They share the problems of basing multiple ratings on a single (or small number of) observation categories. The actual difference between experts and nonexperts may be that the experts may use more of the categories required on the rating scale during observation.

In addition, the choice of memory category (or categories) used during observation, integration, and judgement processes may be outside the awareness of the rater. The rater may add unintended facts and evaluations into the understanding of the rating scale dimension prototypes. Situational factors and unanticipated cognitive factors similar to liking's interference, may further affect rater accuracy. These effects may happen no matter what the level of expertise of the rater. If the experts are better in differential accuracy, this could be due to their better understanding of more of the category prototypes used in the rating scales. Experts may also have a closer link between the categories used during observation and the categories used in the rating scales. The experts may understand more rating dimensions in the manner intended by the scale developers and may also better understand the prototypes within a rating dimension than do nonexperts. A performance rating expert would then be a person who can automatically categorize observed behaviors into the categories used by a specific

measurement device and understand completely all of the prototypes used to define the scale anchors. This definition of an expert may have little to do with a person's job title.

Problems In Past Research and Proposed Improvements

Past research has been done without job analysis

One of the noticeable weaknesses in much of the above research is the absence of a job analysis. Research is typically conducted by borrowing the Borman (1979b) videotapes of managers and recruiters or the Murphy et al. (1982) videotapes of college lecturers. The next step is typically to borrow some performance rating scales. The borrowed videotapes and the borrowed rating scales are then administered within the new design and the data analyzed.

One major improvement in the current research is to conduct a thorough job analysis and test development project prior to conducting the performance rating research. Such a project allows for the measurement of only critical and important dimensions of performance on the job. A job analysis and test development project were completed for the position of first level supervisor of a local welfare department. Based on the analysis, two critical dimensions of performance were assessed: the ability to solve problems and skill in interpersonal relations.

Past research used rating dimensions without theoretical basis

A second noticeable weakness in the above reviewed accuracy research studies is the absence of any theoretical basis for the dimensions being rated. Rating dimensions are usually selected without consideration for whether a category prototype already exists for the concept being measured. Typically, job tasks are written directly into a behavioral observation or behaviorally anchored rating scale. Little effort is made in finding a logical basis for the rating dimensions. In the present study, the two rating dimensions (problem solving and interpersonal relations skills) are defined by researching the relevant literature and establishing firm theoretical foundations for each of the rating dimensions.

What follows is an overview of the rating dimension training provided to the study participants. The training procedures are presented in the following chapter. The detail of the dimension scoring is presented in the Appendices.

Skill in Interpersonal Relations. Supervisory interpersonal relations focuses on the skills necessary for dealing one-on-one with a problem subordinate. It is this specific situation which is simulated in the job simulation exam used in the research study. After the job analysis data pointed out the criticality of this performance dimension, the academic literature on human relations was consulted. A variety of professional training packages were reviewed. Each contained a number of suggestions for improving one's skill in interpersonal relations, but none proved to be able to present a model that

could be used to explain a wide variety of quality levels of performance in the dimension. The clinical psychology and management training literature were then consulted.

The model chosen to be the definition of interpersonal relations for the study is presented primarily in two source books: Interpersonal Living: A Skills/Contract Approach to Human-Relations Training in Groups, by Gerard Egan (1976), and People Skills: How to Assert Yourself, Listen to Others, and Resolve Conflicts, by Robert Bolten (1979). These two books present a wide view of interpersonal relations skills and are able to describe a wide variety of quality levels of performance. In addition, the model is taught in a variety of situations including professional training courses to supervisors and managers.

Interpersonal Relations was dissected into four major components: self expression, responding with empathy, probing and questioning, and challenging. In the typical supervisory problem solving session with a subordinate, the situation presented to the mock candidates in the study, the supervisor typically begins the meeting by telling the subordinate some information about why the meeting is occurring. This, and all other expressions of factual information by the supervisor would be contained in the self expression subcategory.

Once the supervisor has expressed the information, the subordinate would respond. In the problem solving session used in the study, Chester, the problem subordinate, begins to whine and place the blame for any wrong doing on the secretary. To complete the communication loop, the supervisor (or candidate for supervisor in this research) must not

only hear what Chester has to say, but must listen to the meaning. There is no direct way to measure whether the supervisor actually listens to what he or she hears the subordinate say. The second step in the model provides a mechanism for assessing the quality of the candidates listening skills subcomponent of the model. Once the subordinate speaks, the supervisor is required to respond with empathy by repeating back to the subordinate both the information content and the emotional content expressed. This is the second piece of the interpersonal relations model.

These first two pieces of the model establish the communications loop between the supervisor and the subordinate. The supervisor would then be ready to begin to work on the major goal of the meeting: resolving the problem. The third step in the interpersonal relations model is gathering of information that will be useful in resolving the problem. The information content, emotional content, and delivery of these expressions is the third piece of the model.

The final component of the interpersonal relations model is the most difficult. Challenging the subordinate to change his or her inappropriate or below quality behavior should be done only after a solid relationship has been established through the use of the first two steps in the model. The goal of challenging is to provide the subordinate with feedback on his or her past or current behaviors and to seek to motivate the employee to improve future performance.

Each of the four subcomponents of the interpersonal relations model can be divided into three parts. First, each subcomponent will

contain some information content. Second, each subcomponent will contain some emotional content. Third, these two substantive pieces of content must be packaged and delivered to the listener. The third part of each of the four subcomponents is the delivery.

A good metaphor[1] for understanding the three parts of the subcomponents of the interpersonal relations model would be a song on the radio. The song will have some words or lyrics. The lyrics will be the informational content. The tune or the melody that the singer will sing will be the emotional content. Finally the lyrics and the tune are orchestrated into a complete product. The orchestration is the delivery. Each of the four pieces of the interpersonal relations model--self expression, responding with empathy, probing and questioning, and challenging--can each be rated in terms of the quality of their information content, emotional content, and delivery. The interpersonal relations model is presented in detail in Appendix B.

Problem solving using the Scientific Method Definition. The problem solving model is taken from a model discussed by Ernest Archer (1980). The model is a translation of the basic scientific method. Several managerial training programs were reviewed. A major focus of all of these models was training the fledgling supervisor or manager in solving problems. These models proved to be of little value in the present research since they typically presented "helpful hints" or "key principles" to remember when the person is confronted by a problem.

---

1 I am indebted to Richard McGourty (May, 1987) for providing this metaphor of the interpersonal relations process.

What was needed, and what was provided by the Archer (1980) presentation of the scientific method, was a consistent, logical model that described all of the steps that would be involved in solving a problem. This problem solving model proved to be able to describe a wide variety of effective and ineffective methods for attempting to solve problems. The revised version of the problem solving model used here is described in Appendix C.

The problem solving model is analyzed into a five subcomponent, eleven step process. In subcomponent "A," the potential problem must first be identified. This consists of first, monitoring the decision environment to watch for deviations from expectations, and second, defining the problem highlighted in the noticed deviation.

Once the problem is identified, the problem is analyzed (in subcomponent "B") to assess why the problem occurred. In this subcomponent the problem solver also specifies his or her goals and objectives for the situation.

Next, the decision making cycle begins (subcomponent "C"). The supervisor first develops a variety of potential solutions, then develops a set of criteria and methods for appraising the benefits and drawbacks of each possible solution. With the establishment of the criteria, the potential solutions are evaluated, and finally the decision is made on the best solution for the problem.

After the solution is chosen, it is implemented (subcomponent "D"). In subcomponent "E" the implementation of the solution is supervised to make sure the implementations are being carried out according to plan. Finally, in subcomponent "F", which is the same as

subcomponent "A," the decision environment is monitored to watch to see whether the deviations happen again. Each of these steps are discussed in detail in Appendix C.

Past research made no attempt to establish cutoff points

A third weakness in the studies reviewed above is in defining the rating scales. Little effort was made in setting cutoff or passing point scores. The literature suggests that this is a major consideration of the raters. Several studies were cited which showed that the purpose of the rating session proved to have a big effect on the accuracy of the ratings. The present study utilized a group of subject matter experts (SMEs) in setting the cutoff scores. The SMEs reviewed the rating dimension definitions and set minimum performance levels for each subcomponent of the dimensions. The rating scale anchors were set to be logical divisions between the point values assigned by the subject matter experts and the theoretical literature.

Questions on rater expertise

Past research concludes that experts are not more accurate raters than nonexperts. The current study focuses on one reason why this would be true, even if the raters receive the same training. The question is why experts may prove to have more accurate ratings on rating scales than nonexperts when both groups have been given the same amount and quality level of training on the rating scales and the rating process? Two possibilities are suggested.

First, the expert raters have a more thorough understanding of the job for which they will be assigning ratings. The experts will have a better understanding of the dimensions of performance of the job. If a proper job analysis has been conducted, then the dimensions of performance developed should match the dimensions of performance that the experts use. The job analysis should provide information that more closely is linked to the way that the experts think about the job than the way the nonexperts think about the job. Job knowledge should extend to a more thorough understanding of the rating scales and more accurate ratings. Nonexperts will have a variety of idiosyncratic understandings of the performance dimensions of the job.

A second difference between the experts and the nonexperts is in the rater's understanding of the various levels of performance within the rating dimensions. The Foti and Lord (1987) study demonstrated that raters will have different levels of accuracy depending on the degree of prototypicality of the observed behaviors. Even though a rating dimension may be internally consistent (i.e., reliable) the high scoring candidates may be acting out one prototype, while the minimally acceptable performers are acting out a different prototype and the below average candidates are acting out yet another recognizable pattern of behavior. The more knowledgeable raters may be those who understand a wider diversity of prototypical performances, each at different quality levels. For example, there is likely to be a wide variety of methods for handling a particular supervisory situation at a minimally qualified level of performance, and a still wider range of

possibilities for handling the same situation incorrectly. The number of ways for handling the situation absolutely correctly, however, is likely to be very limited.

Summary of the Research Hypotheses

Two variables are assessed in this research. Specifically, do expert raters make more accurate ratings at a variety of levels of performance from below passing to minimally acceptable to highly qualified levels of performance? Second, do raters differ in accuracy on performance dimensions that differ widely from each other, yet are still important components of the job? Can rater expertise be defined operationally as expertise in technical areas such as solving problems or will expert accuracy extend to more broadly applicable performance dimensions such as interpersonal relations? Will expert raters be able to rate more accurately only the more difficult portions of the rating scales? The specific hypotheses are discussed below:

The first hypothesis is that there will be a statistically significant interaction between the two rating dimensions in differential accuracy. This will indicate that a ratee's performance (i.e., the prototype or scale value they display) on the dimension not being rated can have an influencing effect on the dimension being rated. This effect should occur for both rated dimensions.

The second hypothesis is that there will not be an overall main effect for a difference in differential accuracy for expertise of the

rater. Experts will not be more differentially accurate than nonexperts after a training program. However, there may be some interactions between rater expertise and the performance dimension target scores with experts proving to be more accurate than nonexperts at some levels of performance. No specific hypotheses are made as to which scale values will be rated more accurately by expert raters.

Although not technically a research hypothesis, a third finding is also possible. Several studies, especially Borman (1979) suggested that rater accuracy may be influenced not only by the rater's general level of understanding of the rating dimensions, but also the ratee's understanding of the rating dimensions. If a ratee is performing his or her actions with one category prototype in mind, but is being rated on a category with a slightly different prototype definition, then the ratee's actions may not be viewed as categorizable within the rater's understanding. What might typically be viewed as "manipulation checks" on whether multiple performance versions of the same stimulus are in fact the same, may in fact be measures of the overlap of the ratees' understanding of the performance dimensions. For this reason, version differences may also occur if the actors and actresses performing the same script use different prototypes in their performances.

METHOD

Development of the Job Simulation Examination

Job analysis

A job analysis was conducted for the position of entry level supervisor in a welfare department of a local government in a large midwestern city. The job analysis was used both for research and selection purposes. Both projects were done independently and were based on the job analysis data. The selection project was completed and promotions were made before the research project was started.

The purpose of the job analysis was to identify major clusters of frequent and important job tasks and to identify the major knowledges, abilities, and skills required to perform the identified job tasks. Each "Unit Supervisor" is responsible for a unit of eligibility technicians whose job is to determine whether welfare applicants meet the State, Federal or local government requirements for receiving public assistance money. Some units may have a "lead worker" whose job is to either be a technical specialist, a trainer or an assistant supervisor. The lead workers do not perform any hiring, disciplining or firing, but may have input into these processes.

A group of subject matter experts, all long term Unit Supervisors, were gathered into a group to generate a list of job duties and tasks. The list of tasks was divided among the various job duty areas. The list of duty areas was also used as the basis for developing a list of required knowledges, abilities, and skills (KSA's) by another group of

supervisors. These two lists, the duty/tasks list and the KSA's, were reviewed by several other committees and higher level managers for accuracy and thoroughness. Many items were added and/or modified.

The final lists were included in a survey which was administered to the complete set of about 60 Unit Supervisors. The survey was used to select only those tasks which were either frequent or important to the job of the combined supervisors. The importance of this selection was that it cut across the five major divisions in the department (i.e., General Assistance, Neighborhood Assistance, Aide to Families with Dependant Children, and Collection Services). The tasks that were left were those which were either frequent in occurrence or important to the successful performance of the jobs of supervisors in general, across all units in the department.

The survey was also used to select a list of KSA's that were required on entry into the position or were important in distinguishing average from above average supervisors. It is from this list of selected KSA's that this research study was built. The entire job analysis phase of the study was done by the Personnel Representatives of the local government and required about six months to complete. The complete set of data was given to the researcher on which to base this study.

The major dimensions or categories of required job skills on which the research selection test was based were selected from the job analysis information. In addition to the standard frequency and importance considerations, the criteria on which the dimensions were selected included some practical considerations. The research design requires

that several different levels of performance be developed for each selected KSA dimension. In addition, these various performance levels for each KSA dimension must be factorally crossed with those of every other dimension selected. The goal was to select only those KSA dimensions that thoroughly covered the important, required KSA's of the job.

Theoretical basis for the performance dimensions

On the basis the job analysis information two KSA dimension categories were selected around which to develop the selection device. The two selected dimension categories are interpersonal relations and problem solving. These two KSA dimensions covered the range of required and important KSA's quite well. The major KSA dimension that was left out of the research project was technical knowledge. This factor was eliminated from the project because the technical knowledge factor was different for each of the five divisions in the welfare department. No common selection question could be developed as was required for the research design.

Once the KSA dimensions were selected, the dimensions were defined on a theoretical basis. Both dimensions were linked to specific theoretical models. The problem solving dimension definition was based on a definition proposed by Ernest Archer (1980). This definition is based on the scientific method and is very thorough. The interpersonal relations dimension is defined with the model proposed by Gerry Egan (1976) and expanded on by Robert Bolton (1979). The model is based on a clinical psychology process model expanded to include the tasks of an

entry level supervisor. These two performance dimensions are described above and in Appendix B for interpersonal relations and Appendix C for problem solving.

Test development

Once the dimension definitions were drafted, the test development process began. The goal of the test development process was to utilize a job simulation type of examination. These simulations are most easily built based on critical incidents. A series of critical incident generation sessions were held with the supervisors of the entry level supervisors. These Program Managers were all former Unit Supervisors and each supervised over a dozen unit supervisors. Each of three Program Managers was interviewed for two hours. During the interviews the supervisor listed examples of Unit Supervisor actions that were either above average or below average levels of performance.

One consistently recurring theme was dealing with frustrated and "burned out" employees. The job of an eligibility technician is a high volume forms processing job requiring dealing with clients who may hold the worker personally responsible for denying them the complete amount of benefits desired. The workers deal with the stress as well as they can, but after a while even some of the top notch employees become frustrated. Assisting workers in dealing with this frustration is an important aspect of a supervisor's job. A simulation question was developed around a set of such incidences. Several of the common problems were combined into a situation in which a new supervisor was required to discuss some work rule violations with an employee. The

question was written so that the supervisor would be the candidate for the job and would spend a short time discussing a problem situation with a long term lead worker. The simulation item and correct answer are presented in Appendix D.

Once the item was written, an outline version of the problematic employee's situation was also developed. This problem employee (named "Chester" in the question) was built around information gained in the critical incident discussions. A draft of Chester's role was reviewed and modified by the committee of critical incident generating Program Managers. Care was taken not to identify any one individual employee in the department. This was accomplished by manufacturing a cover story for Chester that was the reason for the meeting with the supervisor. Except for the cover story (i.e., a client getting so angry with Chester's lack of service that the client went to the Department Director's home late one evening) all of the details were actual problems. The background information on the item, Chester's script, and the purpose of the items are presented in Appendix E. This information was presented to the participants during the training portion of the data collection process.

The situational question was then pilot tested. Another set of subject matter experts, including Welfare Department Program Managers and long term professionals, supervisors, and managers from the Personnel Department were asked to "take" the test. Twenty-two participants were interviewed, one at a time. In the interview, each participant was given a short time to read the question. When the participant was ready, the "meeting" with Chester began. The "new supervisor" dealt

with Chester's problem in any way he or she chose. The role of Chester was played by the author. Chester was generally resistive to anything the "new supervisor" said. Only strong supervisors were able to "break through" Chester's resistance and accomplish anything meaningful. Each session was tape recorded and typically lasted about twenty minutes. Some sessions were as short as ten minutes. The longest session ran twenty seven minutes. After the "meeting" was finished, the draft correct answer was reviewed. Each participant's responses were discussed and recommendations made for changes in the question or the correct answers.

After the twenty-two sessions were complete the tapes were reviewed and roughly categorized into high, medium, and low quality levels of performance on the two KSA dimensions. The nine tape recorded sessions that most closely met the requirements of the research study were chosen and transcribed. Three tapes were chosen that were subjectively "high" on the problem solving scale and at three different levels of the interpersonal relations scale (i.e., high, medium, and low). Three other tapes were chosen that were subjectively medium on the problem solving scale and at the three different levels of performance on the Interpersonal Relations dimension. Finally, three more tapes were selected that were subjectively poor on problem solving and either high, medium or low on the interpersonal relations dimension. Thus, nine tapes were chosen in all. Each of the nine was chosen to fill a unique combination of performance on the two KSA dimensions.

A committee of Unit Supervisors was formed to develop the correct answers for the supervision simulation question based on the two KSA dimensions. The committee discussed the test question, the dimension definitions, and listened to two selected tape recordings of the "meeting" with Chester performed by participants at subjectively above average and below average levels of performance. The committee then drafted what they considered to be the minimum level of performance of a candidate for the position of unit supervisor on each component of the two KSA dimensions. These unit supervisors also made suggestions for modifying the phrasing of the dimension definitions and for developing ranges of performance on the various KSA dimensions.

Meetings were held until all minimum performance levels were generated and double checked by an independent group of Unit Supervisors. The result of these committee meetings was the development of a "cutoff score" for the test question.

Once the passing points had been thoroughly discussed, the rating scales were drafted. The rating scales were all zero-to-seven point scales with a four point "cutoff" score. The final cutoff scores were written to encompass the information generated by the Unit Supervisor committee and theoretical information. The interpersonal relations dimension theoretical definition also provided information on the degree of effectiveness of particular actions. This theoretical information was utilized in the cutoff scores as much as possible. The problem solving dimension cutoff score was also written to balance the information gathered from the theoretical literature and the cutoff score committee.

The top scores (seven points) on the dimension subcomponents were written to be full performance of the dimension and were essentially a brief summary of the definition of the dimension. The one point level was written to be a performance on the dimension that was the opposite of the dimension definition (i.e., the opposite of the seven point level or performing the required actions incorrectly). The zero point level was written to allow the rater to indicate that a particular dimension subcomponent was not relevant to the rating situation or the rater was unable to rate the dimension. The remaining point values (i.e., two, three, five, and six points) were written logically to fall between the descriptions one point above and below them.

Target script development

The target scores at which the scripts would be written were chosen to be above average but a bit below a perfect score (or six points), an average score but a bit above the cutoff score (or five points), and a below average score, that is, a bit below the cutoff score (or three points). Much debate focused on whether to write a script that was an example of a cutoff score level of performance or a minimally passing score. The difficulty of exactly describing this point forced the choice of the target scores. The scripts written on the basis of these target scores are called the "target scripts."

There were several research and practical considerations that were taken into consideration when developing the target scripts. First, the research design requires that two versions of each script be videotaped. To ensure that each version would have the same content, the

scripts were written down verbatim in advance of the verification process. The research design also required that each research participant view one version of each of the nine possible scripts. Practically, this meant that the scripts had to be kept as short as possible. A target length of about eleven minutes was chosen. The complete set of nine tapes would require nearly two hours to view. This was deemed to be the maximum possible time available to successfully recruit the research participants. Based on these considerations, the nine target scripts were rewritten to about a eleven minute length.

Target script verification

Following the preliminary development of the target scripts, the script verification process was initiated. The purpose of this process was to verify that the target scripts were actually examples of the dimension rating levels that they were targeted to be examples. The first step in this verification process was to modify the rating scales of the dimensions to put them into as logical a progression as possible. The logical progression of the behavioral anchors of each point value would permit easier verification of the scripts. Each scale anchor was written to have directly observable characteristics that were noticeably different from the point values above and below.

The nine, shortened target scripts were then rewritten to match the observable characteristics of the revised scale anchors. The scale anchors were also written into a target script verification rating scale form. The purpose of the form was to explain which paragraphs in

each script were intended as examples of a particular level of performance on each dimension subcomponent. In addition, each verification form listed the target score of which the paragraphs were intended to be examples.

The task of the subject matter experts serving as script verifiers was to read the target score anchor and the linked script paragraphs and decide whether the paragraphs were direct examples of the target scores. In some cases, the scripts were intended to contain examples of behavior either higher or lower than a particular dimension subcomponent. This was true only of the three point (below average) scripts. The overall average score of these scripts would be three points, but some actions were done better and some worse than this target level. Were this tactic not taken, the below average scripts would have all been unrealistically bad. They would not have been examples of the kinds of actions that a candidate would exhibit in such a testing situation.

Once the script verification form was written, the form was used to double check the accuracy of the scripts. The forms were rated by the author and the scripts and verification rating forms modified appropriately. The intention of the double checking was to make the actual verification process as easy as possible.

A procedure for script verification was developed. It was decided to form two committees of target script task verifiers. These committees were formed from the pool of managers and senior level supervisors. Each committee would be responsible for verifying that all nine scripts were accurate examples of one dimension.

Each committee met twice. In the first meeting, the question, the appropriate dimension definition and the verification forms were reviewed. An example of the process to be used during verification was reviewed. Each committee member was then given a complete set of nine scripts, each with a related dimension verification rating scale.

Two weeks later the committee meet again and the results of the verifications was reviewed. If any committee member disagreed with the target score for a particular dimensional subcategory, the matter was discussed. Suggestions were then gathered on how to modify the script to make sure that the dimension would match the target score. Four Program Managers participated in the problem solving dimension script verification process. Four Program Managers and one Personnel Supervisor participated in the verification process of the interpersonal relations dimension.

In each case where a member of the committee rated a particular dimension subcomponent below or above the target score, the member's rationale was discussed. The particular sentences were reviewed by the full committee. If the committee felt that changes were warranted, then the sentence was modified. Every effort was made to ensure that each subject matter expert had his or her suggested changes (or a committee modified version of them) incorporated into the script.

Target script videotaping

Two of the verified scripts were videotaped to be used in initial pilot testing. Each script was taped twice, once with a male and once

with a female role played candidate. Each role player was taped performing two different scripts. Each of the role players was a volunteer. One role player was in her late twenties and the other in his middle thirties. Both of the volunteers were professional staff members of the Personnel Department.

The scripts were videotaped by first printing large sized versions of them and taping them to the walls or to flip charts just out of camera range. All scripts were videotaped in a private office similar to that which might be used by a supervisor in a private problem solving discussion. Emphasis was placed on exact replication of the printed scripts. The author had participated in all of the original scripting sessions, the rewriting, and the verification process. Information gathered from these sources were used by the author in "directing" the reading of the scripts. The author played the role of the subordinate, "Chester" in all performances.

These verified scripts were then incorporated into a pilot study. One version of the two scripts was viewed by a small pool of five subjects. The participants in the data collection pilot study followed the procedure of the main study, except that only two tapes were viewed. At the end of the second tape's rating time, the entire research process was discussed, including the rater training, the rating forms, and the quality of the videotapes.

Suggestions on improvements to the process and forms were incorporated into the final forms used. The data were also analyzed to check whether respondents provided data that followed an understanding of the process.

After the pilot testing on the two versions of the two scripts, the material was revised. The forms were modified to better measure the ratee performance dimensions. The rater training materials and procedure was altered to better allow the candidates to understand the rating dimensions within the half-hour of training time available. The data collection procedure was also modified to allow the candidates sufficient time to make their ratings between the videotapes. The videotaping procedure was also adjusted in an attempt to improve the quality of the recording. The final version of the rater training section of the study was made into an hour long session due to the volume of material to review. Most of this volume was in reviewing and discussing the rating scales.

The remaining videotapes were shot after all other modifications of the data collection process were developed. As was true of the pilot test tapes, the first complete set of videotaped scripts was shot twice, generally once with a male and once with a female role playing candidate-supervisor. The exception was that two males played similar roles on one script. Five males and four females played the roles of the candidates. Most of the role players were professional staff Personnel Department. One role player, however, was an entry level supervisor not employed in the local government unit permitting the research to be conducted. All role players were volunteers.

Role players were assigned to roles at random. There was no attempt to match people to roles. Each role player performed two scripts. The two scripts were at different levels of performance on both rating dimensions (with one exception, where one rater played two

roles at a six point level of performance on rating dimension). Role players were matched on only one script. No two role players were paired on more than one script. Since the research design required all respondents to view a complete set of nine tapes, it was necessary to have nine different actors or actresses play the roles of the candidate supervisors. Had this not been done, one of the persons would have been seen twice performing at two different levels of performance. This is not what would happen in an actual testing situation.

Videotape verification

The set of videotapes was reviewed and modified by another set of subject matter experts. The videotape verification committee was selected from a group of personnel professionals. Each member of the verification committee had been involved in selection development processes and had conducted numerous problem solving sessions with employees similar to the one being simulated in the videotaped exam. The four verification committee members were a personnel supervisor, a senior level administrative assistant, and two senior personnel officers with lead worker and supervisor experience.

The videotape verification committee met twice a week for about two hours over the course of three weeks. The first meeting was devoted to a thorough review of the research process, the simulation exercise, the "true score" answer to the exercise, the rating scales, the typewritten scripts of the videotapes, and the videotape verifica-

tion rating forms. Each of these forms is presented in Appendix F. The remaining five meetings were devoted to reviewing the videotapes, rating the performances, and discussing and modifying the scripts.

The general procedure followed in verifying the videotapes, after the training process, began with a brief review of the target scores for the particular videotape. The target scores were discussed and a general impression of the intended candidate was presented. Once the committee felt comfortable that they knew the intentions of the script, they viewed one randomly chosen videotape of the selected script version. After viewing the tape the committee independently rated the mock candidate on each of the rating scales.

The committee then reviewed their scores. Each time that a rater presented a rating that differed from the target score, the rating was discussed by the group. The videotape verification committee was not restricted to modifying the scripts, they also had input into the wording of the rating scales. The discrepancy in the phrasing of the script or the acting involved in the presentation of the script was modified and agreed on by the complete committee.

The second version of the target script was viewed after the discussion was completed on the first. The committee viewed the second version of the tape and rated the performance independently of the previous version. The second tape was discussed both in terms of the accuracy of performance and the similarity of the performance to the other version. Two different types of errors were possible: differences in both scripts from the target scores and differences between the scripts.

In the first type of error, both versions of the target script could have been found to be divergent from the intended score. Several examples of this type of error were noted in both rating dimensions. In general, errors were noted in the delivery part score of the interpersonal relations dimension subcomponents. Errors were also noted and corrected at the six point level (i.e., outstanding, but not perfect performance) interpersonal relations target score scripts and for the interpersonal relations scores associated with the six point level of problem solving.

For the problem solving dimension, errors in target script accuracy increased as the target score increased. There were no problem solving target score errors noted in the original scripts at the three point level (i.e., clearly below passing levels) of performance. One tentative conclusion reached in watching the pattern of script errors in the videotape verification process was that few accuracy errors occurred in the videotapes performed low quality levels. This was generally true for both the problem solving and interpersonal relations dimensions.

The second style of error found in the videotape verification process was a difference between the performances of the two performers. The script of the more accurate performer was modified until the technical committee felt it would be on target. Then, the performance of the more discrepant performance was discussed. The two versions of the tapes were developed using identical words and phrasings. The differences in the performances would be due to the delivery of the lines. Some actors and actresses were not consistent in their nonverbal and

verbal behaviors. The performers were heard to be saying the correct words, but were not convincing. The committee discussed how the performer could improve his or her performance. Attention was focused specifically on the verbal behaviors and the verbal delivery. Specific examples were highlighted in the scripts where the actor or actress would need to modify phrasings or delivery.

The body language variables proved to be difficult to control. In the actual data collection process raters were asked to tune out these inconsistencies and rate them only in the attribution measures (discussed below). The general feeling of the technical committee was that the nonverbal behaviors were so important that the top quality script could be made into the bottom quality script, and the reverse, by simply changing the manner of delivery. To achieve the goals of the research project, the committee attempted to describe the verbal portions of the body language variable that would need to be consistent with the words. The nonverbal portions were controlled as much as possible and, again, the raters were asked to ignore these differences and focus only on the words.

No general conclusions could be reached about why delivery difference errors were found. Some speculations seem in order, however. The female role players generally were found to have more problems in meeting the delivery necessities of the scripts. Some differences were noted between those scripts that were originally transcribed from male versus female script development pilot test participants. These problems were mostly differences in "male" versus "female" vocabularies.

For example, a male role player referred to Chester's drinking over lunch as "having a few bumps over lunch." This phrase proved to be awkward for female role players.

Another serious problem in meeting the target scores accurately was the difference between the role players natural tendency for speaking and the scripts phrasings. Some of the original scripts were done by managers who felt comfortable in speaking with long sentences, that drifted from topic to topic without noticeable punctuations or pauses. Actors and actresses with different styles of speaking needed to be highly coached before these errors could be corrected.

A third possible source of discrepancy in the delivery of the scripts was familiarity of the role players with the skills of acting. Even though none of the role players were professional actors, several had done semiprofessional, community theater types of acting or had participated in high school plays. These people were much better able to read the lines during the taping without appearing to be doing so and were also better able to look "natural." At times the role players knowledgeable of acting would prove to give performances that were "too good." Much more typical was the finding that the knowledgeable actors and actresses were on target and the others needed to emulate them.

These conclusions and speculations are intended to highlight noted differences. They are not in any way intended to be confused with a systematic and thorough examination of role player accuracy in the videotaping process.

Any script that needed to be modified was done in writing before the script was re-videotaped. The final versions of the videotapes were completed in different ways depending on whether the rewritten script was changed only in wording or if changes in acting were required. The scripts that were only modified in wording were videotaped after only a few practice sessions with the role players. Those revisions that required acting changes began by discussing the changes necessary. Once the changes were explained, the other, correct version of the videotape was watched. Several practice sessions were done and the final reshooting was done. Sometimes several retapings were done. In one case the role player proved to be unable to act out the script as intended. Another role player was recruited and both of the original performer's scripts were reshot and reverified. The final versions of the nine scripts are presented in Appendix G.

Subjects

The participants in the study were from one of two groups. The first group was entry level supervisors. This group is the subject matter expert group. The non-subject matter experts are paraprofessional job class members that do not have any supervisory or lead worker responsibilities. The nonexpert research participants were, however all people who could have applied for the job of entry level supervisor. All participants were volunteers and were not paid for their time. All participants were given time off of the job during the work week to participate in the research. Forty experts and 40 nonexperts participated in the project.

Research Design

The full set of nine videotapes contain each possible combination of the two rating dimensions, each at the three different levels of performance. When raters are permitted to rate all of the videotapes in a set, they may be able to figure out the pattern in the series. True judgements, under these conditions, may not be made. The possibility of this happening is lessened if the tapes are presented in a random order to the candidates.

The above design requires each rater to assess nine different candidates. This is consistent with current professional practice for structured oral examinations. In most oral examination settings the rater may expect to see a series of different candidates over the course of a day. The number of candidates seen in a single day will depend on the length of the examination and the number of applicants. The practice in the jurisdiction performing the study has been to administer five to fifteen oral examinations per day for each oral board.

The nine different videotapes can be classified according to three levels of candidate overall score and two levels of candidate consistency. Table 2 describes the combinations of these variables.

Table 2 Target script scores random orders of presentation

Script Scores	Dimension Actor	Actress/Actor	
		Version One	Version Two
1	PS6IR6	Matt	Barb
2	PS6IR5	Juan	Rachel
3	PS6IR3	Denise	Rick
4	PS5IR6	Nancy	Steve
5	PS5IR5	Sid	Denise
6	PS5IR3	Rachel	Matt
7	PS3IR6	Barb	Sid
8	PS3IR5	Rick	Nancy
9	PS3IR3	Steve	Juan

By having each rater view nine different candidates, actual practice is simulated. Raters are classified into subject matter and non-subject matter experts. A total of 40 raters from each group is required. Each of the videotaped candidates is rated on the two different dimensions: skill in problem solving and skill in interpersonal relations. The 40 raters within each group are divided into groups of five groups of eight raters each. Each of these groups rates the nine candidates in a random combination of the orders of presentation. The design matrix for the study is presented in Table 3. The design is repeated for subject matter expert and non-subject matter expert (between subject variable) categories.

Table 3 The Design Matrix of the Study										
Expert Group	Video Set	Problem Solving (PS) and Interpersonal Relations (IR) Rating Scales								
			IR1			IR2			IR3	
		PS1	PS2	PS3	PS1	PS2	PS3	PS1	PS2	PS3
Expert	1									
Expert	2									
Nonexpert	1									
Nonexpert	2									

Procedure

Participants are categorized as either subject matter experts or non-subject matter experts. Some participants whose jobs included lead worker functions, but no supervisory functions, such as conducting performance appraisal meetings (an inappropriate task at this level), were not used in the data analysis. Since this determination was made in advance and members of these two groups did not interact during data collection, members of the two groups rated the videotapes at the same time.

Subjects within each expertise group were randomly assigned to one of four random orders of presentation. Each of the groups of raters within the expertise groupings rated one of these presentation sequences of one set of videotapes. A second group of ten raters rated the second version of the tapes in a different random order of presentation. Thus, order of presentation was randomized and each cell in

the design contained data from two different videotapes (usually one male and one female role playing candidate-supervisor). This double set of tapes allows for measurement of the effects due to the individual candidate on the videotape.

The two rating dimensions were rated in counterbalanced order within each of the two sets of tapes. Half of the raters in each condition rated the problem solving skill dimension first and the interpersonal relations skill dimension second. The other half of the raters rated the dimensions in the reverse order. Each rater rated the dimensions in the same order over the nine videotapes rated.

When the participants entered the room to begin the study, the researcher gave a brief introduction to the study. The study was explained to be one whose goal is to improve the overall quality of the oral examination process. After the study was explained, the participants were asked to sign the study waver. If subjects chose not to sign the waiver, they were permitted to leave. A total of seven people chose the option of not participating in the study. The general reason was the length of time required to participate (about four hours, with one short break).

Upon receipt of the waiver, subjects were given time to review the test questions and to read the rater guidelines for rating the dimensions. The specific rating scales were examined and the behavioral anchors for each of the rating scales discussed. As is the typical procedure in oral examinations, the response guidelines were reviewed

until there is no more questions from the raters. The rater briefing covers primarily accuracy training. The length of time required for the rater briefing was about one-half hour.

Each videotape was viewed in its entirety and was rated immediately by the raters on the various rating scales. No discussion between the raters was permitted, neither about the content nor the ratings assigned to the videotapes, once the first tape was begun. Each rating was made on a seven point Likert type scale with descriptive anchors at each point value. Each of the eighteen videotapes (two sets of nine) was between nine and thirteen minutes in length. The complete set of training and rating materials and forms is presented in Appendix H.

Two additional sets of ratings were made for each of the dimension ratings on each videotaped candidate. First, the rater was asked to make an overall evaluation of the candidate's success or failure on the exam and rate their perceived confidence in the accuracy of their rating. Second, the rater was asked to make a causal attribution for why they believed that the candidate performed the way that they did.

The rater was asked a series of questions about their causal attributions for the ratings. Items addressed the possible combinations of the location (internal or external to the candidate-supervisor), stability (stable or unstable over time), and controllability (due to factors controllable or uncontrollable by the candidate-supervisor). For example, the raters were asked to state the degree of influence that internal, unstable, and uncontrollable factors

had on their attribution by responding to the item: "How much influence did the following potential causes/factors have on this candidate's performance in this exercise? (e.g., The candidate's good (or bad) mood)".

After rating the last videotape, the candidates were debriefed. They were asked not to discuss the contents of the videotapes with any other people until the last participants completed the research. If they wished, they were sent a brief discussion of the results. The data collection period lasted approximately four hours.

Summary of the Research Variables

The participants in the study were divided into two groups: expert and nonexpert. The experts were all entry level supervisors. The nonexperts were not supervisors or lead workers and would be (or soon would be) eligible to take the examination for entry level supervisor.

The nine different scripts were each videotaped twice. The two performers, generally one male and one female, each said the exact same words. One member of each of the nine versions of the videotapes was randomly assigned to version one or version two. Each version set contained a complete set of possible scores and no actor or actress, other than "Chester," the role-played subordinate, participated in both versions. Candidates viewed the same set of nine actors and actresses, but half of them viewed those actors and actresses doing one version and the other half of the participants viewed the same people perform-

ing a different version on the nine scripts. Any differences between the two versions of the script were due to nonverbal, body language variables ("acting skills").

Each of the videotapes presented one unique combination of the two performance dimensions: problem solving and interpersonal relations. Three tapes were written at a six point rating level (on a seven point scale) for each of the rating parts within the subcomponents for each dimension. Three other tapes were written at the five point level on each dimension. The final three tapes were written at the three point level. The six point level was intended to exemplify an above average performance on the dimension. The five point level was intended to be an example of average performance on the dimension. The three point level was intended to be a below acceptable performance on the two dimensions. (The four point level was the cutoff score.)

Performance ratings on the mock simulation examinations were made at each of three levels of detail by each participant. The problem solving dimension had eleven specific ratings, five subcomponent ratings, and one overall dimension rating. The interpersonal relations dimension had twelve specific part ratings, four subcomponent ratings, and one overall dimension rating. Confidence ratings were also made on each of the overall dimension ratings.

The participants also made overall ratings on each candidate. Finally, participants rated the degree of influence that eleven different possible causal influences had on their assigned scores for the candidate.

RESULTS

Before discussing the findings, a quick review of the mechanics of the study is presented. The eighteen videotapes involved in the research were divided into two groups of nine. Half of the ratees saw version one of tapes and half saw version two. The role of Chester, the problem subordinate, was always played by the same actor. Each version contained the same nine actors and actresses performing the candidate - supervisor's role in the same set of nine scripts. The difference between the two versions was that the same "candidate for supervisor" script was performed by a different actor or actress in the two versions.

Each videotape was rated on a number of different rating scales. In a counterbalanced order each candidate was rated on the problem solving rating scale and the interpersonal relations rating scale. These dimensions contained, among other measures, individual item ratings. The problem solving rating scale contained eleven items. The interpersonal relations rating scale contained twelve item ratings. For each dimension, the item ratings were summed into a dimension total score. These sums were used in the analyses.

The participants also rated their confidence in their ratings in the problem solving and interpersonal relations dimensions. Confidence ratings were made after the two dimension overall ratings. Each participant also rated a number of attribution items. The attribution items asked the rater to indicate the importance of several potential causes of the performance of the videotaped candidate. Finally, raters

made an overall assessment (OAS) rating. The OAS was made after all other dimensional ratings were completed. Raters made a final decision on the overall rating for the candidate by combining the decisions on the candidate's problem solving and interpersonal relations dimensions and whatever other information they had gathered in the testing process.

The reliability of the rating scales was assessed using Winer's (1971, p. 283-296) analysis of variance method of intraclass correlation. This method is appropriate when a number of raters use the same rating scale to rate the same set of candidates. In this case each version of nine tapes was seen by approximately 40 raters (the difference from 40 being due to missing data on some rating scales). The intraclass correlation reliability for the problem solving rating scale was .987 for version one and .983 for version two. The intraclass correlations for the interpersonal relations rating scale was .985 and .979 for versions one and two, respectively.

The hypotheses in this study all relate to differential accuracy. In addition to the analyses related to these hypotheses a number of other supplemental variables were analyzed. These variables--raw scores, rater confidence, rater attributions, and overall assessment ratings--were all added to the study to help clarify the results for the main differential accuracy findings.

The repeated measures analyses of variance summary tables for all analyses run are presented in Appendix A (starting on page 215). In the discussion that follows, "P" is the problem solving target score, "I" is the interpersonal relations target score, "E" is the expertise

of the rater, "V" is the version of the videotapes. The main analyses are performed using SAS procedure GLM for repeated measures. F statistics listed are all for unique variance accounted for (i.e., SAS "type III" sums of squares or BMDP within contrast pool (WCP) mean squares). In most cases in the analyses this provided more conservative results than even the multivariate method of computing the within portion of the design. The main effect cell comparisons were done with planned contrasts using the "profile" method, in which the first cell was compared with the second, and the second with the third, etc. All post hoc simple effect and cell contrast analyses are done using BMDP4V.

Due to the large number of significance tests conducted in the study, a minimum significance level of .01 was chosen. Note that this level does not conform to a strict Bonferroni adjustment of the alpha levels for the simple effects analyses. In the following analyses simple effects analyses were conducted for all of the raw score, differential accuracy, confidence, and OAS analyses. In all cases a maximum of two simple effects analyses were conducted on each set of analyses of variance. For this reason the alpha probability level for the simple effects analyses will be .005.

In the presentation of the results below, first the manipulation checks on the raw scores are presented. This discussion points out that the rating dimensions were successfully manipulated and the differential accuracy hypotheses may be assessed. There were also significant version differences in the raw score analyses. The main differential accuracy hypothesis concerning the interaction of the P and I target scores for both the problem solving and the interpersonal

relations rating scales was supported. The hypothesized effect for rater expertise was not supported. The third hypothesis, relating to version differences, was not supported. Several unanticipated, but hypothesis-consistent results occurred, including main effects for the P and I targets.

Following the presentation of the differential accuracy analyses, the confidence, attribution, and overall assessment findings are presented. The confidence findings modestly support the differential accuracy conclusions. Raters tended to be more confident in the more differentially accurate ratings. The attribution findings proved to be of limited value. Despite the fact that the Feldman model makes firm statements about when attributions should enter the rating process, the attribution measures used had little variance and may not have been correctly used by the respondents. Statistically controlling for the attributions did tend, however, to eliminate the main effect for the opposite dimension's target score (i.e., P main effect in the interpersonal relations analyses, and the I main effect in the problem solving analyses). The overall assessment ratings, on the other hand, tend to support the differential accuracy findings. Raters appeared to combine their ratings on the problem solving and interpersonal relations rating scales by averaging the ratings on the two scores together. The presentation of the results concludes with a summary of the findings and a discussion of the implications of these findings.

Manipulation Checks

The raw score analysis is a manipulation check. If the problem solving rating scale is effectively manipulated in the videotapes, then in an analysis of variance (ANOVA) there will be a main effect for the P target, with significant differences between each of the three P target levels. In a similar manner, if the interpersonal relations rating scale is effectively manipulated, then there should be a main effect for the I target with significant differences between the three I target levels.

Problem solving raw score

For the problem solving rating raw scores the expectation is to find a main effect for the P target with all three levels significantly different from one another. This was found. The main effect for the problem solving target score (P, $\underline{F}(2,142) = 325.3$, $\underline{p}<.0001$) was significant. The results from the repeated measures analysis of variance are summarized, along with the cell means and standard deviations, in Tables 17 and 18 in Appendix A (page 218). (For ease of presentation, all supplemental statistical tables discussed in the results section are presented in Appendix A).

Planned contrasts (using SAS GLM repeated measures "profile" contrasts) on the main effect for P indicate that all three means are significantly different from one another (target scores P3 versus P5, $\underline{F}(1,71) = 395.20$, $\underline{p}<.0001$; for target scores P5 versus P6, $\underline{F}(1,71) =$

24.27, $p \leq .0001$). The mean P scores are 2.97 for P3, 4.78 for P5, and 5.13 for P6. The manipulation for the problem solving rating dimension was successful.

Two other unanticipated effects were also significant: the I main effect ($F(2,142) = 223.2$, $p \leq .0001$) and the I by V interaction ($F(2,142) = 11.7$, $p < .0001$). An I main effect on problem solving raw score ratings is consistent with the main hypothesis of a P by I interaction in differential accuracy. That is, the candidate's skill in interpersonal relations had an effect (whether it be a confound or a statistical interaction) on the ratings of the problem solving dimension.

To investigate this unanticipated, but hypothesis-consistent result, the "profile" contrasts were checked. All three I target levels were significantly different from one another. The difference between I3 and I5 is significant ($F(1,71) = 224.54$, $p \leq .0001$). The I5, I6 difference is also significant ($F(1,71) = 865.28$, $p \leq .0001$). The mean I scores are 3.53, 4.01, and 5.24 for I3, I5, and I6, respectively. A candidate's interpersonal skill clearly had an effect on problem solving ratings. This finding is consistent with the main differential accuracy hypothesis of the study.

Also significant is the interaction between tape version (V) and the interpersonal relations target scores. This finding is more troublesome. Version differences indicate that the manipulation of the problem solving rating dimension were not the same for the two videotapes. The two versions were intended to be replications. These raw score findings suggest that the replications may not have been parallel.

Simple effects analyses were run to assess the extent of the differences between sets of videotapes. The version one and version two problem solving means for I3 ($\underline{F}(1,71) = 2.06$, ns) and I6 ($\underline{F}(1,71) = 0.14$, ns) are not significantly different from one another. The I5 problem solving means for version one and version two, however, are significantly different from one another ($\underline{F}(1,71) = 9.15$, $\underline{p}<.0035$). Thus, for the raw score problem solving ratings, raters viewing the I5 target videotapes performed by two different people performing the same lines, and differing only in nonverbals, saw significantly different problem solving skills being performed.

This potential problem is lessened by the fact that both simple main effects for I were significant (I at Version 1: $\underline{F}(2,142) = 116.70$, $\underline{p}<.0001$; I at Version 2: $\underline{F}(2,142) = 118.28$, $\underline{p}<.0001$). The profiles of the I main effects differ slightly for the two versions, but they both still occur. In both cases the I targets are clearly affecting the rating of the candidate's problem solving skills. The mean scores are shown in Table 4.

Table 4: PS Raw Scores V by I means	I3	I5	I6
Version 1	3.4	4.4	5.2
Version 2	3.6	3.8	5.2

Interpersonal relations raw scores

The expectation for the interpersonal relations rating scale is parallel to that for the problem solving rating raw scores. The expectation is to find a main effect for the I target with all three levels significantly different from one another. This was found ($F(2,142) = 290.7$, $p<.0001$). The ANOVA summary table and the means and standard deviations are presented in Tables 19 and 20 in Appendix A (page 219).

The "profile" contrasts (again using SAS GLM repeated measures "profile" contrasts) for the I target scores show significant differences between all three target values. The difference between the I3 and I5 levels was significant ($F(1,71) = 95.55$, $p<.0001$). The difference between the I5 and I6 levels was also significant ($F(1,71) = 205.77$, $p<.0001$). The mean scores for I3, I5, and I6 are 3.42, 4.40, and 5.69, respectively. The manipulation of the interpersonal relations dimension was successful.

As above, there were also several other unanticipated, but hypothesis-consistent significant effects. A main effect for P ($F(2,142) = 141.5$, $p<.0001$) occurred. The significant P target effect suggests that the candidate's problem solving skills interfered with the rater's ratings of interpersonal relations. Raters saw a difference between the interpersonal relations skills at the target scores of P3 and P5 ($F(1,71) = 201.83$, $p<.0001$). The difference between the interpersonal relations scale ratings at the P5 and P6 levels was ($F(1,71) = 0.24$, ns). The mean scores for P3, P5, and P6 are 3.55,

4.97, and 5.00, respectively. Interpersonal relations ratings were lower when ratees were scripted to be poor problem solvers than when they were above average or top quality problem solvers.

There is also an unanticipated significant P by I two-way interaction ($F(4,284) = 7.9$, $p<.0001$). The means are presented in Table 5. A visual inspection of the cell means shows that for these stimuli, the relationship between problem solving and interpersonal relations raw scores is not clear. The relationship is not a simple addition or average of the two dimension targets, nor are the two target dimensions being multiplied together by the raters. The P3 interpersonal relations ratings are all lower than the P6 ratings at each level of the I targets. The P5 targets, however, do not form a consistent pattern between the P3 and P6 target.

Table 5: IR Raw Score P by I Means	I3	I5	I6
P3	2.3	3.6	4.7
P5	4.2	4.5	6.2
P6	3.7	5.1	6.1

Two troublesome version effects were significant for the interpersonal relations ratings. A three way interaction between the P, I, and version (V) ($F(4,284) = 6.2$, $p<.0001$), as well as a version interactions with the I target ($F(2,142) = 11.2$, $p<.0001$) were significant.

For the V by I interaction, simple effects analyses showed that version differences did occur. The I5 interpersonal relations mean raw scores for version one were significantly different from the I5 mean raw scores for version two ($F(1,71) = 10.67$, $p \leq .0017$). Once again, as was true for the I by V interaction for the problem solving raw scores, the simple main effects for I within version were both significant (I at Version 1: $F(2,142) = 142.93$, $p \leq .0001$; I at Version 2: $F(2,142) = 159.18$, $p \leq .0001$). This lessens the importance of this difference. The mean scores are presented in Table 6.

Table 6: IR Raw Scores V by I means	I3	I5	I6
Version 1	3.3	4.7	5.6
Version 2	3.5	4.1	5.7

The P by I by V significant triple interaction contained a number of significant version differences. Simple effects analyses showed that the simple interaction of P by I continued to be significant within both versions (P by I at Version 1: $F(4,284) = 4.29$, $p \leq .0022$; P by I at Version 2: $F(4,284) = 9.92$, $p \leq .0001$). The reason for the three way interaction was a significant simple effect for version at P6I5 ($F(1,71) = 18.02$, $p \leq .0001$). The cell means are shown in Table 20 in Appendix A (page 219).

Summary of manipulation check analyses

Both the problem solving and interpersonal relations rating dimensions had reliable rating scales. In addition, both dimensions were successfully manipulated. Significant differences were found among all three levels of the P targets for the problem solving rating scale and among the three I target levels for the interpersonal relations dimension. Significant main effects for I in the problem solving ratings scale and P in the interpersonal relations rating scale, as well as a P by I interaction in the interpersonal relations rating were unanticipated. For the raw score ratings, the problem solving and interpersonal relations targets tended to combine in a generally additive fashion for both the problem solving and the interpersonal relations ratings. These effects do not, however, provide complete support for the main hypothesis of the study: a statistical interaction between the differential accuracy ratings for the two rating scales.

Other significant but unanticipated effects also occurred that, on the surface, are more troublesome. These are the version differences. The two versions of the same script were found to be significantly different from one another both for the problem solving and the interpersonal relations rating scales. These differences do not effect the interpretation of the successful manipulation of the rating dimensions, but if carried into the differential accuracy findings, they will put suspicion on the results. The version differences did not, however, cause major differences in the profiles of results within versions. With the successful manipulation of the rating scales, the differential

accuracy analyses are explored.

1.2 The Main Hypotheses Analyses: Differential Accuracy Measures

The accuracy measure used was Cronbach's (1955) differential accuracy. Differential accuracy (DA) is a modified difference score. To calculate the differential accuracy score (see Table 1, page 6, for the equation), the raw score is adjusted by subtracting out the variance due to the rating dimensions and subtracting out the variance due to the ratees. The true score is adjusted for these same factors. The modified true score is then subtracted from the modified raw score. In the unique case created by the experimental design, the true score differential accuracy variance was forced to zero. The differential accuracy measure thus indicates the variance in the rater's ratings due to the individual ratee on the individual rating dimension. Each of the three target scores (3, 5, and 6 points for both rating dimensions) has its differential accuracy calculated in the same fashion.

The first hypothesis for the differential accuracy rating scales is for a significant interaction for the P and I target scores. No specific P by I cell differences are predicted within the interaction. The Feldman model only indicates that a rater will base all ratings on a single observational category. That single category may differ among the raters. Some may have formed their impression based on a clear understanding of only a single rating dimension, with the other dimension confounding their judgements. Other raters may have developed their observational category based on a single quality level (e.g., average) performer on both dimensions. What these raters will have in

common is the influence that the ratee's interpersonal skill will have on the problem solving ratings. The P by I interaction did occur for both rating scales.

Since the P and I interaction is significant, a subsequent hypothesis can be made. The question is whether the two rating scales can be combined into a single "supervisory skill in problem solving meetings" skill factor. One reason why differential accuracy on one dimension would depend on the target level of both the problem solving and interpersonal relations rating scales is that individuals are confusing or combining the two scales. To assess this possibility, a post hoc contrast was run comparing differential accuracy for the P by I cells with the same target scores (i.e., P3I3, P5I5, and P6I6, the main diagonal) with the other, off-diagonal target score combinations. This contrast suggested that the candidates scoring similarly on the two target dimensions were rated with better differential accuracy on problem solving than the other (off-diagonal) candidates.

The P and I main effects would also be consistent with the main hypothesis. A main effect for the P target would indicate that raters had varying amounts of differential accuracy for the three P targets on the problem solving rating scale. The main effect for the I targets would support the main hypothesis of the P by I interaction. The I main effect would indicate that the raters differ in differential accuracy on the problem solving rating scale depending on the I target levels. An I main effect occurred for the problem solving dimension. For interpersonal relations both the P and the I main effects occurred.

A second hypothesis is for an interaction between rater expertise and the P and I targets. This would indicate that a rater's differential accuracy changes depending on their expertise as a supervisor. Experts in general are not more accurate over all rating dimensions that are nonexperts. Instead, it is hypothesized that a rater's expertise will be shown in accurately rating only one dimension. This hypothesis was not supported.

Finally, there may be some unwanted version differences in the ratings of the problem solving and interpersonal relations rating scales. Version differences will indicate that there are serious problems with the stimulus materials due to the nonverbal behavior differences of the actors and actresses performing in the videotapes. Main effect version differences or version interactions will indicate that raters see different amounts of skill in problem solving depending on the particular nonverbals displayed by the actor or actress in the particular videotape. This third "hypothesis" was not supported. There were no significant videotape version differences to cloud the main hypothesis differential accuracy findings.

Problem solving ratings differential accuracy

The first hypothesis for the problem solving rating scales is for a significant interaction of the P and I target scores. The P and I main effects would also be consistent with the main hypothesis.

A second hypothesis is for an interaction between rater expertise and the P and I targets. This would indicate that a rater's differential accuracy changes depending on their expertise as a supervisor.

Any version differences, however, will signal problems in the research. Main effect version differences or version interactions will indicate that raters see different amounts of skill in problem solving depending on the particular nonverbals displayed by the actor or actress in the particular videotape.

The analyses show that the problem solving rating scale differential accuracy measure contained the hypothesized P by I interaction ($F(4,284) = 5.4$, $p \leq .0004$). In addition, a main effect for the I target score occurred ($F(2,142) = 12.2$, $p \leq .0001$). The ANOVA Summary table and the cell means are presented in Tables 21 and 22 in Appendix A (page 220).

The significant interaction was analyzed by performing a simple effects analysis. These analyses showed that a simple main effect for I occurred at the P5 ($F(2,142) = 5.98$, $p \leq .0032$) and the P6 level ($F(2,142) = 15.0$, $p \leq .0001$), but not at the P3 level ($F(2,142) = 0.35$, ns). There was one significant simple main effect for the P targets (at I3): ($F(2,142) = 8.84$, $p \leq .0002$). The I6 simple main effect within P approached the .005 significance level (recall that .005 was chosen as the significance level for all simple effects analyses), but was not significant ($F(2,142) = 3.87$, $p \leq .0231$, ns). The P simple main effects at I5 ($F(2,142) = 0.44$, ns) was not significant. The cell means are presented in Table 7. Note that low differential accuracy scores indicate better accuracy.

Table 7: PS Differential Accuracy P by I Means	I3	I5	I6
P3	.54	.63	.55
P5	.69	.64	.47
P6	.86	.58	.42

The mean scores suggest that the P6I3 and P5I3 cells are driving the interaction. Both had high scores, indicating poor differential accuracy. The best differentially accurate problem solving ratings were at the P6I6 and P5I6 levels. Within the P5 and P6 target levels, the problem solving ratings get more differentially accurate as the target interpersonal relations skills get better. There was no difference among the problem solving ratings at P3.

There was also an interesting difference in the accuracy of problem solving ratings within the I3 and I6 targets. When raters had poor interpersonal relations skills, raters got more accurate as the rated problem solving skills got worse. This effect was reversed for the I6 simple main effect. Candidates who were skilled in interpersonal relations were rated more accurately on problem solving when their problem solving skills got better.

For the I target main effect, the mean scores for I3, I5, and I6 were .70, .62, and .48, respectively. As candidates got more skilled in interpersonal relations, the problem solving ratings got more differentially accurate. This clearly supports the main hypothesis of an

interaction between the two rating dimensions. Problem solving ratings were clearly affected by the interpersonal relation skills of the ratee.

The post hoc hypothesis related to the first hypothesis is that the problem solving and interpersonal relations dimensions should be combined. To assess the possibility a post hoc comparison was run comparing differential accuracy for the P by I cells with the same target scores (i.e., P3I3, P5I5, and P6I6, the main diagonal) with the other, off diagonal target score combinations. For the problem solving rating dimension the main diagonal was significantly more differentially accurate than the off diagonal ($\underline{F}(1,67) = 19.01$, $\underline{p} \leq .0000$). Thus, the tentative conclusion is supported that some combination of the interpersonal relations and problem solving scales can be considered. Differential accuracy is better when the two rating dimensions are at consistent levels of performance than when they are at inconsistent levels of performance.

The second hypothesis concerning expertise differences was not supported. There was no support for an expertise effect, either as a main effect ($\underline{F}(1,71) = 0.03$, ns) or in interaction with the P and I targets ($\underline{F}(4,284) = 2.21$, ns). Expert raters were not more differentially accurate than nonexperts.

More importantly, there were no (unwanted) version differences. The main effect for version was not significant ($\underline{F}(1,71) = 0.66$, ns). In addition, all of the version interactions were at similar $\underline{F}$ and probability levels. See Table 21 and 22 in Appendix A (page 220) for the ANOVA summary table and cell means.

Interpersonal relations ratings differential accuracy

The hypotheses for the interpersonal relations rating scales are parallel to those for the problem solving rating scale. First, a significant interaction for the P and I target scores is expected. As above, no specific hypotheses are developed to predict which P and I target cells will differ in the interaction. Second, the respective P and I main effects and an expertise interaction with them will also be consistent with expectations. Third, version effects will, as previously mentioned, indicate problems in the videotapes due to ratee nonverbals affecting their portrayal of the true score targets.

The data show that the interpersonal relations rating scale differential accuracy measure contained two significant main effects, for the P ($F(2,142) = 9.2$, $p \leq .0002$) and I ($F(2,142) = 5.3$, $p \leq .0062$) targets, and a significant interaction between them (P by I, $F(4,284) = 19.4$, $p \leq .0001$). Once again the main hypothesis of a P by I interaction is supported. The ANOVA summary table and cell means and standard deviations are presented in Tables 23 and 24 in Appendix A (page 221).

Simple effects analyses on the P by I interaction show that two of the three simple main effects for I were significant (I at P3, $F(2,142) = 7.89$, $p \leq .0006$; I at P6, $F(2,142) = 34.40$, $p \leq .0000$), and the other approached significance (I at P5, $F(2,142) = 4.27$, $p \leq .0158$, ns). All three of the simple main effects for P were significant (P at I3, $F(2,142) = 26.65$, $p \leq .0000$; P at I5, $F(2,142) = 5.71$, $p \leq .0041$; P at I6, $F(2,142) = 15.57$, $p \leq .0000$). The cell means are presented in Table 8, below. Again, low scores indicate high differential accuracy.

Table 8: IR DA Scores P by I Means	I3	I5	I6
P3	.48	.88	.89
P5	.63	.59	.38
P6	1.20	.52	.47

The cell means show a strong similarity to the findings for the problem solving differential accuracy ratings. The results for the interpersonal relations dimension are even stronger. For candidates with poor (I3) interpersonal relations skills, their ratings got more accurate when their target problem solving skills got worse. This effect was reversed for the I5 and I6 targets. Interpersonal relations dimension ratings got more accurate as problem solving skills rose.

The same basic pattern emerged within the P target simple main effects. For those candidates with poor problem solving skills, their interpersonal relations ratings got more differentially accurate when their interpersonal relations target skills were weak. For the I5 and I6 simple main effects, interpersonal relations ratings got more differentially accurate when the ratee's problem solving skills also got better.

The most differentially accurate interpersonal relations ratings were made for the I6 targets at P5 and P6, the I5 targets at P5 and P6, and the I3 targets at P3 and P5. By far the least differentially accurate interpersonal relations ratings were made at the P6I3 targets, followed by the P3I5 and P3I6 ratings.

To assess the possibility that the two rating dimensions should be combined, a post hoc comparison was run comparing differential accuracy for the P by I cells on the main diagonal (i.e., P3I3, P5I5, and P6I6) with the off diagonal target score combinations. The interpersonal relations contrast comparing the main diagonal and the off diagonal approached significance at the .01 level, but did not reach it ($\underline{F}(1,71) = 4.85$, $\underline{p} < .0309$). Thus, the tentative conclusion of a combined interpersonal relations and problem solving scale does not receive support from the interpersonal relations findings. More research must be done to assess how the two rating scales interact.

Two other unanticipated effects were significant. Both are consistent with the first hypothesis. For the main effect for the I target the I3, I5, and I6 means are .76, .66, and .58, respectively. Raters were more differentially accurate as the interpersonal skills of the ratees rose.

The main effect for the P target showed an interesting change from the previous findings. Instead of interpersonal relation scores getting more differentially accurate as problem solving targets rose, accuracy was highest for the middle level of problem solving. The P5 target was more accurate than the either the P3 or the P6 target. The P3, P5, and P6 means are .75, .53, and .72. The candidate's problem solving target score clearly had an effect on the accuracy of the interpersonal relations rating, but not in a linear fashion.

The second hypothesis of an effect for rater expertise within the P by I interaction was not supported. The E main effect ($\underline{F}(1,71) =$

2.05, ns) and the P by I by E interaction ($\underline{F}(4,284) = 0.90$, ns) were not significant. Expert raters were not more differentially accurate overall or at some levels of the P by I target combinations.

Also as in the problem solving ratings there were no version differences. The version main effect ($\underline{F}(1,71) = 2.59$, ns) was not significant. In addition, none of the version interactions were significant (see Appendix A, page 219, for the ANOVA summary table).

1.2.3 Summary of the differential accuracy analyses

As hypothesized, a significant P by I target interaction occurred for both the problem solving and interpersonal relations rating scales. For the problem solving rating dimension, the most accurate ratings were given for the P6I6, P5I6, and P3I3 target score combinations. The least accurate problems solving rating was given for the P6I3 target.

A significant, but unanticipated I main effect occurred for the problem solving rating. This was consistent with the overall main hypothesis of a P by I interaction. The non-rated dimension manipulation proved to have a significant effect on differential accuracy. Ratees with high quality interpersonal skills were rated with significantly more differential accuracy on problem solving than were those with average or low skill in interpersonal relations.

Just as important as the support for the main hypothesis is the absence of any version differences in the differential accuracy findings for the problem solving rating scale. The version difference noted in the manipulation checks did not hold in the main analyses.

One other effect was hypothesized. It was anticipated that the expert raters would be significantly different from the nonexpert raters on some of the P by I interaction target cells. In fact, there were no effects for the expertise variable, either alone or in combination with the other variables.

The interpersonal relations ratings were similar in differential accuracy to the problem solving rating scales. The hypothesized interaction between the P and I targets occurred. Again, the most differentially accurate ratings were found for the P6I6, P5I6, and P3I3 targets. The least differentially accurate rating was made for the P6I3 target. In general, differential accuracy was best when both dimensions were either very low or very high and worse if the two dimensions were inconsistent, especially high problem solving combined with low interpersonal relations.

The unanticipated main effects for P and I are both supportive of the main hypothesis. The trend is clear for the I main effect. Overall, raters were more accurate at rating interpersonal relations when the ratee had higher interpersonal skills. The P main effect showed the interfering effect of the P target on the ratings of interpersonal relations. Raters were more accurate when rating the interpersonal relations of a ratee with medium levels of problem solving skills than with high or low quality levels.

The interpersonal relations ratings also contained no significant effects in support of the hypothesis that experts would be more accurate than nonexperts overall or on some of the P by I target combinations.

Finally, there were no version differences within the interpersonal relations ratings. This supports the assertion that the ratee's nonverbal behaviors were not interfering with the differential accuracy of rater's assessment of ratee interpersonal relations skills.

Supplemental Variables: Rater Confidence

The remainder of the analyses serve to clarify and extend the findings for rater differential accuracy. The confidence, attribution, follow-up questionnaire, and overall assessment analyses are all post hoc. Caution should be used in interpreting significant results. The variables are studied only to suggest interpretations of the differential accuracy findings.

Confidence ratings are important because they are a subjective measure of how sure the rater is that he or she made an accurate rating. No specific hypotheses were made about the outcome of the confidence rating analyses. However, whether raters were confident in the ratings for which they were most accurate is of considerable interest. For example, in other types of judgements it has found that experts are no more accurate than inexperienced judges, but are more confident about their accuracy (DePaulo and Pheiffer, 1986). The present findings modestly support the main hypothesis for differential accuracy. Raters were confident in the ratings that were more differentially accurate.

Raters were generally most differentially accurate about the P5I6, P6I6, and P3I3 targets and least accurate about the P6I3 target. Raters were most confident about the same cells. Raters were least confi-

dent about the P6I3, P3I6, and P5I3 targets. This was true both for the problem solving and the interpersonal relations scales. Correlations between rater confidence and differential accuracy showed that these supportive effects were modest at best.

Problem solving rater confidence

If there is a relationship between rater differential accuracy and rater confidence, then the correlation between accuracy and confidence should be significant. Correlations were run between problem solving rater confidence and differential accuracy within each of the nine P by I target score combinations. None of the correlations were significant. The correlations ranged from -.24 for P6I6 to .00 for P3I6. (Note that high differential accuracy scores indicate low problem solving and a hypothesis supporting correlation will be negative.)

Although no specific hypotheses were formulated for the rater confidence analyses, several results would tend to support the main finding for the differential accuracy results. A significant interaction for the P and I targets would indicate that rater confidence changes for specific target score combinations. If rater confidence were higher for the cells with higher differential accuracy (i.e., lower DA score values for P6I6, P5I6, and P3I3) and lower for cells with lower differential accuracy (i.e., higher DA scale values for P6I3 and P3I6), then confidence would support rater differential accuracy.

In an ANOVA run on the confidence data, the problem solving rating scale confidence ratings contained two significant effects: the interaction between P and I ($\underline{F}(4,286) = 14.0$, $\underline{p} \leq .0001$), as well as the three

way interaction between P, I, and V ($\underline{F}(4,286) = 3.5$, $\underline{p} \leq .0078$). The ANOVA summary table and the table of means and standard deviations are presented in Tables 25 and 26 in Appendix A (page 222).

In the P and I interaction, simple effects analyses showed that all of the simple main effects for I were significant (I at P3, $\underline{F}(2,134) = 13.25$, $\underline{p} \leq .0000$; I at P5, $\underline{F}(2,134) = 9.54$, $\underline{p} \leq .0001$; I at P6, $\underline{F}(2,134) = 5.74$, $\underline{p} \leq .0040$). The cell means, shown in Table 9, below, indicate that confidence in problem solving ratings for candidates with poor problem solving skills fell as they got more skilled in interpersonal relations. The trend was reversed for the moderate (P5) and highly skilled (P6) problem solvers. As they also got more skilled in interpersonal relations, the raters' confidence in the problem solving ratings rose.

Table 9: PS Confidence P by I means	I3	I5	I6
P3	5.9	5.2	5.2
P5	4.9	5.2	5.6
P6	5.1	5.3	5.5

One of the three simple main effects for the P targets was significant and one approached significance. The P simple main effect at I3 was significant (P at I3, $\underline{F}(2,134) = 25.03$, $\underline{p} \leq .0000$). The P simple main effect at I6 approached significance (P at I6, $\underline{F}(2,134) = 5.34$, $\underline{p} \leq .0056$). The simple main effect for P at I5 was not significant (P at I5, $\underline{F}(2,134) = 0.29$, ns). Results were parallel to those for the

I simple main effects. When rating the problem solving skills of candidates with poor interpersonal relations skills (I3), raters were more confident in their ratings of poor problem solvers (P3) than of the better problem solvers. This confidence effect was reversed when rating the problem solving skills of candidates with high skills in interpersonal relations (I6). Raters were more confident in their problem solving ratings of medium and high skill problem solvers than of ratees with poor problem solving skills.

Raters were least confident about the ratings made for the targets for which they were least differentially accurate (e.g., P3I6 and P6I3) and showed more confidence in the ratings where they were more differentially accurate (e.g., P3I3, P5I6, and P6I6).

One other effect was significant for problem solving rater confidence. This unanticipated effect is for version differences, just as in the (raw score) manipulation checks. Simple effects analyses showed that the interactions of P and I was significant both within version one ($\underline{F}(1,67) = 4.02$, $\underline{p} \leq .0035$) and within version two ($\underline{F}(1,67) = 13.40$, $\underline{p} \leq .0000$). These findings would not raise major doubts about the confidence rating findings. The simple effects within the P by I targets showed that version differences were found within the P3I6 cell ($\underline{F}(1,67) = 10.13$, $\underline{p} \leq .0022$; means of 4.4 for version one and 5.1 for version two). The version differences are pointed out here only to give the reader an understanding of what was driving the significant three way interaction.

Interpersonal relations rater confidence

No specific hypotheses were formulated for the interpersonal relations rater confidence analyses. As above, a P by I interaction would tend to support the main finding for the differential accuracy results provided that rater confidence were higher for the cells with higher differential accuracy (i.e., P6I6, P5I6, and P3I3) and lower for cells with lower differential accuracy (i.e., P6I3).

Correlations were run between interpersonal relations rater confidence and differential accuracy within each of the nine P by I target score combinations. Again, none of the correlations were significant at the .01 level. The correlations ranged from -.26 for P5I6 to .13 for P6I3. (Note that negative correlations support the hypothesis.)

The confidence ratings for the interpersonal relations rating scale contain two significant effects, the I main effect ($\underline{F}(2,138) = 11.21$, $\underline{p}<.0001$) and the interaction of I with the P target ($\underline{F}(4,276) = 10.85$, $\underline{p}<.0001$). Both are supportive of the main differential accuracy hypothesis. The ANOVA summary and descriptive statistics tables are presented in Tables 27 and 28 in Appendix A (page 223). Rater confidence in their interpersonal relations ratings was consistent with differential accuracy.

The main effect for I was due to the difference between the I5 and I6 confidence ratings ($\underline{F}(1,69) = 23.92$, $\underline{p}<.0001$). The I3 and I5 confidence ratings were not different from one another. The mean confidence ratings for I3, I5, and I6 are 5.2, 5.1, and 5.5, respectively. (As above, these main effect contrasts were carried out with the SAS GLM

"profile" contrasts procedure.) Raters were more confident when rating the interpersonal relations skills of candidates with high (I6) skills than with medium (I5) or low (I3) skills.

Of more importance is the P by I interaction. Simple effects analyses showed that all three simple main effects for the I targets were significant (I at P3, $F(2,138) = 10.71$, $p \leq .0000$; I at P5, $F(2,138) = 13.66$, $p \leq .0000$; I at P6, $F(2,138) = 8.42$, $p \leq .0004$). For the P target simple main effects only those at I3 and I6 were significant (P at I3, $F(2,138) = 7.72$, $p \leq .0007$; P at I5, $F(2,138) = 3.90$, ns; P at I6, $F(2,138) = 12.94$, $p \leq .0000$). The means are presented below, in Table 10.

Table 10: IR Confidence P by I means	I3	I5	I6
P3	5.6	4.8	5.1
P5	4.9	5.2	5.8
P6	5.2	5.2	5.7

The results directly parallel the findings for the problem solving confidence ratings. Raters were more confident rating interpersonal relations for candidates with weak interpersonal skills and weak problem solving skills than for those with weak interpersonal relations skills and strong problem solving skills. On the other hand, when candidates had strong interpersonal relations skills, raters were more confident rating candidates who also had the stronger problem solving skills.

Looking at the means within the P targets, raters were more confident in their interpersonal relations ratings when poor skills in both targets were matched or when moderate or strong skills in each target dimension were matched. Once again rater confidence tended to follow rater differential accuracy.

Supplemental Variables: Rater Attributions

The attribution measures were analyzed to determine whether the rater's personal expectations about the causes of a ratee's actions may have biased the rater's ratings. Again, no specific hypotheses are made. In part due to the absence of specific hypotheses, in part due to the large number of analyses, and in part due to the absence of any solid effects for the attribution measures, these analyses should be viewed with extreme caution. At best they show the absence of an important mediating effect by rater attributions on rater problem solving differential accuracy and a modest effect for the interpersonal relations dimension.

When used as a covariate the four attribution measures proved to have no significant effects, either for the problem solving rating scale or the interpersonal relations scale. For interpersonal relations, the attribution measures tended to eliminate the main effect for the P target. The P main effect in the interpersonal relations analysis was supportive of, but not essential to the main hypothesis of an interaction between the P and I targets. The attribution analyses

also show that ability was more important than effort, effort more than task difficulty, and task difficulty more than luck in the overall attributions made by the raters.

When analyzed as dependent measures the four attribution measures proved to have quite consistent mean scores across the nine P by I target combinations. Ability was the most important attribution, followed by effort, task difficulty, and luck. This ordering of importance ratings held for each of the nine P by I target combinations. Differences among the mean scores were found only for the ability and effort attributions.

What follows is first, a discussion of the calculation of the attribution measures, second, the attribution covariate analyses, and finally the attribution measures as dependent variables in the full design.

The attribution items collected with each set of target ratings were divided into four separate attribution composite measures. All four attributions measure the perceived importance of the factor in the rater's ratings on the particular dimension. The skill in problem solving and skill in interpersonal relations importance rating attributions were averaged into an ability (internal-stable) attribution composite. The hard work and (reverse scored) mood variables were averaged into an effort (internal-unstable) attribution composite measure. The luck, poor acting, inappropriate nonverbals, and unusual assistance by Chester measures were averaged into a luck (external-

unstable) attribution composite. The typical assistance by Chester and the item difficulty measures were averaged into a task difficulty (external-stable) composite measure.

These four composites were analyzed in two ways: as covariates and as dependent variables. In the covariate analyses, each of the four attribution measures were analyzed as separate composites in the original differential accuracy designs. Results are presented to show the changes in the original differential accuracy scores after the effects of the covariates are removed from the linear model.

The second style of analysis was as dependent measures. Each attribution measure was run as the dependent measure in the original experimental design. The intention was to find out if the attribution ratings differed depending on the P by I target levels. The rationale for this set of analyses is similar to those for the rater confidence ratings. Differences in the importance of the attributions for the nine P by I target combinations that match the differential accuracy findings would show the impact of the attributions on differential accuracy. If external attribution importance (to the situation or the item, rather than to ratee ability or effort) was high for ratings with poor differential accuracy, then the attributions may have been interfering with rater accuracy.

Attribution ratings as covariates

The four attribution composite measures were used as covariates in the differential accuracy repeated measures analyses of variance for

the problem solving and interpersonal relations measures. Each target rating had its own set of covariates. Respondents rated the attribution items once for each videotape.

No specific hypotheses were formulated on these variables. The data are presented (see Tables 29-46 in Appendix A, beginning on page 224) to provide light on the reasons for the significant differences in differential accuracy between the various target levels. The raw score averages as well as the adjusted averages of the cells in the design are presented so that comparisons can be made on the relative strength of the covariates. The number of participants included in each analysis differed slightly from those reported above due to the changes in missing values. For these reasons, the results presented below should be interpreted with more caution than is ordinarily placed on the already difficult to interpret analyses of covariance.

Problem solving differential accuracy. The hypothesis to be assessed in these analyses is whether the specific attribution measure covaries with the effects found in the main differential accuracy analyses. Any changes would indicate that the attribution is having an effect. No significant effects for the four attribution measures were found for the problem solving attribution analyses of covariance.

The original analysis on the problem solving differential accuracy data discussed above contained a significant main effect for the I target and a significant interaction between the P and I targets. The means and standard deviations (Tables 24 page 221) as well as the ANOVA (Table 23, page 221) summary table are presented in Appendix A.

The repeated measures analysis of covariance (ANCOVA) with the ability attribution covariate contained the same significant effects (i.e., the I main effect, $F(2,95) = 13.96$, $p \leq .005$) and the P by I interaction ($F(4,191) = 5.62$, $p \leq .005$) as were in the original repeated measures analysis of variance (ANOVA). None of the ability covariates were significant. (See Tables 31 and 32, pages 225-226.)

The repeated measures analysis of covariance (ANCOVA) with the effort attribution covariate also contained the same significant effects (i.e., the I main effect ($F(2,95) = 16.52$, $p \leq .0001$) and the P by I interaction ($F(4,267) = 5.37$, $p \leq .005$) as were in the original repeated measures analysis of variance (ANOVA). None of the covariates were significant. (See Tables 32 and 33, page 226.)

The repeated measures analysis of covariance (ANCOVA) with the luck attribution covariate once again contained the same significant effects as were in the effort ANCOVA. The I main effect ($F(2,95) = 11.24$, $p \leq .0000$) and the P by I interaction ($F(4,191) = 5.26$, $p \leq .0000$) were significant, as were found in the original repeated measures analysis of variance (ANOVA). None of the covariates were significant. (See Tables 34 and 35, page 227.)

The repeated measures analysis of covariance (ANCOVA) with the task difficulty attribution covariate once again contained the same significant and nearly significant effects as were in the other ANCOVAs and the original ANOVAs. The I main effect ($F(2,95) = 16.11$, $p \leq .0000$) and the P by I interaction ($F(4,191) = 5.46$, $p \leq .0008$) were significant. None of the covariates were significant. (See Tables 36 and 37, page 228.)

Interpersonal relations differential accuracy. The hypothesis to be assessed in the interpersonal relations analyses is also whether the specific attribution measure causes the effects found in the main differential accuracy analyses to change. As above, any changes would indicate the importance of the attribution effect. In this case, some changes in the results did occur. The P main effect found in the differential accuracy analyses became nonsignificant in the ANCOVA.

The original analysis of variance (ANOVA, without covariates for the interpersonal relations differential accuracy (see Tables 23 and 24 in Appendix A, page 221) contained three significant effects. Main effects for the P and I targets were significant as well as their interaction. The ANCOVA with the ability attribution covariate showed a significant effect only for the P by I interaction ($F(4,191) = 13.83$, $p \leq .0000$). Both the I and the P main effects were not significant. In addition, the three way interaction between P, I, and V approached significance. None of the covariates were significant. (See Table 38 (page 229) for the interpersonal relations ANCOVA raw scores and Tables 39 and 40 in Appendix A for the ability adjusted results, pages 229-230).

When the effort attribution is added as a covariate (ANCOVA) to the interpersonal relations rating scale model, the results change. Only the main effect for I ($F(2,95) = 4.87$, $p \leq .0097$) and the P by I ($F(4,191) = 14.30$, $p \leq .0000$) interaction remain significant. The P main effect drops out and the P by I by V three way interaction approaches significance. None of the covariate effects are significant. (See Tables 41 and 42 in Appendix A page, 231.)

In the ANCOVA with the luck attribution as a covariate, once again the main effects for I ($F(2,95) = 5.48$, $p \leq .0056$) as well as the P by I interaction ($F(4,191) = 13.97$, $p \leq .0000$) were significant. The P main effect was not significant. None of the covariate effects were significant. (See Tables 43 and 44 in Appendix A, page 232).

For the task difficulty analysis of covariance the results were identical to those above. The main effect for I ($F(2,95) = 5.26$, $p \leq .0068$) and the interaction of P and I ($F(4,191) = 13.82$, $p \leq .0000$) were once again significant. As above, the P main effect was no longer significant. The P, I, and V three way interaction approached significance and none of the covariates were significant. (See Tables 45 and 46 in Appendix A, page 233).

The exact interpretation of these effects would require more research, but the main hypothesis of an interaction of the P and I targets appears to be supported. In general, the P main effect is removed by the four covariates. The P main effect in the differential accuracy ANOVA (without covariates) supports the main hypothesis of an interacting effect of the P and I rating dimensions. The P main effect indicated a clear interference for the problem solving dimension on the differential accuracy of the interpersonal relations ratings. This effect is removed by the attribution covariates. An attribution not related to the interpersonal relations true score is shown to be related to rater differential accuracy. The common variance between the attribution covariates and the removed significant effect is the P target variance. The question that remains to be answered is why the

removal of attribution variance would tend to eliminate the P main effect from one dimension, but not the other. Further research must be done to address this issue.

Another interesting comparison is the change in differential accuracy for the two rating dimensions after the attribution variance is removed. By comparing Tables 29 and 30 in Appendix A (page 224), the differences in problem solving differential accuracy can be seen (recall that higher scores indicate poorer accuracy). By removing the ability attribution bias, accuracy tends to improve for the P3I5, P3I6, P5I6, and P6I6 target score combinations. Differential accuracy is worse for the P3I3, P5I3, P6I3, and P6I5 target score combinations. Differential accuracy appears to get worse for problem solving ratings combined with poor interpersonal relations skills and somewhat better when combined with above average interpersonal relations skills. Again, further research in this area is recommended.

Ability, effort, luck, and task difficulty as dependent variables

Further analyses were run to assess whether the four attributions had any differential effect in the full model in the study. Each of the four attribution measures were used as dependent variables in the model crossing the P and I targets with expertise and version. Significant effects were found only for the ability and effort analyses. Note that all of these analyses are post hoc. Care should be taken not to over-interpret these analyses due to the high "fishing rate."

Ability (an internal-stable attribution; see Table 47 in Appendix A (page 234) was given the highest mean ratings of the four attributions. Two significant main effects were found. The P ($F(2,134) = 6.48$, $p \leq .0021$) main effect was due to the P3 rating being significantly lower than the P5 and the P6 ratings. The mean scores are 4.56, 4.88, and 5.01 for P3, P5, and P6 respectively. For the I main effect ($F(2,134) = 12.06$, $p \leq .0001$) the I5 target was significantly lower than the I6, but not the I3 target. Mean scores for I3, I5, and I6 are 4.72, 4.54, and 5.20. Ratees with higher P targets and higher I targets were attributed with more ability than raters with lower targets. Mean scores for the four attribution ratings are presented in Table 48 Appendix A (page 234).

Effort (internal-unstable) attributions were assigned the second highest importance ratings for all P by I target combinations, behind the ability attributions. The effort attribution importance ratings also showed significant P and I main effects. For the P main effect ($F(2,136) = 9.12$, $p \leq .0002$), once again the P3 targets were significantly lower than either the P5 or the P6 targets. Means are 4.23, 4.50, and 4.50 for P3, P5, and P6 respectively. For the I main effect ($F(2,136) = 8.14$, $p \leq .0005$) the I6 target was rated as significantly more important than either the I3 or the I5 ratings. Means are for I3, I5, and I6 are 4.28, 4.34, and 4.61. Again, as was true for the ability analyses above, ratees with generally higher target scores were attributed as having exhibited more effort than ratees with lower target scores. (See Table 49 in Appendix A, page 235).

For the task difficulty (external-stable) attribution, the third largest mean rating for all P by I target combinations, there were no significant effects (see Table 50 in Appendix A, page 236). Finally for the luck attributions there were also no significant differences. Luck attributions were assigned the lowest importance of any of the four attributions (see Table 51 in Appendix A, page 237).

Summary of the attribution analyses

Rater attributions were measured both as covariates and as dependent measures. Both of these post hoc analyses tended to mildly support the main finding of an interaction between the P and I targets for differential accuracy. In the covariate analyses no significant effects were found for the four attribution measures for the problem solving attributions. For the interpersonal relations attributions, however, the covariates tended to eliminate the effect of the P target. The P main effect supported the main hypothesis by interfering with the interpersonal relations dimension differential accuracy. An attribution was rated as important and this attribution tended to have variance in common with an effect that supported the main hypothesis. The ability covariate also eliminated the significant I main effect in the interpersonal relations analysis of covariance.

In the analyses of the attributions as dependent measures, ability and effort attributions were assigned the first and second highest importance ratings for all P by I target combinations. In addition, ratees with generally higher target scores were attributed as having exhibited more ability and effort than ratees with lower target scores.

This is consistent with, but does not support the major hypothesis that the problem solving and interpersonal relations dimensions interact. The findings are, however, consistent with attribution theory and research. The ratee's performance was attributed more often to internal (ability and effort) than to external (task difficulty and luck) circumstances. This is the "fundamental attribution error" (Jones and Nisbett, 1971).

Supplemental Variables: Overall Assessment (OAS) Ratings

An important question arises with the discovery that raters tend to be most accurate when the two performance dimensions are being performed at similar qualities of performance and are least accurate when the performance qualities differ widely. The question concerns how the raters will view each candidate as a whole. The overall assessment ratings (OAS) provide a means of assessing this question.

The OAS is a raw score (not a differential accuracy measure) made by each rater for each candidate after rating both of the dimensions separately. Ratings were made based on a subjective combination of the problem solving and interpersonal relations dimensions along with any other information that the rater thought to be relevant to an overall score. Candidates were free to subjectively weight the two scales, rather than average their two sets of ratings. The ANOVA summary table for OAS is presented in Table 52 in Appendix A (page 238), along with a table (Table 53, page 239) of means and standard deviations.

OAS, attribution correlations

One important measure of the effect of the attribution ratings on the raters' overall assessment of the candidates is the correlations among these measures. That correlation matrix is presented below (see Table 11). Attribution measures were rated as the importance of the various attributions in the rater's ratings of the candidate. High scores indicate that the attribution was rated as important. Ability, effort, and task difficulty all correlated significantly with the overall assessment (OAS) rating. The correlations are not high. The luck attribution, although not significantly correlated with the OAS, did correlate r=.65 with the task difficulty attribution. None of this luck attribution covariance seems to be related to OAS. The other three attributions: ability (internal-stable) and effort (internal-unstable) as well as task difficulty (external-stable) impact on the OAS.

Table 11: Correlations of Overall Rating (OAS) and Attribution Ratings

	OAS	Ability	Effort	Luck
Ability	.23*			
Effort	.22*	.19*		
Luck	-.01	.14*	-.09	
Task Difficulty	.11*	.22*	.08	.65*

Note: * is significant at the .01 alpha level

If an attribution correlates with OAS at the global level, the next question is its degree of influence on the individual target score combinations. The correlation matrix below (Table 12) presents these correlations. Note that the attribution measures are importance ratings. Using the conservative .01 alpha significance level, very few of the correlations reach statistical significance. Luck and task difficulty are correlated with OAS for the P3I3 target, but with no other targets. Ability attributions correlate significantly with OAS for the P5I6 targets. Effort does not correlate significantly with OAS for any of the nine targets. Note that high attribution ratings indicate more importance assigned to them.

Table 12: Within Cell Correlations of OAS and Attributions									
OAS Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
Ab	.08	.08	.26	.16	.28	.38*	-.13	.28	.17
Ef	.07	.02	.15	.07	.20	.19	.15	.24	.14
Lu	.31*	-.01	-.11	.02	.11	-.07	.20	-.00	-.14
TD	.41*	.21	.06	.06	.17	.12	.25	.05	.05
Note: * is Significant at the .01 alpha level									

Overall, the OAS-attribution correlations were rather varied across the nine cells for each attribution. For example, Table 11 show a significant correlation for OAS and perceived ability, but the breakdown in Table 12 shows only one significant correlation. Despite the few number of significant correlations, each of them is in a consistent direction. For the P3I3 candidate, a person performing poorly on both rating dimensions, a significant relationship occurs between the overall score assigned to the candidate and the amount of luck or task difficulty attributed as causes for the action. For the P5I6 target, a high scoring candidate, overall scores are seen as relating to the candidate's ability, an internal-stable attribution.

OAS raw scores

The main question to be answered by the analysis of the OAS data is where the various P by I target score combinations were rated. The differential accuracy analyses showed that raters did not assign the

same rank order to scores for the nine P and I target score combinations for the problem solving and interpersonal relations ratings. For the OAS ratings the P by I interaction was not significant ($F(2,272) = 2.4$, ns). There were, however, main effects for the P and the I targets, as well as an interaction between the P, I, and V variables. Finally, there was a significant interaction between the I and V variables. In general, raters were able to make distinctions in overall scores between the bottom level (P3) problem solving target scores and the other (P5 and P6) problem solving target scores. On the other hand, raters made distinctions between all three levels of the interpersonal relations targets. (The means and standard deviations and the ANOVA summary table for the ability corrected OAS raw scores is presented in Tables 54 and 55 in Appendix A, pages 239-240).

The P main effect ($F(2,136) = 197.22$, $p<.0000$) is due to the difference between the P3 and the P5 ratings ($F(1,68) = 245.26$, $p<.0001$). There is no difference between the P5 and P6 ratings ($F(1,68) = 1.26$, ns). The mean ratings for P3, P5, and P6 are 3.18, 4.97, and 5.07, respectively. Raters appear to be more accurate at distinguishing among poor and better performing candidates. Raters do not appear to be able to clearly distinguish between the skill levels of average and above average performers.

The main effect for the I ratings ($F(2,136) = 204.94$, $p<.0000$) is due to differences between each of the three I targets. The I3 and I5 means were significantly different (means of 3.47 and 4.26, $F(1,68) = 69.45$, $p<.0001$). The I5 and I6 means were also significantly different from one another (the I6 mean is 5.49, $F(1,68) = 150.33$, $p<.0001$).

These two main effects both indicate that when the candidates' performances are viewed as a whole, their ratings increased as the target score increased, except that no overall differences were seen between the P5 and P6 targets. This result suggests that the raters did use an averaging strategy for combining the ratings from the problem solving and interpersonal relations rating scales, at least when overall raw score were analyzed.

The P by I by V three way interaction ($F(4,272) = 4.79$, $p<.0000$) points out the difference between the two version raw OAS scores. The I by V interaction ($F(2,136) = 13.40$, $p<.0000$) also points out this potential problem. The cell means for version one are 3.3, 4.6, and 5.4 for I3, I5, and I6 respectively. For version two the I target cell means are 3.6, 3.9, and 5.6. These findings are of limited importance, however, in light of the previous results on version differences.

OAS raw scores with attribution covariates

Each of the four attribution measures are used as covariates with the overall assessment rating (OAS). No specific hypotheses are evaluated. Instead, the measures are assessed to find out whether there is any significant variance in the final scores that can be accounted for by ability, effort, luck, or task difficulty attributions.

The answer is no for the ability, effort, and luck, but yes for the task difficulty attributions. Ability, effort, and luck did not reach significance as covariates in the ANCOVAs with the OAS. For task difficulty, the covariate reached significance within two parts of the design. The attribution covariates did tend to cause the three way

interaction of the P and I targets with version to drop from significance. This would suggest that the version differences are attribution related. The adjusted means and ANCOVA summary tables for the four attribution measures are presented in Tables 54 through 61 in Appendix A (starting on page 239).

The ANOVA for the OAS raw scores (i.e., without covariates) contained main effects for the P and I targets, an I by V interaction, and a P by I by V interaction. When run with ability as a covariate (Tables 54 and 55 in Appendix A, pages 239-240) the P, I, and I by V effects remain, but the P by I by V interaction becomes a P by I two way interaction. None of the covariate effects are significant. For the effort analysis of covariance (Tables 56 and 57 in Appendix A, page 241) the P, I, and I by V effects remain only. The P by I by V effect approaches significance $\underline{F}(4,195) = 5.13$, $\underline{p}<.0104$). The effort covariate within the P by I interaction also approaches significance $\underline{F}(1,195) = 8.82$, $\underline{p}<.0166$). For the luck ANCOVA (Tables 58 and 59 in Appendix A, page 242) all of the ANOVA effects are duplicated without any significant covariate effects.

For the task difficulty ANCOVA (Tables 60 and 61 in Appendix A, page 243) the P, I, and I by V effects are replicated. The P by I by V effect is not significant. More importantly, two of the covariate effects are significant. The task difficulty covariate within the I target ($\underline{F}(1,95) = 11.16$, $\underline{p}<.0010$) as well as the covariate within the P by I interaction ($\underline{F}(1,191) = 9.89$, $\underline{p}<.0019$) are both significant. For the covariate within the I effect the regression coefficient is 0.35,

for the covariate within the P by I interaction the regression coefficient is .30. A review of the sphericity test for these analyses shows that the covariate within the I target falls within acceptable limits, but the covariate within the P by I interaction does not. For this reason the latter significant covariate is suspect and will not be discussed further.

Follow Up Analyses

Each participant completed a follow up questionnaire. The items were used to double check on the expertise classification of the ratee, to assess the respondent's reaction toward the videotapes, to find out if the respondent understood what was going on, and to determine whether the raters followed instructions. It turns out that some of the nonexperts may have in fact been experts. Second, raters saw the videotapes as having face validity. Finally, raters reported following the instructions. However, some of the findings suggest that they may have in fact failed to do so in some instances. The follow up analysis suggests that the lack of expertise effects may have been due to misclassification of respondents.

Expertise classification

One potential problem in the research is whether the raters were correctly classified as experts and nonexperts. Raters had been assigned as experts or nonexperts according to their job classification. First line supervisors were experts. Employees of the first line supervisors were nonexperts. To assess the effectiveness of the

expertise classification each respondent was asked to give the number of months of experience they had in counselor, leadworker, supervisor, and manager jobs, and the number of performance appraisal sessions they had participated in as a supervisor. Although the experts had significantly more supervisory experience than nonexperts, some of the nonexperts, it turns out, did have substantial experience as supervisors.

There were few differences between the experts and the nonexperts in the number of months of management experience possessed. Most had none. One expert (who viewed version one), reported having six months of management experience. Three nonexperts, all viewing version two, reported having management experience of one year, 16 months, and two years.

In the analyses on the number of months of supervisory experience (the most important distinction between the experts and nonexperts) there was a significant effect for expertise ($F(1,76) = 58.41$, $p \leq .0001$). Experts reported having a mean of 71.8 months while the nonexperts had a mean of 6.0 months of supervisory experience. Ten of the 38 nonexperts reported having between two months and four years of supervisory experience. None of the experience was in the job class of the experts, and none was within the past year. Apparently, these ten nonexperts had supervisory experience while working in college or in previous jobs.

There were few differences between experts and nonexperts in the number of months of leadworker experience and in the number of performance appraisals conducted while working as a leadworker. The nonexperts had a mean of 0.26 months of leadworker experience and had

conducted a mean of 3.1 performance appraisals as a leadworker. The experts had a mean of 0.18 months of leadworker experience and had performed a mean of 1.7 performance appraisals as leadworkers. Seven of the nonexperts reported performing between 2 and 52 performance appraisals as leadworkers.

Finally, since a good deal of the interpersonal relations dimension was developed from clinical psychology, respondents were asked to state the number of months of clinical or therapeutic experience they had. Nonexperts had a mean of 11.2 months of counseling experience and experts had a mean of 5.2 months. Ten of the nonexperts reported having between 2 months and 4 years of counseling experience.

In sum, the experts did have substantially more direct supervisory experience than did the nonexperts. Some of the nonexperts, however, did have substantial experience as experts. The effect of this potential misclassification has an unknown effect on the differential accuracy results. The obvious conclusion is that results would be stronger with a cleaner classification of expertise. Other research (cf Cline, 1964 and Wiggins 1973 for summaries), however, suggests that experts are not better than nonexperts on judgement tasks. Further research will be required to clarify this issue.

Face validity of the videotapes

Each videotape rating scale contained questions directed at whether there were any technical problems in the videotapes. First, respondents were asked if they could see and hear all of the tapes sufficiently well to make their ratings. There were no difference

between experts and nonexperts, nor between the two versions on any of the ratings. In addition, there were no differences between any of the tapes individually.

In an attempt to assess the quality and believability of the acting, respondents were asked to rate the importance of the performer's acting skills in the scores assigned. In assessing the effects of poor acting, there was a significant difference between rated acting importance for the I targets ($\underline{F}(2,132) = 6.56$, $\underline{p} \leq .0019$). Acting skill on the I6 targets was rated less important than for the I3 or I5 targets. Mean importance ratings (on a one [low] to seven [high] scale) were 2.1, 1.9, and 1.6 for I3, I5, and I6 respectively.

Raters had been told prior to initial viewing that the persons in the videotapes are actually role players reading scripts taped on a wall out of view of the camera. They were not professional actors and actresses. For this reason, their nonverbal behaviors may have been inconsistent with their words. Raters were also asked to rate the importance on their ratings of the actors or actresses words possibly being inconsistent with the nonverbal behaviors. Data analyses showed that there was a significant three way interaction for the P and I target levels with the expertise of the rater. Mean scores (on a one [low] to seven [high] rating scale) are presented in Table 13.

Table 13: Consistency of Nonverbal behavior		Means I3	Means I5	Means I6
P3	Expert	1.6	2.2	1.8
	Nonexpert	2.1	1.9	1.6
P5	Expert	1.6	1.7	1.7
	Nonexpert	1.8	1.7	1.7
P6	Expert	2.1	1.6	1.6
	Nonexpert	1.8	1.9	1.8

Respondents were asked if they knew any of the actors or actresses performing in the videotapes. Thirty-five of the experts and 15 of the nonexperts knew one of the actors. One nonexpert knew five of the nine actors and actresses. The one actor known by most of the respondents had worked as an employee development trainer in their department. Of the two tapes created by this actor, one was at the P3I3 level and the other at the P5I6 level. In addition, while 24% of the nonexperts in version one knew at least this one person, 84% of the nonexperts in version two knew him. All of the experts knew at least this one actor. None of the respondents reported that this knowledge had any effect on their ratings.

One follow up item asked whether respondents recognized any patterns within the presentation of the videotapes, 45% of the nonexperts and 58% of the experts reported seeing some sort of pattern. Again, when asked whether this had any effects on their ratings, none of the

respondents reported that it did have an effect. None of the explanations had any connection with the presentation pattern of the videotapes.

The responses to the pattern recognition questions should be taken with some skepticism. One entire data collection session of raters (six in all) failed to report inadvertently viewing the same target level script performed twice. Several hours after the group had completed their session, one of the respondents telephoned the researcher and asked the purpose of showing the same script twice, but performed by two different candidate supervisors. None of the group, including the telephone caller mentioned this in the follow up questionnaire or in the discussion following the session. All but one of the respondents (the caller) agreed to meet again to view one additional videotape. The person not viewing the tape was given missing (i.e., blank) values for the unseen videotape.

Following directions and understanding the rating dimensions

When asked whether they had followed the directions given by the researcher--to rate all of the items in the (counter balanced by rating dimension) order they were directed--three of the nonexperts reported failing to follow the directions. All of the experts reported following the directions.

Raters were also asked whether they understood each of the rating dimensions and whether the dimension's definition fit in with their own understanding of the concept being described. Responses were given on a one (low) to seven (high) rating scale. The mean understanding for

the problem solving rating scale was 5.2 for nonexperts and 5.3 for experts. Five respondents, one expert and four nonexperts, reported levels of knowledge of the problem solving rating scale below the mid-point of the knowledge rating scale. The mean interpersonal relations rating scale understanding ratings were 5.6 for nonexperts and 5.4 for experts.

Raters were also asked to judge the similarity between their own and the researcher's definitions of the two rating dimensions. For problem solving the mean expert similarity level was 5.5 on a seven point scale. The nonexpert similarity level was 5.2. On the interpersonal relations rating scale the mean definition similarity judgements were 5.2 for nonexperts and 5.5 for experts.

Summary of the Analyses

The repeated measures analyses of variance on the problem solving and interpersonal relations raw scores (i.e., the manipulation checks), differential accuracy measures, and confidence ratings are summarized in Table 14, below. The effects that did not reach at least the .01 level of statistical significance are deleted from the table. All effects listed are significant unique (SAS type III sums of squares) variance, after all other component's variances in the linear model have been removed.

Table 14: Summary of Significant Effects

		P	I	I*V	P*I	P*I*V
Manipulation Checks	PS	X	X	X		
	IR	X	X	X	X	X
Differential Accuracy	PS		X		X	
	IR	X	X		X	
Rater Confidence	PS				X	X
	IR		X		X	

note: PS = problem solving, IR = interpersonal relations, X = probability less than .01.

Table 14 points out the complicated nature of the results. The P and I target score effects were significant in every analysis, either as main effects or in an interaction. Confidence ratings also had significant P and I target interactions, plus a three way interaction with version for the problem solving scale. The version of the videotapes seen by the ratees had significant interactions, but no main effect for the raw scores. There were no effects for rater expertise.

The raw scores are the method of checking the effectiveness of the videotapes. For the tapes to be effective, differences must be found for the rating dimension's own target score, that is, a P target main effect for problem solving and an I target for interpersonal relations. These effects were significant. In addition, several other effects were observed. Perhaps the most important is the version interaction. The two versions were intended to provide similar effects. Although

there were no significant main effects for version, it did interact with the P and I targets. These effects will be summarized in the following sections.

Table 15: Summary of the Attribution findings

		P	I	I*V	P*I	P*I*V
Problem Solving Attribution Covariates	Ab		X		X	
	Ef		X		X	
	Lu		X		X	
	TD		X		X	
Interpersonal Relations Attribution Covariates	Ab				X	
	Ef		X		X	
	Lu		X		X	
	TD		X		X	
Attributions as Dependent Measures	Ab	X	X			
	Ef	X	X			
	Lu					
	TD					

note: PS = problem solving, IR = interpersonal relations, Ab = ability, Ef = effort, Lu = luck, TD = task difficulty, X = probability less than .01.

The attribution analyses are summarized in Table 15. The problem solving attribution covariates proved to have little influence in the differential accuracy analyses. They did not reach significance in any analysis, nor did they modify any of the original findings. For the

interpersonal relations analyses, the covariates did modify the differential accuracy findings slightly. All four covariates eliminated the main effect for the P target. The ability attribution also eliminates the I main effect. All four covariates thus had variance in common with the P main effect. This effect demonstrates that changes in the candidate's problem solving skill had an effect on the interpersonal relations ratings. The attribution covariates seemed to have accounted for similar variance. Further research is necessary to assess more clearly the influence of attributions on rater differential accuracy.

As dependent measures, only the ability and effort attributions proved to have any effects. Both had significant main effects for the P and the I targets. In both cases the importance rating of the attribution rose when the candidate's problem solving ability rose. For the interpersonal relations attribution ratings, ability importance ratings fell and effort importance ratings rose as interpersonal relations skills rose. The individual findings are summarized below.

Raw scores

Problem solving raw scores. For the problem solving raw scores, the three levels of the P target were all significantly different from one another. The manipulation of the problem solving rating scale was successful. The P3 and P5 means were approximately at their intended true score targets. The P6 mean, however, fell nearly a full point below its intended true score target. The P targets did not interact with version, even though the I targets did interact significantly with version.

The I targets also were significantly different from one another. Even though the three I target problem solving ratings should have averaged out to be the same for each of the I targets, they did not. I target problem solving scores were lower for I3 than for I5, and lower for I5 than for I6.

The I targets also interacted significantly with version. The problem cases were the I5 targets for the two versions. In the second version, the three I5 targets on average failed to be significantly different from the I3 targets. They were significantly different from the I5 targets for version one.

Interpersonal relations raw scores. The interpersonal relations rating scale also had the necessary differences between the three levels of the related dimension targets (I in this case) . The I3 mean target score was about a half a point above the true score and the I5 target a half a point short of the target true score. The I6 mean interpersonal relations score was short of the target true score by 0.3.

The interpersonal relations rating scales had parallel results for the opposite target as did the problem solving rating scale. In this case, the mean interpersonal relations ratings for the P3 targets were significantly different from both the P5 and the P6 targets. The latter two sets of ratings were not significantly different.

The P by I interaction occurred for the interpersonal relations ratings. This interaction, and the P main effect were both unanticipated. The I by V interaction was identical to that for the problem solving rating scale.

The three way interaction of the P and I targets with videotape version, although potentially problematic, proved to be of relatively little concern. The P by I interaction was significant within each of the two versions. The three way interaction was driven by a significant simple effect for version at the P6I5 target combination. The profile of the P by I interaction was a bit different between the two versions, but it still occurred.

Differential accuracy

Problem solving differential accuracy. For the problem solving differential accuracy analyses the hypothesis of an interaction between the P and the I target was supported. The second hypothesis, of an interacting effect for rater expertise, was not supported. One other significant effect (the I main effect) was unanticipated, but supportive of the main hypothesis.

There were no version differences in the problem solving differential accuracy analyses. Although version interacted significantly with the I target in the problem solving raw scores (at the I5 level) there were no differences here. Instead, there were differences within the I target and within the P by I interaction.

Problem solving ratings were significantly more accurate at the I6 target level--where videotaped candidates were skilled in interpersonal relations--than for the other skill levels of interpersonal relations. No matter which level of skill in problem solving was being portrayed, raters could rate that level with significantly more differential accuracy when the candidate being viewed was also interpersonally skilled. For candidates who were weak in interpersonal skills (clearly below an acceptable passing level), problem solving skills were rated a half of a point above the target level. For candidates with an average (i.e., five points out of seven, with a four point score being passing), problem solving ratings were nearly a point below the true score target.

The P by I interaction shows that problem solving differential accuracy scores fell (i.e., raters got more accurate) for the above average (P5) and high quality (P6) problem solving target scores as the candidate's interpersonal relations skills rose. Also, when the candidate's interpersonal relations skills were poor (I3) the problem solving ratings got worse (i.e., the differential accuracy measures got higher) as the candidate's problem solving skill got better.

The least accurate problem solving rating was at the P6I3 level. The most accurate rating was at the P6I6 level. As problem solving target skills got stronger, problem solving rating scale rater accuracy became more influenced by the I targets. Differential accuracy was at its best for ratees with strong skills in both the P and the I targets, worst with a strong P target and a weak I target.

Interpersonal relations differential accuracy. The interpersonal relations rating scale contained significant differential accuracy effects for the P targets, the I targets, and their interaction. These findings once again support the main hypothesis of an interacting effect for problem solving and interpersonal relations targets, but do not support the hypothesized effect for rater expertise, and fortunately they do not reveal the potentially harmful effect of version.

For the P targets, raters were most differentially accurate when rating interpersonal relations skills of candidates with above average, but not outstanding problem solving skills (the P5 targets). They were equally less accurate when rating the weak and the strong problem solvers.

In the I target main effect, raters were least accurate in rating interpersonal relations at the lowest level of skill in interpersonal relations. Raters were equally accurate when rating the I5 and the I6 targets. The I3 targets were candidates whose ratings, incidentally, were among those containing version differences in the raw scores.

The P by I interaction once again fine tunes the findings of the I target main effects. At the P3 level, low levels of I targets lead to the highest accuracy. However, at the P5 and P6 levels, the best accuracy was at the high skilled levels.

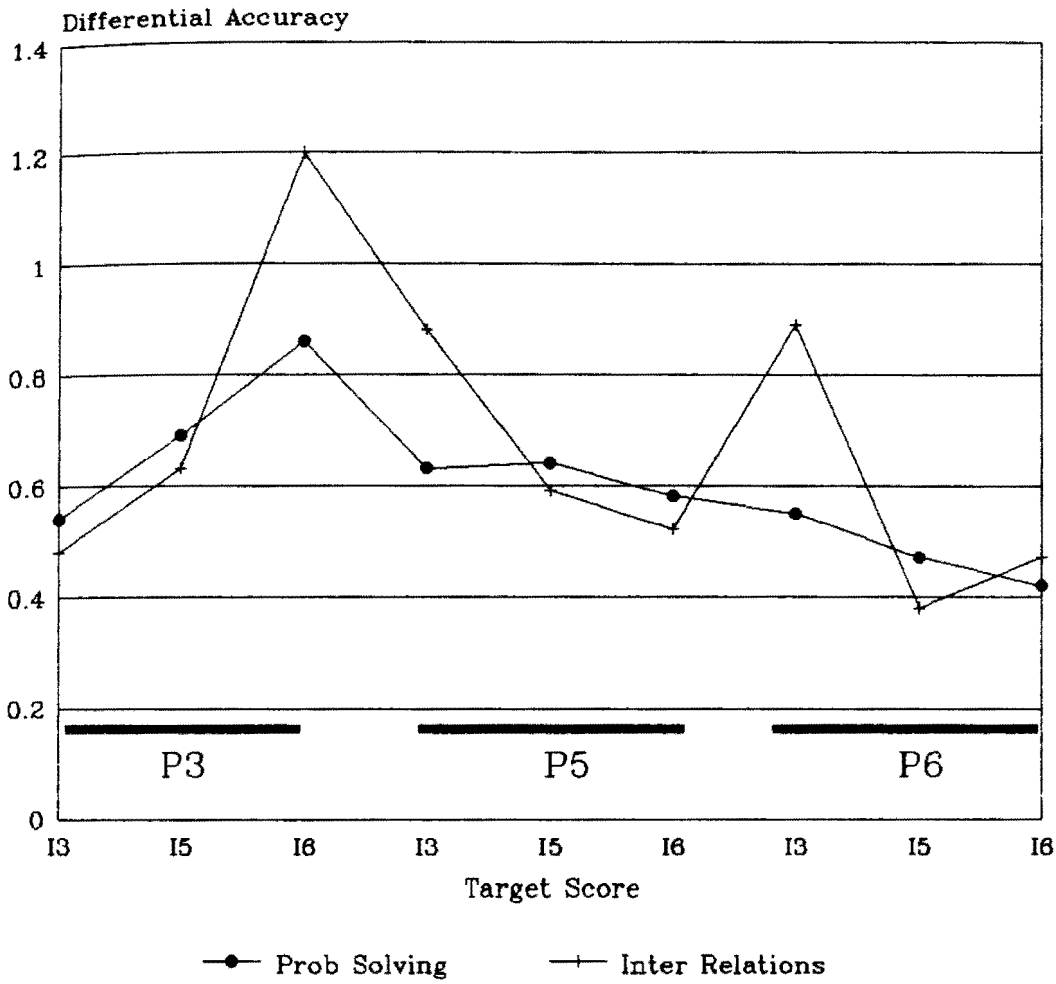
There were also interesting significant differences between differential accuracy measures within I target true scores. Ratings at the P6 level for I3 were larger (less accurate) than the P3 and P5 levels. For the I5 and I6 targets, the effect was reversed. The P3 targets were less accurate than the P5 and P6 target ratings.

The least accurate interpersonal relations ratings were made at the extreme opposite combinations of problem solving and interpersonal relations true score targets; raters were less accurate at rating interpersonal relations when poor interpersonal skills were crossed with strong problem solving skills and when above average or strong interpersonal skills were combined with poor problem solving skills. Raters were most accurate when the skill levels of the two rating dimensions were at similar levels: both weak, both a bit above the mid point, or both strong.

Differential accuracy for the two measures combined. The following two tables highlight the similarities and differences between the problem solving and the interpersonal relations differential accuracy ratings. The figures actually present the same differential accuracy data from two perspectives. Figure 1 presents the data sorted in groups of three by increasing problem solving target scores. Within each group of problem solving target scores, the three interpersonal relations targets are presented. The second figure presents the same data in the opposite ordering. The data are grouped by increasing interpersonal relations. Within each group of three interpersonal relations ratings are the three problem solving target scores.

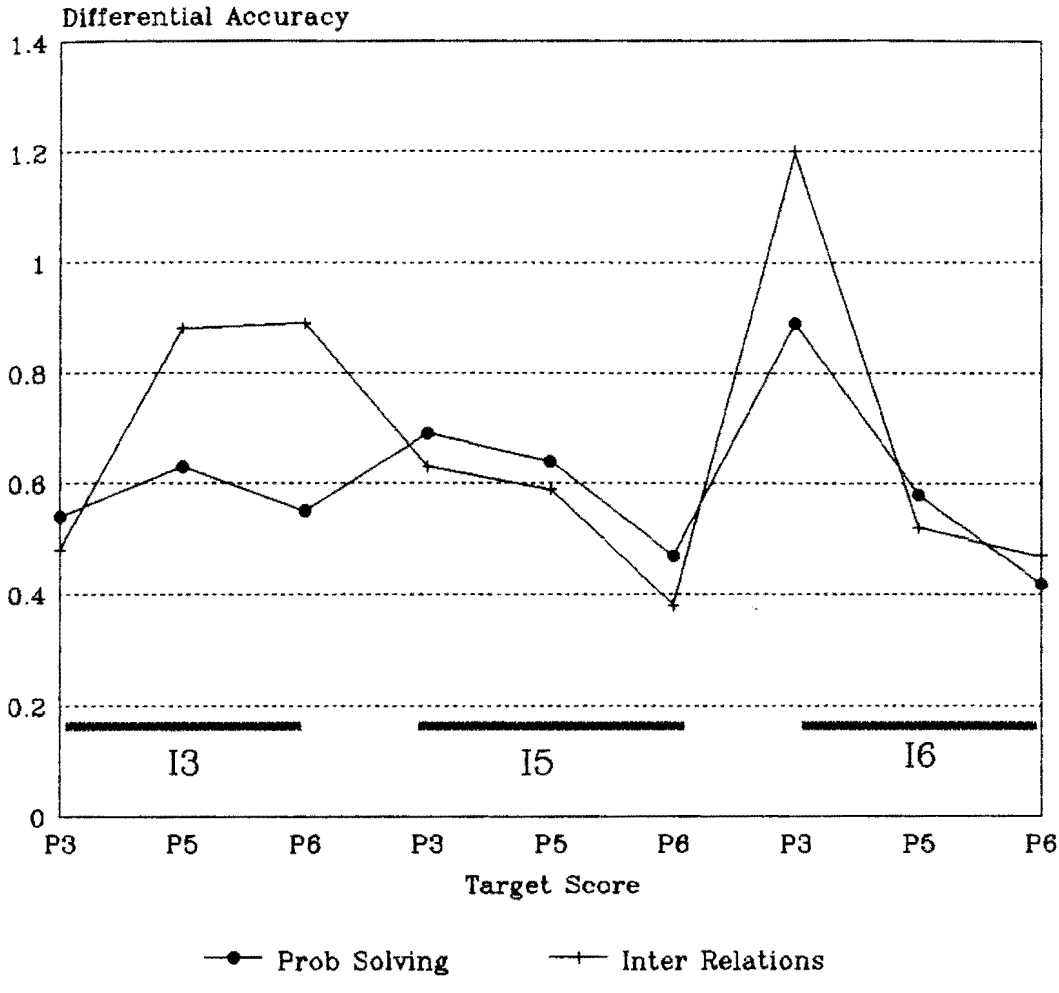
Looking at the Figure 1, sorted by the P targets, the results are remarkably parallel except for the P6I3 target. Differential accuracy of the problem solving and the interpersonal relations rating scales tends to rise and fall together.

Figure 1: Differential Accuracy
Sorted by Problem Solving
Means of the P by I interaction



In Figure 2, sorted by the I targets, the major difference in differential accuracy is clarified. The major differences in differential accuracy between the two rating scales fall within the I3 targets. Interpersonal relations differential accuracy for the P5I3 and P6I3 targets are widely different from the problem solving differential accuracy ratings.

Figure 2: Differential Accuracy
Sorted by Interpersonal Relations
Mean P by I Interaction



Larger differential accuracy scores (lower accuracy) appear to be occurring for the interpersonal relation rating dimension at the three problem targets mentioned above.

Another important piece of information to consider when exploring whether the problem solving and interpersonal relations rating scales could be combined is the intercorrelations of the differential accuracy

measures. Table 16 presents these intercorrelations. Problem solving target scores change across the rows. Interpersonal relations target scores change down the columns.

Table 16: Differential Accuracy Rating Intercorrelations

\PS IR \	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
P3I3	.12	-.02	-.02	-.02	-.08	-.08	.11	-.01	.03
P3I5	.32*	.12	-.06	.13	.12	.35*	.22	.37*	.20
P3I6	-.17	.17	.28	.08	.14	.01	.31*	.23	.53*
P5I3	.08	.20	.03	.11	.14	.26	.35*	.34*	.18
P5I5	.39*	.14	.06	.04	.02	.18	.17	.32*	.33*
P5I6	.18	.06	.05	.11	.22	.27	.20	.40*	.17
P6I3	-.09	.25	.21	.18	.07	.09	.22	.33*	.15
P6I5	.26	-.05	-.01	-.10	-.01	.08	.16	.27	.19
P6I6	-.06	.09	-.04	-.00	.01	.19	.20	.38*	.16

Note: * = probability less than or equal to .01

Thirteen of the 81 correlations reach significance at the .01 level. Ten of the 13 significant correlations occur among the P6 (top quality) problem solving rating correlations with interpersonal relations (the last three columns of Table 16). Eleven of the 13 significant correlations occur in the last four columns of Table 16. These

are the top problem solving dimension accuracy measures. Individuals who were differentially accurate at rating problem solving for ratees at the P6I5 level also tended to be differentially accurate at rating all levels of interpersonal relations skill. The intercorrelations provide no solid support for the hypothesis that the problem solving and interpersonal relations rating scales can be combined.

Confidence ratings

The confidence ratings are used as supplemental analyses. Their value is in supporting, expanding and explaining the differential accuracy results. The same is true for the analyses on the attributions, follow up items, and the overall assessment ratings.

Problem solving confidence ratings. There were two significant effects for rater confidence with the problem solving rating scale, a P by I interaction, and a P by I by V interaction. Raters were generally confident about their ratings. Mean confidence ratings in the nine cells of the P by I interaction ranged from 4.9 to 5.9. In general, raters were least confident in those ratings that proved to be least accurate and most confident in those ratings that were the most differentially accurate. Significantly the lowest confidence ratings were in the above average and top scoring problem solving targets combined with below average interpersonal relations skills, below average problem solving skills combined with top interpersonal skills, and for the raters with above average (P5) interpersonal relations skills at all problem solving skill levels.

Raters were most confident when rating the problem solving skills of the candidates for whom they were also the most accurate: the candidates with consistent top or consistent below average skills on both dimensions at the same time. The three way interaction with version showed that raters rating the P3I6 candidates were significantly different between the two versions. Version two had the higher scores. More importantly, the P by I interaction was significant within both versions.

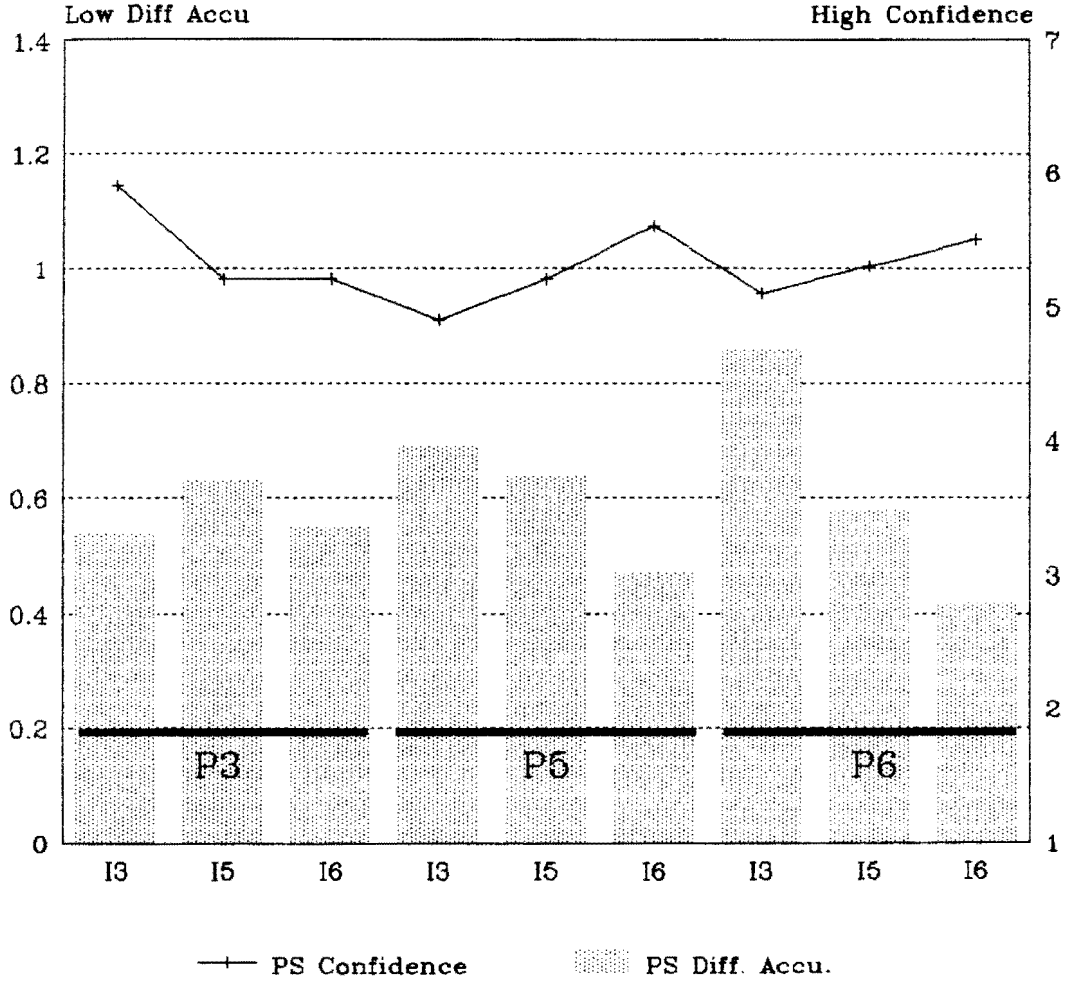
Interpersonal relations confidence ratings. For the interpersonal relations confidence ratings raters were most confident in the ratings that were the most accurate: the I6 target ratings. Raters were least confident in the ratings for which they were least accurate.

The P by I interaction shows that the highest confidence ratings were for candidates with top skills in interpersonal relations combined with top scores on problem solving. High confidence ratings also were given for the opposite end of the quality of performance scale, for the candidates with below average problem solving and interpersonal relations skills. The lowest confidence was assigned to ratings for divergent combinations of problem solving and interpersonal relations skills: weak problem solving target skills with above average and top interpersonal relations skills, or below average interpersonal relations skills with above average problem solving skills.

Confidence ratings and differential accuracy. The figures below present the two sets of differential accuracy and confidence ratings simultaneously. Both sets of measures rise and fall in near unison and

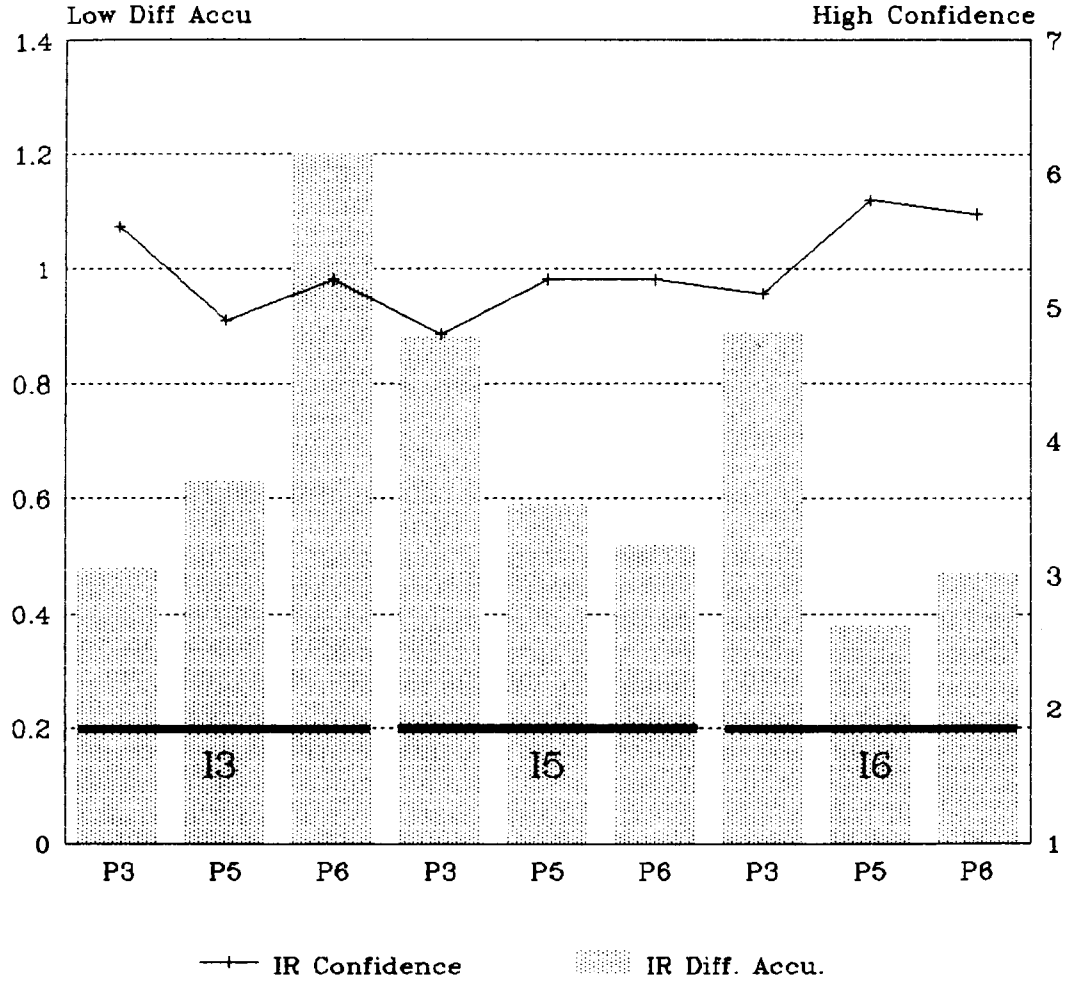
fall between the five and six point scale values. Figure 3 shows the problem solving rating scale differential accuracy paired with rater confidence in the problem solving ratings. Figure 4 pairs the same information for the interpersonal relations rating scale.

Figure 3: Problem Solving Dimension
Differential Accuracy and Confidence
P by I Target combinations



For the problem solving rating scale, the P3 target level confidence ratings drop from I3 to I6, yet the differential accuracy rises and then falls. For the P5 and P6 target combinations, confidence rises as accuracy improves.

Figure 4: Interpersonal Relations
Differential Accuracy and Confidence
P by I Target combinations



For the interpersonal relations ratings, confidence falls and rises for the I3 targets as differential accuracy scores rise (i.e.,

ratings get less accurate). Differential accuracy improves (values get smaller) and confidence ratings get stronger. The same generally happens in the I6 target level series.

Rater attributions

Attribution ratings were analyzed to assess whether a rater's differential accuracy could be accounted for by his or her ratings of the importance of ability, effort, luck, and task difficulty in the ratings of the dimensions. To a minimal degree this covariation was found for the interpersonal relations dimension, but not for the problems solving dimension.

Attribution ratings were analyzed in two ways, first as covariates within the main problem solving and interpersonal relations differential accuracy designs and second as independent variables. The covariate analyses for problem solving had little effect. For interpersonal relations, the significant P main effect in the differential accuracy analyses tended to fall to nonsignificance, indicating covariance with the various attribution ratings. As dependent variables, only the ability and effort covariates had any mean differences within the P and I effects. Mean scores for the covariates decreased from ability to effort to task difficulty to luck. Both ability and effort had main effects for P and I. The luck and task difficulty attributions had no significant effects within them.

For problem solving ratings, the ability, effort, luck, and task difficulty attributions did not account for any significant variance as covariates in the differential accuracy analysis. In addition, the

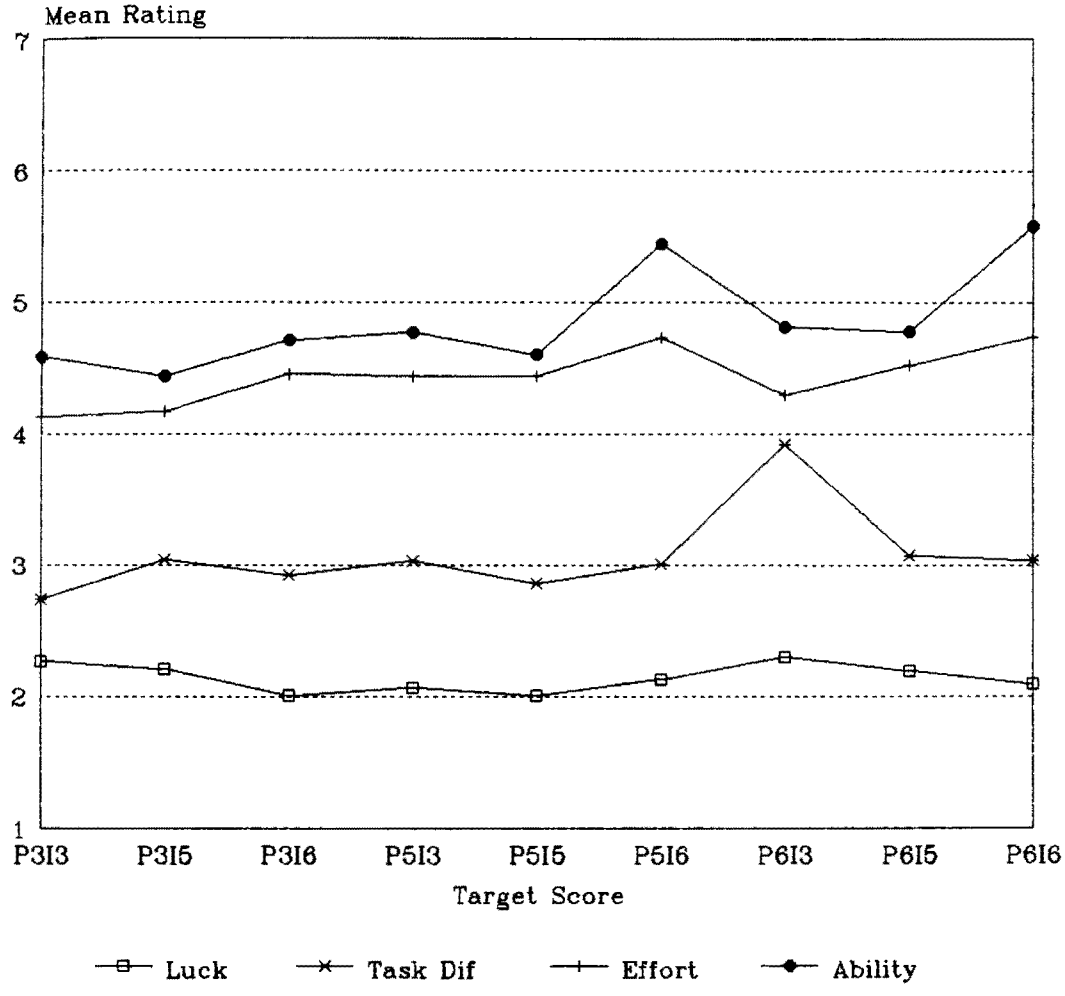
same effects were significant in the covariate and original models, the I main effect and the P by I interaction. The comparisons of differential accuracy means before and after the removal of the variance due to ability attributions suggests that some effects may have been taking place. To some degree, problem solving differential accuracy for candidates with poor interpersonal relations skills got worse when ability attribution variance was removed from the analyses. Similarly, problem solving differential accuracy improved for candidates with high interpersonal relations after the ability attribution variance was removed. The interpretation of these findings remains to be clarified by future research.

For the interpersonal relations ratings, the attribution covariates changed the differential accuracy findings. Adding the ability covariate to the differential accuracy model caused the P and I main effects to drop from significance. For the effort, luck, and task difficulty attribution covariates the P main effect dropped from significance. These effects showed that the rater's attributions tended to covary with hypothesis consistent interpersonal relations differential accuracy effects.

Rater attributions and rater differential accuracy. Figure 5 below presents the four rater attribution composite values at each of the nine target combinations. The figure highlights the importance of each attribution. Luck had the weakest effect and did not change in importance over any of the nine targets. Task difficulty also had a relatively small effect, but values were higher than for luck. The

P6I3 target had a larger task difficulty attribution. The two most important attributions were for effort and for ability. Each had significant differences and were rated as higher on the P5I6 and P6I6 targets.

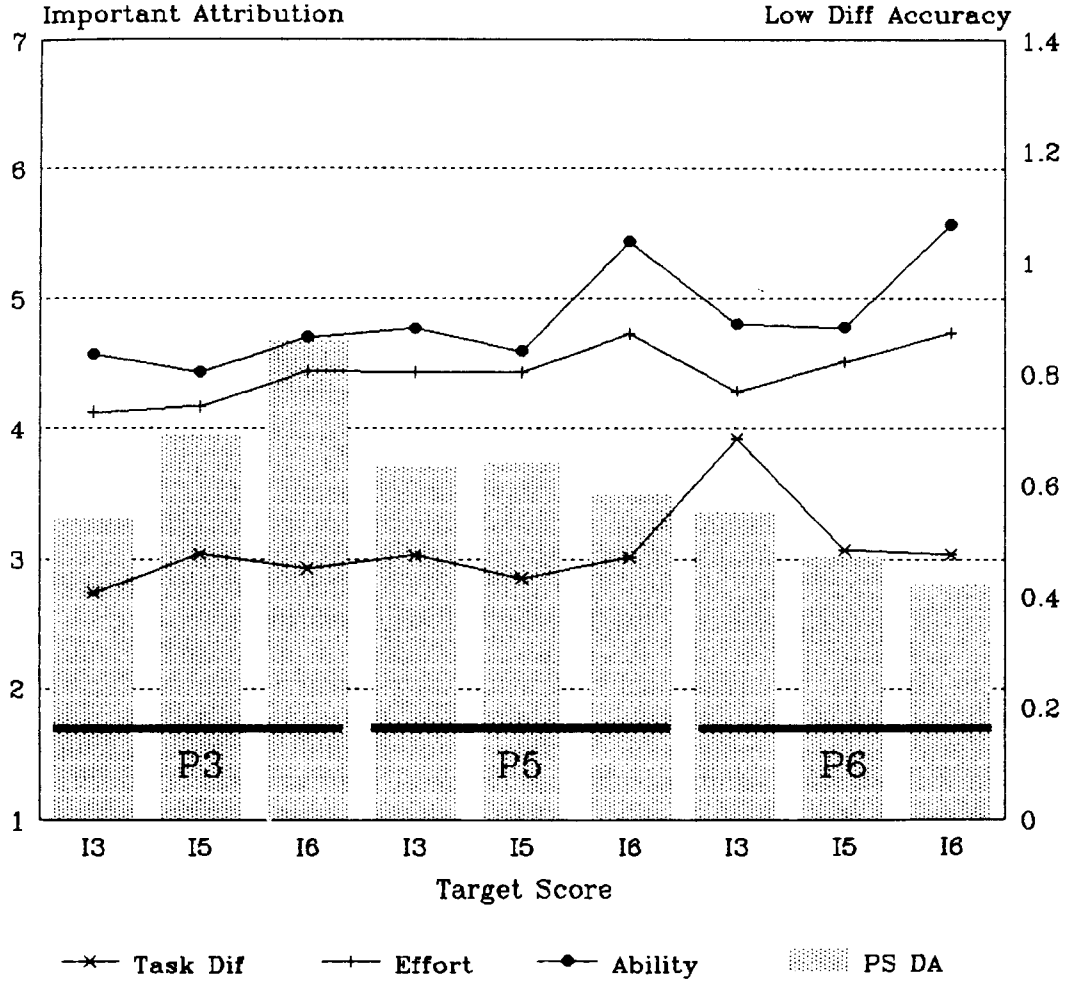
Figure 5: Attribution Ratings
P by I Target Combinations



The next two figures (Figures 6 and 7) present data for the interpersonal relations and problem solving differential accuracy measures and the attributions. Differential accuracy is present as bars. The

rater attributions are presented as lines. The purpose for the figures is to highlight the interrelationship between rater differential accuracy and rater use of various attributions. The luck attribution is not presented because it is essentially a flat line with weak values at each target.

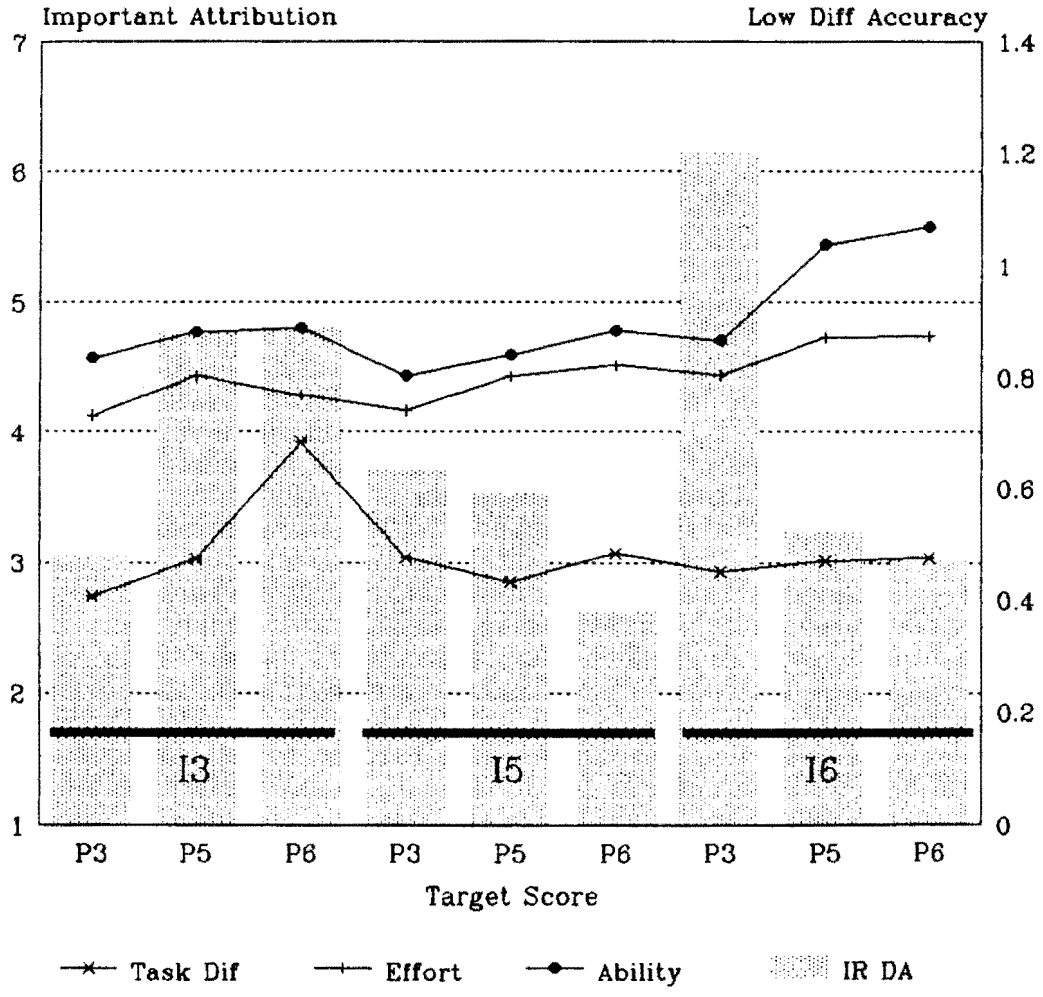
Figure 6: Problem Solving
Differential Accuracy & Attributions
Sorted By P Targets



For the problem solving rating scale, inaccuracy rises over the P3 values from the low to the high I targets. Accuracy stays constant

over the P5 targets and falls over the P6 targets. The ability, effort, and task difficulty attributions remain constant at the P3 level. At the P5 target level, the ability and effort attributions rise at P5I6. For the P6 target levels, the ability and effort attributions rise while the task difficulty rating drops as the I targets rise.

Figure 7: Interpersonal Relations
Differential Accuracy & Attributions
Sorted By I Targets



For the interpersonal relations target ratings the task difficulty and ability attributions rise in the I3 targets as the P target reaches P6. Rater accuracy also declines (i.e., the DA values rise). In the I5 and the I6 target series accuracy improves as rater attributions in ability and effort increase. Task difficulty attributions remain constant.

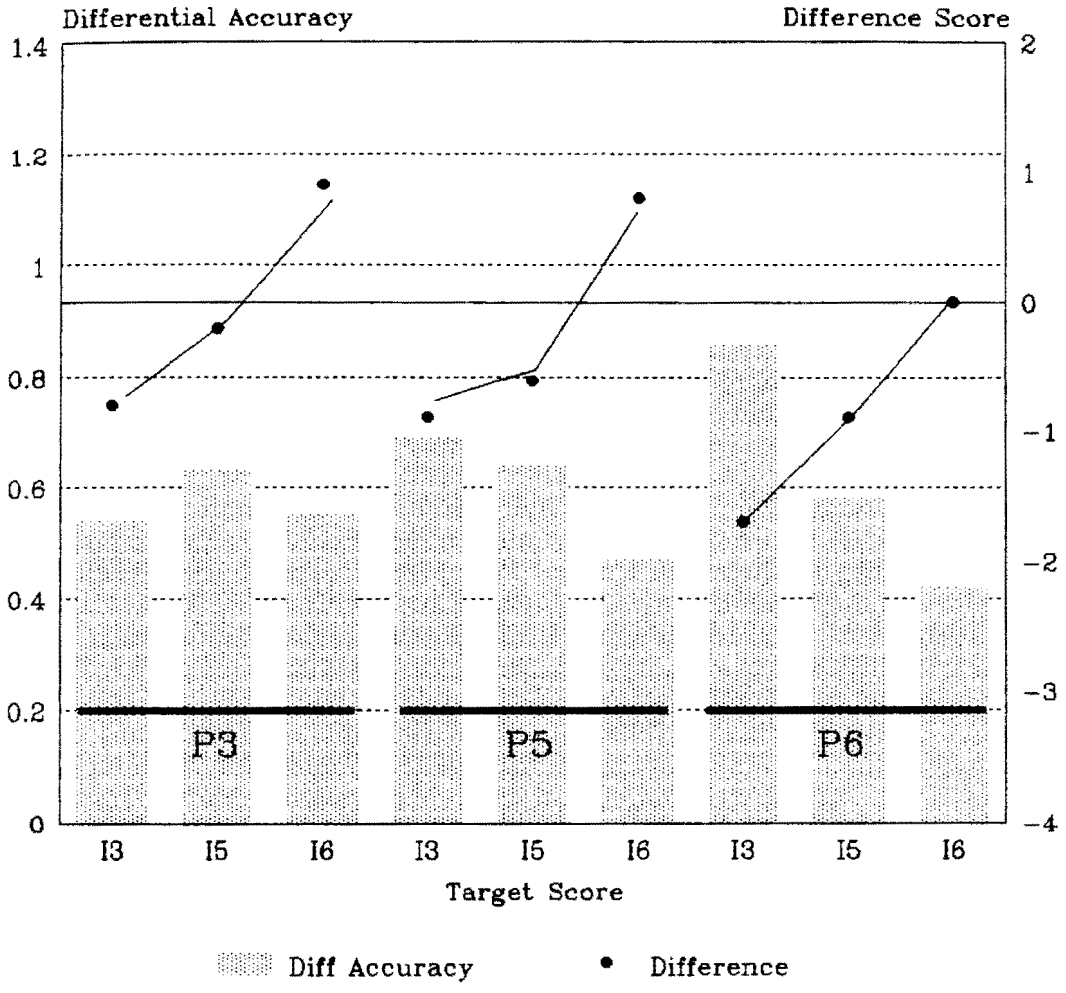
Differential accuracy and difference scores

With the finding that the two rating dimensions interact with one another, the question of the practical significance of this finding arises. It is of great practical concern whether the ratee's scores rise or fall when the rater becomes less accurate. If a candidate is benefited by the differentially inaccurate rating, and receives a higher score than deserved, then criteria will be less strict than appropriate. People will be hired who should not be hired. On the other hand, if inaccurate ratings are a sign of lower scores than appropriate, people who deserve a chance at the position may not get it.

Figure 8 presents the data simultaneously for differential accuracy and rater difference scores from the targets. The difference scores are centered around the zero point using the scale values on the Y axis to the right of the figure. Differential accuracy values for the bars are listed on the left hand side of the figure.

For the problem solving ratings the results are clear and direct. The ratings over the P3 targets are most accurate at the I3 and I6 targets. Ratees who are also weak on interpersonal relations are rated

Figure 8: Problem Solving Differential Accuracy & Difference Scores Sorted by P targets



worse than they should be rated. Ratees who are strong on interpersonal relations are rated better than they should be rated on problem solving ratings, although raters are about equal in differential accuracy.

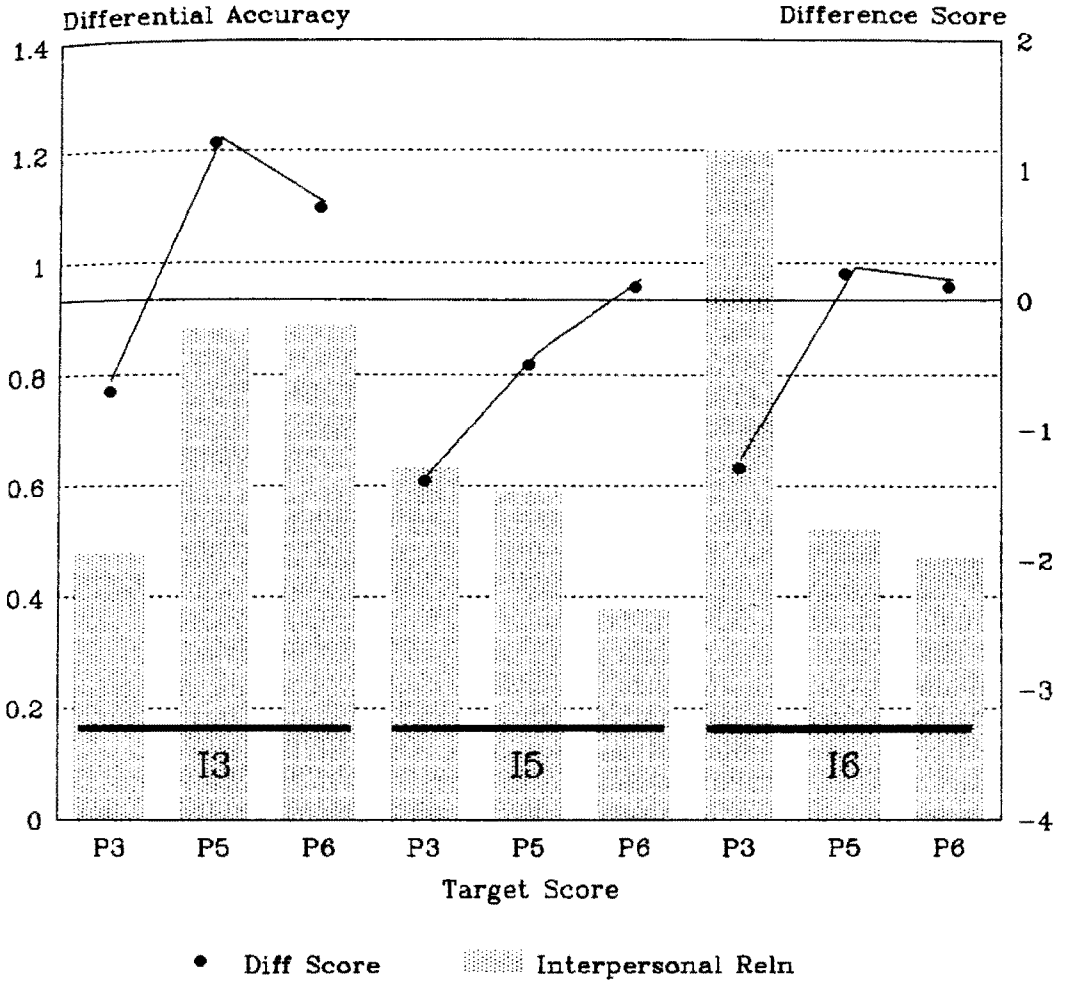
The P5 group of ratings points out the relationship between difference scores and differential accuracy. The same pattern in difference scores occurs as was in the P3 target series, but accuracy

improves linearly. Raw scores are higher than they should be at the P5I6 target, but the raters are most differentially accurate here. Generally, as raw scores rise, raters get more differentially accurate at the P5 level of problem solving true scores.

This same generalization also occurs at the P6 target level. This time, raters are most accurate and have difference scores closer to zero (not different from the target value) at the highest level of interpersonal relations.

The interpersonal relations ratings (Figure 9) generally tell a similar story at the I5 and the I6 groups of target ratings. As raters become more differentially accurate, their raw score ratings become larger in value and closer to the target. At the lowest level of interpersonal relations skills, raters get less differentially accurate as the ratees' skills in problem solving rise. Raters with weak interpersonal skills and above average or top quality problem solving skills have their interpersonal relations skills rated better than they should be rated. At the P6I3 target the difference scores actually start to drop back toward the target true score, but the raters remain differentially inaccurate.

Figure 9: Interpersonal Relations
Differential Accuracy & Difference
Scores Sorted by I targets



The Feldman model states that ratings will be accurate to the degree that the category prototype used by the rater in observing the behavior are the same as the category prototypes used in the rating scales. Several hypotheses were made based on this model.

First, it was hypothesized, and subsequently supported, that the rating dimensions will interact with one another. The Feldman model assumes that raters use a single category within which to observe, store, and recall information about any one interaction. If this is the case, then forcing the rater to rate the candidates on two rating scales will cause the information blended in the observation category to be split onto two rating scales. This is what was observed. Once blended together, the information was not cleanly separable again for all ratees.

The second hypothesis was that the experts should be more differentially accurate than are the nonexperts on some of the rating dimensions than on others. Their expertise was defined by having higher levels of skill for the dimensions being rated. This should allow them to have a more thorough understanding and be more differentially accurate than nonexperts for some quality levels of performance. This hypothesis is not supported. The result is no difference between experts and nonexperts. One potential reason for the failure to support the hypothesis is a significant amount of supervisory experience shown by the some members of the nonexpert group.

The third hypothesis centers around the two versions of the target videotapes. The versions are replications. There should be no difference between the replications if the above two hypotheses are to be firmly supported. However, if the actors and actresses performing as candidates for supervisor in the videotapes have a different understanding of the skills being portrayed, then they will enact the skills using slightly different prototypes. This may cause their performances to match their personal level of understanding rather than the true score level. If this is the case, then there will be different ratings assigned to the two versions of the target videotapes at a single level. The results were that their were differences (mainly interactions) involving versions for the raw scores, but not for the differential accuracy scores. The raw score version differences proved to be of limited importance, showing differences in the relative amount of the effect in both versions, not the presence or absence of the effect in one of the versions. The support for the first hypothesis is not threatened by version differences.

Ratee Problem Solving and Interpersonal Relations Skills

Perhaps the strongest finding of the study is the interaction between the problem solving and interpersonal relations targets for both rating dimensions. A person's skill in one dimension clearly affects how he or she is rated in the other dimension.

This raises the question of whether the two rating scales should be combined into a single "supervisory skill in problem solving meetings" skill factor. This factor would definitely not be a small,

focused set of key actions and accountabilities so popular in current management training courses. Rather, the combined problem solving and interpersonal relations rating scales would be a complex combination of analytical and interpersonal tactics used to solve problems in this specific one-on-one situation.

This possibility was assessed in two post hoc comparisons run comparing differential accuracy for the P by I cells with the same target scores (i.e., P3I3, P5I5, and P6I6, the main diagonal) with the other, off diagonal target score combinations. Problem solving dimension main diagonal ratings were more differentially accurate than off-diagonal ratings. Interpersonal relations main diagonal ratings approached being significant. Thus, the tentative conclusion of further considering whether to combine the interpersonal relations and problem solving dimensions does receive some modest support. Differential accuracy may be better when the two rating dimensions are at consistent levels of performance than when they are at inconsistent levels of performance. More research must be done to more thoroughly assess how the two rating scales interact.

Combining problem solving and interpersonal relations into one rating dimension will not solve all of the issues presented in these data. Some information contained in the inconsistent performance levels would be lost if the problem solving and interpersonal relations dimension ratings are combined, given their present definitions. Not only did the two rating scale target levels interfere with one another, but raters were not consistently inaccurate at the six off-diagonal

performance levels. Raters tended to be less accurate when ratees performed with high problem solving skills and poor interpersonal relations skills. This happened for both rating dimensions. Thus, ratees also performed inconsistently on the two performance dimensions.

Combining the two rating scales into a single one may only lead to greater differential accuracy for the consistently performing candidates. The inconsistently performing candidates, those with high skills in problem solving, but not interpersonal relations, or the reverse, would still not be rated accurately.

What these data suggest is an alternate way of looking at the two problems. Supervisors (the ratees in this study) who can follow exactly the steps involved in solving a problem but who cannot interact with interpersonal skill with a subordinate may not be skilled in problem solving after all. They are technically sound, but may, for example, lack the skill to gather the information in a productive manner. The facts come through, but the subordinate feels so attacked that he or she may quit or file a grievance against the supervisor. One issue is resolved only to be replaced by another. In the same manner, a supervisor who has high quality interpersonal skills, yet manages to resolve the wrong problem, will be of little value as a supervisor. The definitions of the rating dimensions, which were theory-derived, may need to be update.

In developing the videotapes for the extreme opposite values of performance on the two rating dimensions, considerable time was spent

on developing a mental picture of who these people might be. The person who is highly skilled in problem solving, yet ineffective at interpersonal relations was built to be like a worst nightmare of an attorney cross-examining a hostile witness. The candidate who was developed to be strong in interpersonal skill, yet poor in problem solving was pictured as an ineffective social worker, that is, someone who could conduct a skillful conversation with a person who is hostile and living with undiscovered problems. The ineffective social worker spends too much time setting the mood and as a result avoids working on the "real" problems. Side issues, which are brought up by the subordinate as an excuse, are the focus of the meeting.

In both of these cases, the candidate-supervisor may in fact be highly skilled in the rating dimension on which they are scoring poorly. They may intentionally be misusing the skills, as in the case of the attorney. They may be applying the skills to the wrong problem, as did the ineffective social worker.

Both of these candidate-supervisors may also, in a sense, be considered ineffective on both rating dimensions. How can a social worker be considered to be interpersonally skilled, as defined in this study, and yet not assist people in resolving problems? How too can a person be a highly effective problem solver and still create more problems in the process through their interpersonal abruptness? It may be that the attacking attorney and the ineffective social worker are not effective at all in the skills for which they are rated the most highly.

Feldman's Cognitive Process Model

All of this discussion focuses directly on the Feldman (1980, Ilgen and Feldman 1983) cognitive model. The model proposes that an observer, recorder, and evaluator of behaviors forms a single cognitive picture of a person. This single cognitive model may contain a variety of attributes. The less the amount of actual information, the greater the reliance on preexisting attribution types of information.

The rater attributions may provide the most interesting evidence in the accuracy research. Although no specific hypotheses were developed about rater attributions outside of those presented in the Feldman model, the variables can help link the raters' intentions to their respective outcomes.

In this research the attribution ratings did not live up to their expectations. Despite preparation on the phrasing of the attribution items, many of the participants in the research did not understand the meaning and/or the purpose of the questions. Extra time in nearly every data collection session needed to be taken, after the first videotape was seen and most of the ratings of the interpersonal relations and problem solving dimensions were completed, to explain for a second or a third time the method of answering the attribution items.

Even with this extra work, many of the variables had little variance and some confidence must be withheld from the interpretation of the outcomes of these variables due to the confusion of the respondents. Many respondents chose a single number and rated all of the attributions with that number. The number picked was the scale

midpoint for some and the lowest point on the scale for others.

Raters did, however, tend to rate the attributions in a manner that is consistent with expectations. Ratee's ability (internal, stable) was the most important, with effort (internal, unstable) close behind. Luck--an external, unstable attribution--was unimportant. Task difficulty, an external, stable attribution was also relatively unimportant. In only one case did it approach the midpoint of the importance rating scale.

With the high importance attributed to ability and effort, both internal characteristics of the candidates, one may assume that the candidates were not intentionally assigning ratings based on factors external to the performance of the ratees. This does not mean that the Feldman model is not supported in its claim that candidates who are not able to rate an observation with differential accuracy are instead using stereotypes or other attributional biases. On the contrary, it may support the claim.

If the ratees' behaviors were seen as being simply scripted into the performance, a fact that was understood by all raters prior to viewing the first candidate, then the importance of external factors would have had higher values. Instead, the internal factors were rated as more important. In an actual selection situation, raters would presumably make only stable attributions unless situational factors such as the weather (external, unstable attributions) or candidate nervousness (internal, unstable attributions) were apparent. Any task difficulty attributions should have been removed by the test developer. A task that is too difficult for the candidate would mean that either

the candidate was not properly screened or the simulation exam was not job related. In future research measuring these attributions, more open ended, less obtrusive attributions may be more useful.

The interference of the two rating dimensions with one another may make the strongest statement about the effect of rater attributions on rater differential accuracy. When there was consistency of performance, as with the P3I3, the P5I5, and the P6I6 target ratings, raters tended to be more differentially accurate than when the target ratings were widely divergent. If raters formed only one internal picture of the candidate (i.e., as an interpersonal problem solver) and the candidate was actually performing on two different dimensions, raters could not observe the behavior with their preset cognitive category.

Instead, they were forced to rely on the single cognitive category they had used during observation to make decisions about a candidate who did not display behaviors that were associated with that cognitive category. The ratings were less accurate as a result. When no information was available from memory on a person who was an "attacking attorney" style problem solver or as a "ineffective social worker," then the ratings were made on the available skills.

When rating problem solving, candidates with poor problem solving skills were rated higher than they deserved when they also displayed high skills in interpersonal relations. Candidates were assigned lower problem solving dimension ratings than they deserved when their strong problem solving skills were paired with weak interpersonal skills. When rating interpersonal relations, candidates with poor interpersonal skills were assigned scores that were too high when they exhibited

above average or top quality problem solving skills. Candidates were assigned interpersonal relations ratings that were too low when their high interpersonal skills were paired with poor problem solving skills. Raters may have been making attributions about the candidate's skills in one dimension based on attributions developed from information from the other rating dimension.

An interesting further study would be to investigate the relationship between halo effect and rater differential accuracy. Halo effect occurs when a rater fails to distinguish among a ratee's weaker and stronger skills. All ratings are inflated by the rater's positive regard for the ratee. The hypothesis suggested by the present study is that halo effect occurs (i.e., raw scores are higher than appropriate) when rating problem solving skill (or perhaps other, more technical skills) without accurately considering the confounding effect of strong interpersonal skills. This and related hypotheses deserve additional research.

Expertise and Differential Accuracy

One of the thrusts of this research was to assess the difference between the rating accuracy of experts and nonexperts. No differences in accuracy were found. Instead, some evidence was presented that some of the nonexperts may have been qualified as supervision experts. The question of whether expert raters are more accurate than nonexperts must be answered negatively in this study. This is in general agreement with the findings of the clinical psychology literature (cf. Wiggins, 1973). Einhorn and Hogarth (1978) suggest a variety of reasons

why experts might be no more accurate than nonexperts. The biggest reason fits directly in with the Feldman (1980) model. Raters have a tendency to notice items that support their existing schemas and ignore contrary information. Future research must be conducted to find out why this was the case. The current findings suggest that experts are no more accurate than nonexperts, given equivalent rater training.

An alternative, less likely solution is that the expertise of the expert raters was in an area other than that tapped by the two rating dimensions. The one supervisory performance dimension not studied in this research was technical expertise in the job of welfare eligibility technician. If this was the actual basis for the expertise classification, then none of the supervisors would be qualified for their jobs, an extremely improbable possibility.

Importance of the Version Differences in the Raw Scores

The utility of the videotapes may be called into question by the discovery of significant version differences. Version differences were found in the interpersonal relations raw scores. Most of the version differences were at scale midpoints.

These findings question the believability of the tapes. The tapes may have been seen mostly as research instruments not relating to the actual ability of the people seen in the tapes. This was generally the case. Long hours were spent with many of the actors and actresses in training on the phrasing of the scripts. One actress was replaced when she could not present the role randomly assigned to her in the tone and style of the original version of the role. Another actress commented

that the role randomly assigned to her was clearly a "male" role (and it was originally). The script contained references to Chester's drinking over lunch as "having a few 'bumps' over lunch." This is a phrasing that could have sexual rather than the intended alcohol related connotations.

These speculations, however, do not call into question the differential accuracy findings. Despite these differences in raw scores assigned to candidates for some P by I target combinations in the two versions, the hypothesized P by I interaction still occurred in the raw scores within both versions. There were no version effects within the differential accuracy findings. The raters are still rating the videotapes with the measured amount of differential accuracy.

What it does suggest is that the raters clearly understand that the tapes are staged. Several comments by participants support this speculation. One of the actors used in the tapes was a Personnel Department Staff and Management Development Instructor and an actor in community theater. In both roles he was known to many research participants. In one of the two videotaped roles this actor portrayed, the target scores were P3 and I3, the lowest scores on the tapes. Personally, this actor was known as a capable and professional speaker and trainer of the skills he was demonstrating at a poor quality. Laughter and jokes would sometimes rise from the research participants when he made his first supervisory blunder in the videotape. Many participants commented verbally, though not on paper, that they knew that this actor (Steve) was actually a much better supervisor than he

was portraying. Of course, these problems will not arise in any future research using these videotapes using individuals who are unfamiliar with these actors and actresses.

The Videotape Stimulus Research Methodology

The cost of doing interviewing or performance appraisal research with videotaped, scripted interactions between role players is very high. The process is also quite time consuming. Before deciding to utilize this research methodology it is important to consider the costs and benefits.

The research methodology used here included a job analysis of supervisors jobs, a test development process of items related to simulation types of items and then the typical research preparatory functions. These included extensive rating scale development activities, true score development on the two rating scales, target script development, verification, and pilot testing, and research item phrasing pilot testing. With the conclusion of these activities, the two sets of tapes were finished.

These tapes are now useful for future research. The development process is the longest portion. The review of the literature shows that the use of videotapes is well published, but very few researchers take the time to develop them and typically borrow tapes from other researchers developed for other settings. None have been put together in as practical a setting as the current research. The current data

and the future research on these tapes can shed a great deal of light on the processes and the dimensions involved in having one person rate the performance of another person.

There are high costs associated with the research that must be taken into account. The current research required nearly four and a half hours of time in a single block from many small groups of participants. Long before the project was well underway, people recruited to participate in the project would call and try to rescind their volunteered participation. Many people felt that it would be more fun to stay on their jobs than participate in the grueling study. It is unlikely that this many videotapes (nine per version) would be useful in future research.

This in itself is an important finding. Typically, raters in a structured oral exam or assessment center are expected to review the performance of this many or more participants in a single session. The raters can do this, and with some degree of accuracy, but they will not enjoy the experience nor recommend it to their colleagues.

These speculations point out the need for more research in this area. The process involved in rating the response of a simulation (or live) performance of a supervisor are far from being totally understood. There is a clear interaction in performance quality between a person's skill in resolving problems and the person's skill in interpersonal relations. The interaction of these seemingly unrelated performance dimensions includes components in addition to those measured here. The largest addition would be a nonverbal behavior or body language variable.

The interaction between the candidates' skills in the performance dimensions and rating accuracy is the main point in the study. Research is typically done with little or no thought given to what is being measured. Most time is spent by the researchers on the factors influencing the accuracy. One of the strong factors pointed out by this study is job relatedness of the rating dimensions. It is entirely possible that much of the debate on rater accuracy could be resolved by focusing clearly on the content validity (i.e., proper instrument development) and on the skill-focused nature of the rating scales used.

Both of the rating scales used in this research were skill based rating scales that drove the development of the stimulus videotapes. These two dimensions are clearly job related and important skills for the candidates to possess if they want to become successful supervisors. This should be the starting point for research into rater accuracy, not the reuse of another researcher's or some company's internal performance appraisal trait rating scale. The focus should be on skills that can be seen and documented (and trained).

Post-hoc analyses were done to assess the correspondence between raw scores and differential accuracy. When compared with the raw scores for the two dimensions, raters with weak differential accuracy were found to be rating candidates lower than they deserved. When the overall assessment rating raw scores were analyzed, the interaction of the interpersonal relations and problem solving rating dimensions was found to be resolved by the raters by averaging the scores of the two

rating scales together. This would cause the lower scores of the persons weak in interpersonal relations to be even lower and the higher scores of the persons strong in interpersonal relations to be even higher. The accuracy problems would be exaggerated.

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Conclusions

The Feldman model was supported. Raters were asked to rate two dimensions for each of nine different videotape candidates. Each candidate was targeted to perform the two dimensions at different combinations of below average, above average, or top quality performance. Respondents were found to be able to rate with high differential accuracy the candidates whose performance on the two rating dimensions were consistent, either both low, both above average or both top quality scores. The candidates with widely divergent skills on the two rating dimensions were not accurately rated. Lowest accuracy on both rating dimensions was for the candidates with high problem solving skills and poor interpersonal relations skills (i.e., the P6I3 target). Future research should carefully consider combining the two rating scales into a single skill dimension.

This combination should be done with great care. The rater's understanding of these two critically important dimensions must take into account the possibility of divergent performance. Some candidates may be highly technically skilled, but weak on interpersonal relations. Others may be strongly skilled interpersonally, but be weak on problem solving. Future research must be done to clarify understanding on the quality level of these types of performance.

On the second hypothesis, both expert and nonexpert raters were found to be able to use the two rating dimensions with equal differen-

tial accuracy. Some of the nonexperts however, were found to have substantial experience that might cause them to be classified as experts.

For the third hypothesis, on differences between rated performance between the two versions of the same target level performance, there were no differences in differential accuracy found. There were differences, however, between versions in the raw scores. These differences may be due to different understandings of the rating dimensions by the actors and actresses involved in the videotapes. These differences between performances of the same scripts by two different people need to be studied in future research. They should also inspire skepticism about results from other studies that have not utilized replications of the stimulus materials.

Suggestions for Improved Rater Accuracy

Accuracy may be increased by careful preparation of observation and rating scale category prototypes

If the performance dimensions do interact in a manner that lowers rater differential accuracy, then this suggests that the performance dimensions of interest should be established and trained with great care. A single picture of a high quality candidate should be formulated. Alternately, clear pictures of each level of performance quality of interest should be formulated. For example, one important level of performance quality for selection research is the passing point. Careful examination of the behaviors of a "just passing" candidate must

be clearly understood by all raters.

Ratings may be accurate on one or a unique combination of multiple prototypes (level of performance) contained within a single or multiple rating dimensions. The other levels of performance may have less well understood category prototypes and, as a consequence, lead to lower differential accuracy. High rater differential accuracy may depend on raters using the exact same category prototypes to store observed performance as are used in the performance rating scale. In particular, candidates with high skills in some areas and low skills in other areas are particularly prone to ratings low in differential accuracy.

Over-training on a single rating dimension may have the largest improvement on rating ac iracy

The simplest path to rater accuracy may be to over-train all raters on a single performance category with a single carefully defined prototype of most desirable behavior. Raters would be trained until they could automatically map observations onto this single category. Only then would they be permitted to make observations and ratings on this same prototype. The degree of overlap of the observed performance with the prototype of most desired performance may be the most accurate measure of the quality of ratee performance.

REFERENCES

Barnes-Farrell J. L. And Courture, K. A. (1984) Effects of appraisal salience on immediate and memory-based judgements (Tech. Rep. No. 84-1). Honolulu: University of Hawaii, Department of Psychology. Cited in Ilgen, D. R., Barnes-Farrell, J. L. and McKellin, D. B. Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986).

Becker and Cardy (1986) Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration, Journal of Applied Psychology, 71, 662-671.

Bernardin, H. J., Cardy, R. L. and Abbott, J. G. (1982) The effects of individual performance schemata, familiarization with the rating scales and rater motivation in rating effectiveness. Cited in Ilgen, D. R., Barnes-Farrell, J. L. and McKellin, D. B. Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986).

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.

Borman, W, C. (1979a). Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 3, 103-115.

Borman, W. C. (1979b). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.

Borman, W. C., Hough, L. M. and Dunnett, M. D. (1978) Performance ratings: An investigation of reliability, accuracy and relationships between individual differences and rater effort. (Army Research Institute Tech Rep. # TR-78-A12)

Brown, F. G. (1976). Principles of educational and psychological testing (2nd ed.). New York: Holt, Rinehart, and Winston.

Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Cardy, R. L. and Dobbins, G. H. (1986) Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. Journal of Applied Psychology, 71, 672-678.

Cardy, R. L. and Kehoe, J. F. (1984) Rater selective ability and appraisal effectiveness: The effect of a cognitive style on the accuracy of differentiation among ratees. Journal of Applied Psychology, 66, 589-594

Cline, V. B. (1964) Interpersonal perception. In Maher, B. A. (Ed) Progress in Experimental Personality Research (Vol. 1), New York: Academic Press.

Cline, V. B. and Richards, J. M., Jr. (1960) Accuracy of interpersonal perception--A general trait? Journal of Abnormal and Social Psychology, 60, 1-7.

Cline, V. B. and Richards, J. M., Jr. (1961) The generality of accuracy of interpersonal perception. Journal of Abnormal and Social Psychology, 62, 446-449.

Cline, V. B. and Richards, J. M., Jr. (1962) Components of accuracy of interpersonal perception scores and the clinical and statistical prediction controversy. Psychological Record, 12, 373-379.

Cline, V. B. and Richards, J. M., Jr. (1963) A methodological note on Cline and Richard's studies of accuracy of interpersonal perception: A reply. Journal of Abnormal and Social Psychology, 66, 195-196.

Constantin, S. W. (1976). An investigation of information favorability in the employment interview. Journal of Applied Psychology, 6, 743-749.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 54, 177-193.

Crow, W. J. (1954) A methodological study of social perceptiveness. Unpublished doctoral dissertation, University of Colorado, Boulder.

DePaulo, B. M., and Pfeiffer, R. L. (1986) On-the-job experience and skill at detecting deception. Journal of Applied Social Psychology, 16, 249-267.

DeNisi, A. S. and Summers, T. P. (1986). Rating forms and the organization of information: A cognitive role for appraisal instruments. Organizational Behavior and Human Decision Making Processes.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgement: Persistence of the illusion of validity. Psychological Review, 85, 395-416.

Farr, J. L. (1973). Response requirements and primacy-recency effects in a simulated selection interview. Journal of Applied Psychology, 57, 228-232.

Favaro, J. L. and Ilgen, D. R. (1983). The effects of ratee characteristics on rater performance appraisal behavior. (Tech. Rep. No. 83-5). East Lansing: Michigan State University, department of Psychology. Cited in Ilgen, D. R., Barnes-Farrell, J. L. and McKellin, D. B. Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986)

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Foti, R. L. and Lord R. G. (1987). Prototypes and scripts: The effects of alternative methods of processing information on rating accuracy. Organizational Behavior and Human Decision Making Processes, 39, 318-340.

Gage, N. L. and Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. Psychological Review, 62, 411-422.

Goldberg L. R. (1959) Simple models or simple processes?: Some research on clinical judgements. American Psychologist, 23, 483-496.

Golden, M. (1964) Some effects of combining psychological tests on clinical Inference. Journal of Consulting Psychology, 28, 440-446.

Ilgen, D. R., Barnes-Farrell and McKillin, D. B. (1987). Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986)

Johnson and Reys (1982) Cited in Ilgen, D. R., Barnes-Farrell, J. L. and McKellin, D. B. Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986)

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol 2). New York: Academic Press.

Jones, E. E., & Nisbett, R. E. (1971). The Actor and the Observer: Divergent perceptions of the causes of behavior. Morristown N.J.: General Learning Press.

Kelley, H. H. (1971a). Attribution in social interaction. In E. E. Jones et al. (Eds.) Attribution: Perceiving the causes of Behavior. Morristown, N. J.: General Learning Press.

Kelley, H. H. (1971b). Causal schemata and the attribution process. In E. E. Jones, et al., (Eds.), Attribution: Perceiving the causes of Behavior. Morristown, N. J.: General Learning Press.

Kelly, G. A. (1955). A theory of personality: The psychology of personal constructs New York: Norton.

Kelly and Fiske (1954) Cited in Wiggins, J. S. (1973) Personality and Prediction: Principles of Personality Assessment. Reading, Mass.: Addison Wesley.

Kostlan (1954) Cited in Wiggins, J. S. (1973) Personality and Prediction: Principles of Personality Assessment. Reading, Mass.: Addison Wesley.

Kozlowski, S. W. J. and Kirsch, M. P. (1987) The systematic distortion hypothesis, halo and accuracy: An individual-level analysis, Journal of Applied Psychology, 72, 252-261.

Kozlowski, S. W. J., Kirsch. M. P. and Chao, G. T. (1986). Job knowledge, ratee familiarity, conceptual similarity and halo error: An exploration. Journal of Applied Psychology, 71, 45-49.

Kremers (1960) Cited in Wiggins, J. S. (1973) Personality and Prediction: Principles of Personality Assessment. Reading, Mass.: Addison Wesley.

Kruglanski, A. W. (1970). Attributing trustworthiness in supervisor-worker relations. Journal of Experimental Social psychology, 6, 214-232.

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

Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.

London, M., & Hakel, M. D. (1974). Effects of applicant stereotypes, order, and information on interview impression. Journal of Applied Psychology, 59, 157-162.

Lord, R. G. (1985) Accuracy in Behavioral Measurement: An alternative definition based on raters' cognitive schema and signal detection theory. Journal of Applied Psychology, 70, 66-71

Meehl, P. (1954). Clinical versus statistical prediction; a theoretical analysis and review of the evidence. Minneapolis: University of Minnesota Press.

Mount, M. K. and Thompson, D. E. (1987) Cognitive categorization and quality of performance ratings. Journal of Applied Psychology, 72, 240-246.

Murphy K. R. and Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. Journal of Applied Psychology, 71, 39-44.

Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W. K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 67, 320-325.

Murphy, K. R., Martin, C., & Garcia, M. (1982). Do behavioral observation scales measure observation? Journal of Applied Psychology, 67, 562-567.

Nathan, R. B. and Alexander, R.A. (1985) the role of inferential accuracy in performance rating. Academy of management Review, 10, 109-115

Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68, 102-114.

Nunnally J. C. (1967). Psychometric theory. New York: McGraw-Hill.

Osburn, H. G., Timmreck, C., & Bigby, D. (1981). Effect of dimensional relevance on accuracy of simulated hiring decisions by employment interviewers. Journal of Applied Psychology, 66, 159-165.

Ostroff, C. and Ilgen, D. R. (1985a). The effect of rater training on rater's accuracy and cognitive categories (Tech Rep No. 85-5). East Lansing: Michigan State University, Department of Psychology and the Department of Management. Cited in Ilgen, D. R., Barnes-Farrell and McKillin, D. B. (1987) Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986)

Ostroff, C. and Ilgen, D. R. (1985b). The relationship between cognitive categories of raters and rating accuracy (Tech Rep No. 85-4). East Lansing: Michigan State University, Department of Psychology and the Department of Management. Cited in Ilgen, D. R., Barnes-Farrell and McKillin, D. B. (1987) Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986)

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.

Pulakos, E. D. (1986). Development of Training Programs to Increase Accuracy with Different Rating Tasks. Organizational Behavior and Human Decision Making Processes, 38, 76-91.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in Experimental Social psychology (Vol 10). New York: Academic Press.

Schmitt, N. A. and Lappin, M. (1982). Race and sex as determinants of the mean and variance of performance ratings. Journal of Applied Psychology, 65, 428-435.

Schneider, E. C., & Shiffren, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 84, 1-66.

Shiffren, R. M., & Schneider, E. C., (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Sines (1959) Cited in Wiggins, J. S. (1973) Personality and Prediction: Principles of Personality Assessment. Reading, Mass.: Addison Wesley.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Springbett, B. M. (1958). Factors affecting the final decision in the employment interview. Canadian Journal of Psychology, 12, 13-22.

Taft (1955) The ability to judge people. Psychological Bulletin, 52, 1-23

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A., & Kahneman, D. (1971). Belief in the laws of small numbers. Psychological Bulletin, 76, 105-110.

Valenzi, E., & Andrews, I. R. (1973). Individual differences in the decision process of employment interviewers. Journal of Applied Psychology, 58, 49-53.

Wiggins, J. S. (1973) Personality and Prediction: Principles of Personality Assessment. Reading, Mass.: Addison Wesley.

Winer, B. J. (1971) Statistical Principles in Experimental Design (2nd Ed). New York: McGraw-Hill

Williams, K. J., DeNisi, A. S., Blencoe, A. G., and Cafferty, T. P. (1985). The role of appraisal purpose: Effects of purpose in information acquisition and utilization. Journal of Applied Psychology, 35, 314-339.

Yountz, M. A., And Ilgen, D. R. (1986). The impact of performance level on alternative measures of rater accuracy (Tech. Rep. No. 86-1). East Lansing: Michigan State University, Department of psychology. Cited in Ilgen, D. R., Barnes-Farrell and McKillin, D. B. (1987) Performance Rating Accuracy. (Paper presented at the 21st International congress of Applied Psychology, Jerusalem, Israel, July, 1986)

Zedeck, S. and Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67, 752-758.

APPENDIX A

ANOVA and ANCOVA Summary Tables

In the ANOVA summary tables presented below, "P" is the Problem Solving Target Score, "I" is the Interpersonal Relations Target Score, "E" is the Expertise of the Rater, and "V" is the Version of the Videotapes. The analyses are performed using SAS procedure GLM and BMDP programs P2V and P4V. *F* statistics listed are all for unique variance accounted for (i.e., SAS "type III" sums of squares or BMDP within contrast pool (WCP) mean squares). In most cases in the analyses this provided more conservative results than even the multivariate method of computing the within portion of the design. The SAS procedure GLM was used to make the initial contrasts (using the Helmert contrast option) on the significant repeated measures factors. All simple effects and individual cell comparisons were done using planned contrasts in BMDP4V. Due to the large number of significance tests conducted in the study, a conservative significance level of .01 was chosen.

The mean ratings and the standard deviations for each cell in each analysis are presented below. In each matrix of numbers the top number is the mean and the bottom the standard deviation of the combination of factors. Note that "E" represents the Experts, "N" is Non-Experts, "V1" is Version one, "V2" is version two, "P" and "I" are the

problem solving and interpersonal relations target scores. If any respondent had a missing value for an item, the case was deleted from the analysis.

Table 17: Problem Solving Raw Scores Repeated Measures ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
P	325.3	2,142	<.0001
P*E	.5	2,142	ns
P*V	1.1	2,142	ns
P*E*V	.1	2,142	ns
I	223.2	2,142	<.0001
I*E	1.2	2,142	ns
I*V	11.7	2,142	<.0001
I*E*V	1.0	2,142	ns
P*I	3.1	4,284	ns
P*I*E	.5	4,284	ns
P*I*V	2.1	4,284	ns
P*I*E*V	.02	4,284	ns
E	4.24	1,71	ns
V	.3	1,71	ns
E*V	1.0	1,71	ns

Table 18: Problem Solving Raw Scores									
Mean Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E V1	1.96	2.71	3.47	3.56	4.36	5.67	3.80	5.52	5.90
n=17	.54	1.00	1.68	1.15	1.24	.93	1.11	1.08	.70
E V2	2.05	2.44	4.05	4.23	3.87	5.88	4.30	4.73	5.96
n=18	.46	.98	1.62	1.11	.73	.84	1.13	1.28	.78
N V1	2.50	3.02	3.95	4.20	4.95	5.99	4.40	5.60	6.22
n=20	.88	.80	1.55	1.14	1.14	.94	1.01	.91	.79
N V2	2.30	2.84	4.13	4.46	4.32	5.68	4.44	4.79	5.87
n=20	.95	.97	1.16	1.10	1.27	.70	1.12	1.16	.59

Table 19: Interpersonal Relations Raw Scores Repeated Measures ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
P	141.49	2,142	<.0001
P*E	.15	2,142	ns
P*V	2.07	2,142	ns
P*E*V	.84	2,142	ns
I	290.70	2,142	<.0001
I*E	3.21	2,142	ns
I*V	11.16	2,142	<.0001
I*E*V	1.36	2,142	ns
P*I	7.88	4,284	<.0001
P*I*E	1.60	4,284	ns
P*I*V	6.25	4,284	<.0001
P*I*E*V	.78	4,284	ns
E	2.70	1,71	ns
V	.35	1,71	ns
E*V	1.08	1,71	ns

Table 20: Interpersonal Relations Raw Scores									
Means Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E V1	2.09 0.97	3.71 1.16	3.98 1.22	3.90 1.04	4.64 1.16	6.10 0.72	3.14 1.01	5.48 0.83	6.17 0.59
E V2	1.96 0.75	3.51 1.29	5.11 1.57	4.30 1.38	3.91 0.98	6.31 1.08	3.56 1.24	4.71 1.36	6.36 0.81
N V1	2.86 1.08	3.61 0.97	4.67 1.52	3.93 1.26	5.02 1.00	6.25 0.96	4.00 1.28	5.73 1.86	6.36 0.67
N V2	2.17 0.88	3.71 1.18	5.03 1.33	4.79 1.43	4.39 1.25	5.92 0.82	4.14 1.14	4.36 1.19	5.95 0.90

Table 21: Problem Solving Differential Accuracy ANOVA Summary Table			
Effect	F	Degrees of Freedom	Alpha probability
P	1.22	2,142	ns
P*E	.06	2,142	ns
P*V	.35	2,142	ns
P*E*V	.19	2,142	ns
I	12.25	2,142	<.0001
I*E	1.47	2,142	ns
I*V	.86	2,142	ns
I*E*V	.38	2,142	ns
P*I	5.43	4,284	<.0004
P*I*E	2.21	4,284	ns
P*I*V	.52	4,284	ns
P*I*E*V	.90	4,284	ns
E	.03	1,71	ns
V	.66	1,71	ns
E*V	.16	1,71	ns

Table 22: Problem Solving Differential Accuracy									
Mean Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E V1	0.57 0.31	0.63 0.47	0.57 0.50	0.55 0.44	0.80 0.66	0.35 0.29	0.73 0.38	0.62 0.60	0.47 0.29
E V2	0.57 0.35	0.57 0.42	0.64 0.52	0.79 0.58	0.72 0.55	0.53 0.32	0.84 0.47	0.60 0.48	0.41 0.22
N V1	0.52 0.41	0.66 0.57	0.41 0.28	0.71 0.53	0.45 0.32	0.48 0.40	0.82 0.69	0.60 0.67	0.37 0.23
N V2	0.53 0.43	0.65 0.71	0.60 0.32	0.69 0.44	0.62 0.37	0.52 0.39	1.03 0.65	0.52 0.51	0.44 0.26

Table 23: Interpersonal Relations Differential Accuracy ANOVA Summary Table			
Effect	F	Degrees of Freedom	Alpha probability
P	9.18	2,142	<.0002
P*E	.00	2,142	ns
P*V	.10	2,142	ns
P*E*V	1.63	2,142	ns
I	5.27	2,142	<.0062
I*E	2.77	2,142	ns
I*V	.33	2,142	ns
I*E*V	2.31	2,142	ns
P*I	19.37	4,284	<.0001
P*I*E	.90	4,284	ns
P*I*V	1.52	4,284	ns
P*I*E*V	.23	4,284	ns
E	2.05	1,71	ns
V	2.59	1,71	ns
E*V	.10	1,71	ns

Table 24: Interpersonal Relations Differential Accuracy									
Means Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E V1	0.49 0.24	0.82 0.80	0.54 0.37	0.52 0.35	0.65 1.08	0.29 0.29	0.87 0.86	0.47 0.33	0.37 0.31
E V2	0.41 0.34	0.92 0.65	0.95 1.11	0.37 0.25	0.55 0.39	0.44 0.53	1.36 1.13	0.58 0.52	0.42 0.32
N V1	0.48 0.41	0.79 0.85	1.05 0.88	0.72 0.72	0.41 0.32	0.36 0.25	1.09 0.71	0.50 0.41	0.58 0.81
N V2	0.53 0.47	0.97 0.94	1.07 1.20	0.95 0.95	0.74 0.74	0.42 0.44	1.48 0.98	0.53 0.36	0.53 0.78

Table 25: Problem Solving Confidence Rating ANOVA Summary Table			
Effect	F	Degrees of Freedom	Alpha probability
P	2.33	2,134	ns
P*E	1.64	2,134	ns
P*V	.01	2,134	ns
P*E*V	.01	2,134	ns
I	2.06	2,134	ns
I*E	2.17	2,134	ns
I*V	.52	2,134	ns
I*E*V	.51	2,134	ns
P*I	13.96	4,268	<.0001
P*I*E	2.07	4,268	ns
P*I*V	3.54	4,268	<.0078
P*I*E*V	1.07	4,268	ns
E	.44	1,67	ns
V	2.85	1,67	ns
E*V	.20	1,67	ns

Table 26: Problem Solving Rating Scale Confidence Ratings									
Means Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E, V1	5.81 0.83	5.56 0.96	5.37 0.81	5.06 1.18	5.37 0.88	5.37 0.96	5.19 1.05	5.44 1.03	5.37 0.89
E, V2	5.76 1.30	5.47 1.17	4.76 1.15	4.82 1.33	4.71 1.49	5.53 1.12	4.53 1.37	5.41 1.28	5.41 1.37
N, V1	6.10 1.12	5.00 0.97	5.80 0.89	5.30 1.22	5.55 1.15	6.00 1.08	5.55 1.05	5.55 0.94	5.65 1.04
N, V2	5.78 1.11	5.06 1.47	4.83 1.25	4.67 1.33	5.33 1.37	5.61 1.09	4.89 1.49	5.00 1.03	5.61 1.04

Table 27: Interpersonal Relations Confidence Rating ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
P	2.68	2,138	ns
P*E	.74	2,138	ns
P*V	.24	2,138	ns
P*E*V	.44	2,138	ns
I	11.21	2,138	<.0001
I*E	.62	2,138	ns
I*V	2.87	2,138	ns
I*E*V	2.89	2,138	ns
P*I	10.85	4,276	<.0001
P*I*E	1.04	4,276	ns
P*I*V	1.27	4,276	ns
P*I*E*V	1.43	4,276	ns
E	.77	1,69	ns
V	1.57	1,69	ns
E*V	.16	1,69	ns

Table 28: Interpersonal Relations Confidence Ratings									
Means Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E, V1	5.29 1.16	5.06 1.34	4.82 1.01	4.71 1.31	5.53 0.87	5.53 0.80	5.06 1.03	5.47 1.18	5.71 0.77
E, V2	5.41 1.73	4.59 1.70	4.82 1.29	4.88 1.41	4.47 1.18	5.82 1.29	5.00 1.32	5.00 1.54	5.65 1.45
N, V1	5.74 1.19	5.00 1.20	5.63 1.01	5.05 1.22	5.47 1.07	5.89 1.10	5.68 0.95	5.32 0.95	5.68 1.00
N, V2	5.75 1.12	4.55 1.50	4.80 1.43	4.95 1.43	5.40 1.31	5.60 1.14	4.85 1.60	4.85 1.27	5.75 1.16

Table 29: Problem Solving Differential Accuracy ANCOVA Raw Scores

Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.32	0.63	0.82	0.42	0.41	0.66	0.43	0.57	0.44
N,V2	0.44	0.53	0.95	0.52	0.59	0.72	0.61	0.63	0.52
E,V1	0.50	0.66	0.73	0.38	0.85	0.47	0.58	0.57	0.59
E,V2	0.42	0.62	0.84	0.50	0.71	0.75	0.67	0.55	0.52

Table 30: Problem Solving Differential Accuracy Ability ANCOVA Means

Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.44	0.57	0.43	0.66	0.41	0.42	0.82	0.63	0.33
N,V2	0.52	0.63	0.62	0.72	0.58	0.53	0.95	0.53	0.45
E,V1	0.59	0.56	0.57	0.47	0.84	0.38	0.73	0.65	0.50
E,V2	0.52	0.54	0.67	0.75	0.70	0.51	0.84	0.62	0.43

Table 31: PS Differential Accuracy Ability ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	.46	1,47	ns
V	1.50	1,47	ns
E*V	.15	1,47	ns
Cov(E*V)	1.13	1,47	ns
P	.69	2,95	ns
P*E	1.06	2,95	ns
P*V	.17	2,95	ns
P*E*V	.19	2,95	ns
Cov(P*E*V)	.43	1,95	ns
I	13.96	2,95	<.005
I*E	1.24	2,95	ns
I*V	.47	2,95	ns
I*E*V	.19	2,95	ns
Cov(I*E*V)	4.28	1,95	ns
P*I	5.62	4,191	<.005
P*I*E	.93	4,191	ns
P*I*V	.11	4,191	ns
P*I*E*V	.79	4,191	ns
Cov(P*I*E*V)	.59	1,191	ns

Table 32: Problem Solving Differential Accuracy Effort Adjusted Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.44	0.59	0.43	0.67	0.43	0.43	0.85	0.54	0.31
N,V2	0.44	0.48	0.64	0.72	0.54	0.48	0.96	0.46	0.44
E,V1	0.61	0.60	0.58	0.48	0.79	0.39	0.72	0.69	0.46
E,V2	0.52	0.58	0.69	0.75	0.75	0.52	0.85	0.65	0.43

Table 33: PS DA Effort ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	.46	1,47	ns
V	1.90	1,47	ns
E*V	.08	1,47	ns
Cov(E*V)	.01	1,47	ns
P	.24	2,95	ns
P*E	.81	2,95	ns
P*V	.02	2,95	ns
P*E*V	.11	2,95	ns
Cov(P*E*V)	1.50	1,95	ns
I	16.52	2,95	<.0001
I*E	1.37	2,95	ns
I*V	.60	2,95	ns
I*E*V	.16	2,95	ns
Cov(I*E*V)	1.75	1,95	ns
P*I	5.37	4,267	<.0001
P*I*E	1.08	4,191	ns
P*I*V	.71	4,191	ns
P*I*E*V	.77	4,191	ns
Cov(P*I*E*V)	.04	1,191	ns

Table 34: Problem Solving Differential Accuracy Luck Adjusted Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.55	0.58	0.43	0.72	0.45	0.45	0.83	0.62	0.34
N,V2	0.52	0.47	0.61	0.68	0.54	0.46	0.92	0.45	0.43
E,V1	0.61	0.59	0.56	0.47	0.79	0.39	0.72	0.69	0.46
E,V2	0.52	0.55	0.67	0.75	0.70	0.50	0.84	0.62	0.42

Table 35: PS DA Luck ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	.32	1,47	ns
V	1.30	1,47	ns
E*V	.01	1,47	ns
Cov(E*V)	.63	1,47	ns
P	.91	2,95	ns
P*E	1.28	2,95	ns
P*V	.14	2,95	ns
P*E*V	.13	2,95	ns
Cov(P*E*V)	2.59	1,95	ns
I	11.24	2,95	<.0000
I*E	2.71	2,95	ns
I*V	1.66	2,95	ns
I*E*V	.30	2,95	ns
Cov(I*E*V)	.03	1,95	ns
P*I	5.26	4,191	<.0001
P*I*E	1.04	4,191	ns
P*I*V	.79	4,191	ns
P*I*E*V	.73	4,191	ns
Cov(P*I*E*V)	1.14	1,191	ns

Table 36: Problem Solving Differential Accuracy Task Difficulty Adjusted Means

Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.53	0.57	0.43	0.71	0.45	0.45	0.82	0.61	0.33
N,V2	0.53	0.47	0.62	0.69	0.54	0.46	0.92	0.45	0.44
E,V1	0.61	0.60	0.57	0.48	0.79	0.40	0.73	0.69	0.46
E,V2	0.52	0.55	0.67	0.75	0.71	0.50	0.84	0.62	0.42

Table 37: PS DA Task Difficulty ANCOVA Summary

Effect	F	Degrees of freedom	Alpha probability
E	.36	1,47	ns
V	1.80	1,47	ns
E*V	.04	1,47	ns
Cov(E*V)	.31	1,47	ns
P	.53	2,95	ns
P*E	1.23	2,95	ns
P*V	.15	2,95	ns
P*E*V	.06	2,95	ns
Cov(P*E*V)	2.42	1,95	ns
I	16.11	2,95	<.0000
I*E	1.70	2,95	ns
I*V	.52	2,95	ns
I*E*V	.18	2,95	ns
Cov(I*E*V)	1.56	1,95	ns
P*I	5.46	4,191	<.0008
P*I*E	1.07	4,191	ns
P*I*V	.71	4,191	ns
P*I*E*V	.75	4,191	ns
Cov(P*I*E*V)	.00	1,191	ns

Table 38: Interpersonal Relations Differential Accuracy ANCOVA Raw Scores									
Raw Scores	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.48	0.80	0.52	0.52	0.63	0.25	0.73	0.46	0.32
N,V2	0.42	0.95	0.95	0.40	0.55	0.46	1.47	0.59	0.41
E,V1	0.48	0.64	1.03	0.55	0.44	0.38	1.14	0.51	0.51
E,V2	0.46	0.96	1.11	0.96	0.77	0.46	1.54	0.56	0.54

Table 39: Interpersonal Relations Differential Accuracy Ability ANCOVA Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.49	0.81	0.55	0.51	0.70	0.25	0.76	0.48	0.35
N,V2	0.42	0.95	0.99	0.39	0.56	0.48	1.41	0.59	0.44
E,V1	0.47	0.67	1.03	0.57	0.42	0.40	1.09	0.49	0.66
E,V2	0.49	0.97	1.18	0.92	0.76	0.47	1.50	0.53	0.56

Table 40: Interpersonal Relations Differential Ability Adjusted ANCOVA			
Effect	F	Degrees of Freedom	Alpha probability
E	2.69	1,47	ns
V	5.21	1,47	ns
E*V	.43	1,47	ns
Cov(E*V)	.00	1,47	ns
P	3.89	2,95	ns
P*E	.06	2,95	ns
P*V	.97	2,95	ns
P*E*V	2.11	2,95	ns
Cov(P*E*V)	3.56	1,95	ns
I	3.91	2,95	ns
I*E	3.57	2,95	ns
I*V	.03	2,95	ns
I*E*V	.72	2,95	ns
Cov(I*E*V)	3.92	1,95	ns
P*I	13.83	4,191	<.0000
P*I*E	.71	4,191	ns
P*I*V	3.23	4,191	<.0136, ns
P*I*E*V	.81	4,191	ns
Cov(P*I*E*V)	.00	1,191	ns

Table 41: Interpersonal Relations Differential Accuracy Effort Adjusted ANCOVA			
Effect	F	Degrees of Freedom	Alpha probability
E	3.26	1,47	ns
V	6.06	1,47	ns
E*V	.22	1,47	ns
Cov(E*V)	1.96	1,47	ns
P	3.62	2,95	ns
P*E	.05	2,95	ns
P*V	1.14	2,95	ns
P*E*V	2.07	2,95	ns
Cov(P*E*V)	.03	1,95	ns
I	4.87	2,95	<.0097
I*E	2.89	2,95	ns
I*V	.01	2,95	ns
I*E*V	.67	2,95	ns
Cov(I*E*V)	.02	1,95	ns
P*I	14.30	4,191	<.0000
P*I*E	.84	4,191	ns
P*I*V	3.20	4,191	<.0143, ns
P*I*E*V	.50	4,191	ns
Cov(P*I*E*V)	.20	1,191	ns

Table 42: Interpersonal Relations Differential Accuracy Effort Adjusted Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.49	0.85	0.49	0.53	0.73	0.23	0.77	0.48	0.32
N,V2	0.34	0.96	0.99	0.39	0.53	0.38	1.37	0.52	0.37
E,V1	0.50	0.72	1.10	0.58	0.47	0.39	1.14	0.47	0.65
E,V2	0.50	1.06	1.25	0.95	0.79	0.47	1.59	0.56	0.56

Table 43: Interpersonal Relations Differential Accuracy Luck Adjusted Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.49	0.82	0.54	0.50	0.68	0.24	0.84	0.50	0.33
N,V2	0.35	0.94	0.98	0.38	0.51	0.37	1.32	0.51	0.37
E,V1	0.50	0.72	1.10	0.58	0.47	0.39	1.14	0.47	0.65
E,V2	0.48	1.01	1.18	0.93	0.77	0.45	1.51	0.54	0.53

Table 44: Interpersonal Relations Differential Luck Adjusted ANCOVA			
Effect	F	Degrees of Freedom	Alpha probability
E	2.65	1,47	ns
V	5.01	1,47	ns
E*V	.40	1,47	ns
Cov(E*V)	.00	1,47	ns
P	3.06	2,95	ns
P*E	.06	2,95	ns
P*V	1.22	2,95	ns
P*E*V	2.05	2,95	ns
Cov(P*E*V)	.71	1,95	ns
I	5.48	2,95	<.0056
I*E	2.94	2,95	ns
I*V	.02	2,95	ns
I*E*V	2.09	2,95	ns
Cov(I*E*V)	.31	1,95	ns
P*I	13.97	4,191	<.0000
P*I*E	.70	4,191	ns
P*I*V	3.07	4,191	ns
P*I*E*V	.80	4,191	ns
Cov(P*I*E*V)	.36	1,191	ns

Table 45: Interpersonal Relations Differential Accuracy Task Difficulty Adjusted Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	0.49	0.82	0.54	0.50	0.66	0.24	0.84	0.50	0.33
N,V2	0.36	0.94	0.98	0.38	0.51	0.37	1.32	0.51	0.37
E,V1	0.50	0.72	1.10	0.58	0.47	0.39	1.14	0.47	0.65
E,V2	0.48	1.01	1.18	0.93	0.77	0.45	1.51	0.54	0.54

Table 46: Interpersonal Relations Differential Task Difficulty Adjusted ANCOVA			
Effect	F	Degrees of Freedom	Alpha probability
E	2.37	1,47	ns
V	5.13	1,47	ns
E*V	.54	1,47	ns
Cov(E*V)	.54	1,47	ns
P	3.60	2,95	ns
P*E	.07	2,95	ns
P*V	1.20	2,95	ns
P*E*V	1.90	2,95	ns
Cov(P*E*V)	3.13	1,95	ns
I	5.26	2,95	<.0068
I*E	3.01	2,95	ns
I*V	.02	2,95	ns
I*E*V	.66	2,95	ns
Cov(I*E*V)	.26	1,95	ns
P*I	13.82	4,191	<.0000
P*I*E	.70	4,191	ns
P*I*V	3.20	4,191	<.0143, ns
P*I*E*V	.52	4,191	ns
Cov(P*I*E*V)	.38	1,191	ns

Table 47: Ability ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha Probability
E	1.06	1,67	ns
V	.05	1,67	ns
E*V	.00	1,67	ns
P	6.48	2,134	<.0021
P*E	.30	2,134	ns
P*V	1.45	2,134	ns
P*E*V	.03	2,134	ns
I	12.06	2,134	<.0001
I*E	.91	2,134	ns
I*V	.23	2,134	ns
I*E*V	.05	2,134	ns
P*I	2.06	4,268	ns
P*I*E	1.08	4,268	ns
P*I*V	1.17	4,268	ns
P*I*E*V	.93	4,268	ns

Table 48: Attribution Means									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
Ability	4.58	4.44	4.71	4.77	4.60	5.44	4.81	4.78	5.58
Effort	4.13	4.17	4.45	4.44	4.44	4.73	4.29	4.52	4.74
Task Dif	2.74	3.04	2.93	3.03	2.86	3.01	3.92	3.07	3.04
Luck	2.27	2.21	2.01	2.07	2.01	2.13	2.30	2.20	2.10

Table 49: Effort ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha Probability
E	.00	1,68	ns
V	.14	1,68	ns
E*V	1.28	1,68	ns
P	9.12	2,136	<.0002
P*E	1.96	2,136	ns
P*V	2.27	2,136	ns
P*E*V	1.81	2,136	ns
I	8.14	2,136	<.0005
I*E	.26	2,136	ns
I*V	1.02	2,136	ns
I*E*V	.22	2,136	ns
P*I	.62	4,272	ns
P*I*E	1.98	4,272	ns
P*I*V	.42	4,272	ns
P*I*E*V	.89	4,272	ns

Table 50: Task Difficulty ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha Probability
E	3.14	1,66	ns
V	.33	1,66	ns
E*V	.05	1,66	ns
P	4.05	2,132	ns
P*E	.48	2,132	ns
P*V	.32	2,132	ns
P*E*V	1.10	2,132	ns
I	.24	2,132	ns
I*E	1.18	2,132	ns
I*V	.74	2,132	ns
I*E*V	.02	2,132	ns
P*I	3.02	4,264	ns
P*I*E	1.35	4,264	ns
P*I*V	1.37	4,264	ns
P*I*E*V	.69	4,264	ns

Table 51: Luck ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha Probability
E	2.30	1,66	ns
V	.68	1,66	ns
E*V	2.84	1,66	ns
P	2.53	2,132	ns
P*E	.03	2,132	ns
P*V	.98	2,132	ns
P*E*V	.38	2,132	ns
I	1.68	2,132	ns
I*E	.28	2,132	ns
I*V	2.22	2,132	ns
I*E*V	.82	2,132	ns
P*I	1.38	4,264	ns
P*I*E	1.61	4,264	ns
P*I*V	2.23	4,264	ns
P*I*E*V	.86	4,264	ns

Table 52 Overall Assessment Raw Scores ANOVA Summary			
Effect	F	Degrees of Freedom	Alpha Probability
P	197.20	2,136	<.0001
P*E	.05	2,136	ns
P*V	1.01	2,136	ns
P*E*V	.12	2,136	ns
I	204.90	2,136	<.0001
I*E	3.44	2,136	ns
I*V	13.40	2,136	<.0001
I*E*V	.99	2,136	ns
P*I	2.40	4,272	ns
P*I*E	.31	4,272	ns
P*I*V	4.79	4,272	<.0015
P*I*E*V	.09	4,272	ns
E	3.58	1,68	ns
V	0.00	1,68	ns
E*V	1.90	1,68	ns

Table 53: Overall Assessment Raw Scores Item Statistics

Means Std	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
E V1	1.87 0.20	2.87 0.31	3.75 0.37	3.81 0.30	4.75 0.30	5.75 0.21	3.25 0.31	5.56 0.22	6.00 0.20
E V2	1.65 0.12	3.12 0.21	4.88 0.30	4.29 0.36	4.23 0.25	6.23 0.25	4.12 0.35	4.59 0.38	6.18 0.21
N V1	2.53 0.23	3.21 0.26	4.32 0.39	4.16 0.27	5.16 0.23	6.05 0.19	4.00 0.30	5.84 0.24	6.21 .21
N V2	2.25 0.34	2.95 .025	4.60 0.30	4.70 0.33	4.45 0.34	5.90 0.20	4.65 0.27	4.45 0.31	5.95 0.21

Table 54: OAS Raw Scores Corrected for Ability Attributions

Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	2.74	3.28	4.47	4.35	5.27	6.09	3.85	5.67	6.13
N,V2	2.26	2.87	4.65	4.74	4.20	5.66	4.47	4.55	5.85
E,V1	1.90	2.74	3.80	3.82	4.99	5.51	2.72	5.59	5.79
E,V2	1.66	3.24	4.89	4.47	4.37	6.16	3.78	4.85	6.10

Table 55: OAS Raw Scores Corrected for Ability Attributions ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	2.26	1,47	ns
V	.02	1,47	ns
E*V	2.42	1,47	ns
Cov(E*V)	1.12	1,47	ns
P	121.63	2,95	<.0000
P*E	.25	2,95	ns
P*V	.26	2,95	ns
P*E*V	.09	2,95	ns
Cov(P*E*V)	1.89	1,95	ns
I	117.88	2,95	<.0000
I*E	4.31	2,95	ns
I*V	7.96	2,95	<.0006
I*E*V	.53	2,95	ns
Cov(I*E*V)	5.35	1,95	ns
P*I	3.68	4,191	<.0065
P*I*E	.32	4,191	ns
P*I*V	3.23	4,191	ns
P*I*E*V	.15	4,191	ns
Cov(P*I*E*V)	6.07	1,191	ns

Table 56: OAS Raw Scores Corrected for Effort Attributions									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	2.76	3.11	4.65	4.32	5.38	6.21	3.85	5.79	6.08
N,V2	2.24	2.90	4.65	4.72	4.14	5.70	4.47	4.52	5.80
E,V1	2.00	2.75	3.86	3.84	4.99	5.59	2.85	5.57	5.85
E,V2	1.65	3.07	4.83	4.34	4.30	6.19	3.84	4.78	6.12

Table 57: OAS Raw Scores Corrected for Effort Attributions ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	2.30	1,48	ns
V	.11	1,48	ns
E*V	2.20	1,48	ns
Cov(E*V)	.01	1,48	ns
P	124.03	2,97	<.0000
P*E	.22	2,97	ns
P*V	.29	2,97	ns
P*E*V	.06	2,97	ns
Cov(P*E*V)	.05	1,97	ns
I	117.79	2,97	<.0000
I*E	3.23	2,97	ns
I*V	8.00	2,97	<.0006
I*E*V	.85	2,97	ns
Cov(I*E*V)	.71	1,97	ns
P*I	2.92	4,195	<.0065
P*I*E	.13	4,195	ns
P*I*V	3.39	4,195	<.0104, ns
P*I*E*V	.09	4,195	ns
Cov(P*I*E*V)	5.84	1,195	ns

Table 58: OAS Raw Scores Corrected for Luck Attributions									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	2.66	3.04	4.62	4.24	5.34	6.19	3.84	5.74	6.04
N,V2	2.17	2.95	4.74	4.67	4.24	5.80	4.57	4.44	5.88
E,V1	2.00	2.72	3.84	3.84	5.00	5.65	2.82	5.64	5.92
E,V2	1.63	3.08	4.85	4.41	4.33	6.23	3.77	4.78	6.15

Table 59: OAS Raw Scores Corrected for Luck Attributions ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	2.52	1,47	ns
V	.17	1,47	ns
E*V	2.33	1,47	ns
Cov(E*V)	.26	1,47	ns
P	132.53	2,95	<.0000
P*E	.22	2,95	ns
P*V	.28	2,95	ns
P*E*V	.07	2,95	ns
Cov(P*E*V)	.00	1,95	ns
I	142.41	2,95	<.0000
I*E	2.81	2,95	ns
I*V	7.42	2,95	<.0010
I*E*V	.68	2,95	ns
Cov(I*E*V)	4.68	1,95	ns
P*I	2.90	4,191	ns
P*I*E	.56	4,191	ns
P*I*V	3.69	4,191	<.0064
P*I*E*V	.08	4,191	ns
Cov(P*I*E*V)	.84	1,191	ns

Table 60: OAS Raw Scores Corrected for Task Difficulty Attributions									
Means	P3 I3	P3 I5	P3 I6	P5 I3	P5 I5	P5 I6	P6 I3	P6 I5	P6 I6
N,V1	2.67	2.96	4.59	4.20	5.34	6.18	3.83	5.67	6.03
N,V2	2.18	2.94	4.77	4.68	4.23	5.80	4.52	4.43	5.85
E,V1	2.09	2.77	3.86	3.87	5.06	5.66	2.83	5.65	5.90
E,V2	1.66	3.09	4.85	4.42	4.34	6.23	3.75	4.82	6.15

Table 61: OAS Raw Scores Corrected for Task Difficulty Attributions ANCOVA Summary			
Effect	F	Degrees of Freedom	Alpha probability
E	2.22	1,47	ns
V	.08	1,47	ns
E*V	1.99	1,47	ns
Cov(E*V)	.01	1,47	ns
P	134.22	2,95	<.0000
P*E	.25	2,95	ns
P*V	.26	2,95	ns
P*E*V	.13	2,95	ns
Cov(P*E*V)	1.00	1,95	ns
I	146.21	2,95	<.0000
I*E	3.32	2,95	ns
I*V	7.70	2,95	<.0008
I*E*V	.85	2,95	ns
Cov(I*E*V)	11.61	1,95	<.0010
P*I	3.09	4,191	ns
P*I*E	.71	4,191	ns
P*I*V	3.95	4,191	ns
P*I*E*V	.09	4,191	ns
Cov(P*I*E*V)	9.89	1,191	<.0019

APPENDIX B

APPENDIX B

INTERPERSONAL RELATIONS

Self expression is defined as disclosing or communicating one's point of view. This includes expressing the point of view of the department. The information content of the self expression involves expressing information concretely and directing it toward one's goals for the conversation. The top scoring people will state the information succinctly and actively, using the word "I." Stating things actively means owning the information personally, rather than trying to place the words in someone else's control. The information should be goal directed. The average quality candidates will state their information in ways that are understandable in most cases, but not succinctly and not actively. The below acceptable people will be vague in their expression of the information, whether stated actively or passively.

The emotional content of the self expression is stating one's own positive and negative emotions. Top quality candidates will express their positive and negative emotions genuinely, constructively and immediately. Average quality individuals express their emotions, according to the model, in the same way as top quality people, but not as consistently. Below acceptable candidates express their emotions either too aggressively or too passively.

The delivery of the information and feelings should be appropriate in depth and amount to the goals of the conversation, gradually increased and friendly and shared. The average candidates will deliver their information assertively and gradually increased in intensity, but not always in a friendly and shared manner. The below average candidates will deliver their information inappropriately, either in depth or amount or duration. The information may not be relevant to the target person or situation. In some way the disclosure will be detracting or distracting from the ongoing nature of the relationship.

The response with empathy is communicating accurate understanding of the core meaning of the other person's words. Responding with empathy serves two purposes: making sure that the other person is understood and giving the other person the feeling of being understood. The information content should be expressed succinctly and actively. In all cases, for all four components of the interpersonal relations model, the rating scales for information content will always be the same.

The emotional content of the response with empathy is expressing the core meaning of the other person's words. This is done by stating succinctly the correct type and intensity level of the other person's feelings. The average quality level response with empathy expresses the correct type of the other person's emotion, but the expressed

intensity level may be a bit off target. The unacceptable level of emotional content feedback would consist of the candidate being grossly off target in expressing the emotional intensity level or misidentifying the type of emotion.

The delivery of the response with empathy should be appropriate in depth and amount to the feelings and information expressed. Top quality candidates will deliver the feedback clearly and succinctly, without rambling using an assertive, but not aggressive or passive delivery. The top quality person uses good metaphor and humor to feed back the other person's words and feelings, and does not parrot the exact same words. The average respondent will follow the lead of the top quality people, but may be a bit disruptive at times. Some parroting may be done. The below average person will deliver the response in a way that is disruptive to the goals of the conversation. The feedback will become a center of attention, rather than a tool to promote the ongoing problem solving discussion.

The probes and questions begin the steps designed to reach the major purpose of the problem solving discussion. The information gathering should once again be done with clear and succinct questions and probes. The emotional content of the questions and probes should be carefully controlled by asking open-ended, rather than, closed-ended, one word answer questions. Top quality candidates will ask

only open-ended questions. Average quality respondents will have a mixture of one word answer and open ended items. The below acceptable quality candidates will use predominantly closed-ended, one word answer, yes, no type items. The lower the quality of the questioning, generally the less clear are the items or probes.

The delivery of the probes and questions should be done in a way that facilitates or supplements the discussion and should not dominate the interaction. Average level candidates may dominate the discussion at times with questioning, but the items will usually facilitate the interaction. The items may not always be friendly. The below average candidates will sometimes be either aggressive or passive in their information gathering, sometimes dominating the discussion with questions that are clearly not friendly.

The challenging subcomponent of the interpersonal relations model is done only after establishing a positive relationship with the subordinate. The intention is to provide feedback to the person on past performance and seek to motivate them to improve future performance. This can be done in a variety of ways. The supervisor can describe the person's actions inconsistent with standards, or state implications or conclusions that were unrecognized by the subordinate, or confront the person to examine their actions more closely.

The information content should once again be stated succinctly and actively. The challenges themselves should be behaviorally centered on observable or reasonably inferred facts and should not be personality centered. The emotional content of the challenge should be positive and motivating. The emotional reaction of the subordinate should be controlled by asking for his or her reaction to the challenge. The average candidate may not attempt to consistently motivate the candidate to act constructively and instead at times attempt to force the correct action. The average candidate will not usually solicit feedback from the subordinate. The below average candidate either does not motivate or negatively motivates the subordinate to act. Negative motivation is trying to coerce or force the candidate to act constructively of else face some veiled threats.

The delivery of the challenge should be done tentatively, but concretely. The challenging statement should not be stated assertively, but gradually increased in intensity. This tentativeness will allow the subordinate time to reach conclusions him or herself without feeling threatened. The delivery of the challenge is the most difficult component and should only be attempted with caution. The top level candidates will express their challenges in a friendly way. The average quality challenges will be tentative, but not concrete and not always friendly. The

below average quality challenges will be delivered either aggressively or passively. They may not be gradually increased or tentative and not always friendly.

What follows in Table 62 is a listing of the role players and their target ratings for each of the versions in which they acted. The table also lists the tape positions on the videotapes.

Table 62: Random Sequences, Associated Scores and videotape positions (Script, Candidate one, tape position, candidate two, tape position)

Order	1	2	3	4	5
1	8 Rick 306	6 Rachel 365	2 Juan 676	9 Steve 002	3 Denise 002
	Nancy 623	Matt 631	Rachel 596	Juan 001	Rick 002
2	1 Matt 522	2 Juan 676	8 Rick 306	3 Denise 002	6 Rachel 365
	Barb 448	Rachel 596	Nancy 623	Rick 002	Matt 631
3	3 Denise 002	9 Steve 002	4 Nancy 286	1 Matt 522	5 Sid 002
	Rick 002	Juan 001	Steve 250	Barb 448	Denise 290
4	6 Rachel 365	3 Denise 002	1 Matt 522	7 Barb 567	2 Juan 676
	Matt 631	Rick 002	Barb 448	Sid 439	Rachel 596

Order Cont.	6	7	8	9
1	7 Barb 567	1 Matt 522	4 Nancy 286	5 Sid 002
	Sid 439	Barb 448	Steve 250	Denise 290
2	7 Barb 567	5 Sid 002	9 Steve 002	4 Nancy 286
	Sid 439	Denise 290	Juan 001	Steve 250
3	2 Juan 676	7 Barb 567	6 Rachel 365	8 Rick 306
	Rachel 596	Sid 439	Matt 631	Nancy 623
4	8 Rick 306	9 Steve 002	4 Nancy 286	5 Sid 002
	Nancy 623	Juan 001	Steve 250	Denise 290

Dimension Definitions
Interpersonal Relations

A. Self Expression: disclosing or communicating one's point of view.

1. Information Content: stating information by expressing it concretely and directing it toward one's goals for the conversation
   (7) succinct, stated actively ("owned"), goal directed.
   (6) clear, not succinct, stated actively, goal directed.
   (5) clear, not succinct, not stated actively, mostly goal directed.
   (4) goals and info. understandable in most statements, not stated actively.
   (3) goal (purpose) or info. vague, stated either actively or passively.
   (2) goal (purpose) or info. unclear or irrelevant, stated passively.
   (1) information or point of view is not expressed.
   (0) not relevant; unable to rate.

2. Emotional Content: expressing positive (+) and negative (-) emotions genuinely, constructively and assertively
   (7) +/- emotion is genuine, expressed constructively, dealt with rather than held back, "owned" and stated actively.
   (6) +/- emotion is usually genuine, always expressed constructively, dealt with rather than held back.
   (5) +/- emotion is nearly always expressed constructively, dealt with rather than held back, sometimes not succinct.
   (4) +/- emotion is sometimes expressed bluntly or vaguely, but usually constructive.
   (3) +/- emotion is sometimes clearly not expressed constructively by being expressed aggressively or passively.
   (2) +/- emotion is clearly disruptive, unclear or irrelevant.
   (1) feelings are not expressed.
   (0) not relevant; unable to rate.

3. Delivery: disclosing information and feelings appropriate in depth and amount to the goals of the conversation.
   (7) appropriate to goals, gradually increased; friendly; shared.
   (6) appropriate to goals, gradually increased; friendly, not always shared.
   (5) usually appropriate to goals, assertive, gradually increased; not always friendly and shared.
   (4) sometimes not appropriate to goals by being blunt (not gradually increased) or vague, unfriendly or not shared.
   (3) sometimes inappropriate in depth, amount or duration, or irrelevant to target person, situation or detracting from the ongoing nature of the relationship.
   (2) clearly disruptive or inappropriately passive.
   (1) self expression not done.
   (0) not relevant; unable to rate.

Dimension Rating scales

E. Supervision:

1. Ensuring that the decision is carried out according to plan by planning clear goals and tasks, specifying evaluation criteria and time lines; organizing the tasks to be accomplished and directing and motivating the employee.

(7) implements solutions by clearly stating the goals, tasks, evaluation criteria and time lines; organizes the tasks and people to accomplish the tasks; directs and motivates subordinates to accomplish the goals.
(6) states a general plan and directs action toward the goal, provides some positive motivation.
(5) states the plan as a goal, may do some directing, without organizing or motivating.
(4) (3) some statements of hope are expressed, some weak motivating or negative incentives with few supervisory actions mentioned
(2) (1) no supervisory actions are mentioned.
(0) dimension subcomponent is not relevant or unable to rate.

F. Problem identification:

1. Monitoring the Decision Environment: Maintaining a clear understanding of how things should be and recognizing the deviations from normal, planned and expected states.

(7) specifies what actions are to be taken and why; sets up and initiates a method for monitoring the solutions chosen.
(6) initiates a monitoring plan, may not insure that the actions to be taken are understood.
(5) mentions a monitoring plan but does not necessarily initiate it.
(4) (3) (2) (1) monitoring is not mentioned or may be done in a way that detracts from the solution.
(0) dimension subcomponent is not relevant or unable to rate.

Dimension Rating scales

C. Decision Making (concluded)

4. Choose the best potential solution or course of action.

(7) decides on and specifies a course of action based on a systematic and thorough evaluation for each identified problem.
(6) decides on and specifies a course of action for the full set of problems but not based on a systematic and thorough evaluation.
(5) decides on and specifies a course of action for the major identified problems.
(4) decides on and specifies a course of action only for crises issues; longer term issues are continued.
(3) decides on and specifies a course of action, but a decision is not reached on at least one important issue.
(2) decisions are made to continue to gather information, specific actions are not made.
(1) decisions are not made; may express hope that the deviations will not occur again.
(0) dimension subcomponent is not relevant or unable to rate.

D. Action:

1. Implementing the best potential solution or course of action.

(7) takes specific actions to resolve the identified problems.
(6) takes specific actions to resolve the crisis issues and the most important identified problems, discusses but does not necessarily take specific actions for all issues discussed.
(5) takes specific actions to resolve the crisis issues; discusses and takes some actions for the other issues discussed, but does not necessarily take specific actions for all issues discussed.
(4) takes specific actions to resolve crisis issues, specifies that longer term issues will be discussed again in the future meetings.
(3) does not take specific actions to resolve the crisis issues, but some actions are taken.
(2) inappropriate, irrelevant or useless actions are implemented some of which may be counterproductive.
(1) no actions are taken.
(0) dimension subcomponent is not relevant or unable to rate.

C. Decision Making (continued)

2. Establish a methodology or criteria that is the most appropriate for appraisal of the possible solutions.

(7) discusses the criteria independently of development of possible solutions; discussion includes whether the problem will be resolved, the degree of satisfaction with the solution, and the amount of time required
(6) discusses criteria at the time the possible solutions are developed; discussion includes whether or not the problem will be resolved, and the amount of time required.
(5) develops and presents criteria, but does not discuss; criteria may fail to evaluate an important aspect of a problem solution or may be unable to evaluate a possible solution; process is linked with development of possible solutions.
(4) (3) (2) (1) criteria not discussed or evaluations are made without objective standards being stated or criteria are not developed.
(0) dimension subcomponent is not relevant or unable to rate.

3. Appraising possible solutions using the same set of criteria.

(7) clearly states the decision rules that are used to evaluate possible solutions; appraisal is systematic and thorough; appraises the possible solutions by stating the degree to which the criteria will be met.
(6) clearly states the decision rules that are used to evaluate possible solutions; appraisal may not be systematic and thorough; appraises only preselected solutions by stating the degree to which they meet the criteria (i.e., the "If this, then this" logic is clear).
(5) Appraises possible solutions by evaluating them as they are developed; states decision rules when pressed. ("If this, then this" logic is not clear).
(4) appraises possible solutions by default; does not discuss criteria or rationales, but states solutions as if appraisal had been done.
(3) appraises possible solutions incorrectly or inconsistently.
(2) (1) appraisals not done.
(0) dimension subcomponent is not relevant or unable to rate.

Dimension Rating scales

C. Decision Making:

1. Developing potential solutions or courses of action without judging or appraising them as they are developed.

(7) develops at least three possible solutions for each identified problem; seeks and persists in seeking help from the employee in the development of possible solutions; possibilities are developed over the full range of quality; returns to earlier problem solving steps if necessary.
(6) develops at least three possible solutions for each for each identified problem; seeks but does not persist in obtaining help in possibility development; returns to earlier problem solving steps if necessary.
(5) develops at least two possible solutions for each identified problem; no assistance is sought; returns to earlier problem solving steps if necessary.
(4) develops at least one possible action for each identified problem; no assistance is sought; returns to earlier problem solving steps if necessary for the most important issues only.
(3) possible solutions are not consistently developed for all identified problems; may not seek help when developing possible solutions; may not return to previous problem solving steps if a new problem arises.
(2) possible solutions are not typically developed, but some solutions are specified, most of which are irrelevant, inappropriate or ineffective.
(1) development of possible solutions is not done or solutions specified are not effective or even harmful.
(0) dimension subcomponent is not relevant or unable to rate.

Dimension Rating scales

B. Problem Analysis (continued)

2. Diagnosing the causes of the problem by digging beneath the symptoms: who and what caused the problem; when and where did the problem occur; why and how was the problem caused

(7) asks questions seeking the cause of the problem and all the details about who, what, where, when, why and how the problem occurred; seeks interrelationships among the causal details and the deviation facts.
(6) asks questions seeking the cause of the problem and all of the essential details; may not seek interrelationships among causal details or among deviation facts
(5) asks questions seeking the cause of the problem: what caused the problem and why did the problem occur.
(4) asks questions seeking the cause of the problem and what caused the problem; mentions other issues, but may not identify the causes.
(3) does not ask questions that are intended to identify the cause of the problem; seeks only symptoms (i.e., who was involved and affected, what was the impact on relations and actions, where and when did the problem occur and what were the costs).
(2) asks questions,but does not dig beneath the symptoms and identifies the wrong causes.
(1) does not ask questions seeking to identify the causes, problem diagnosis is not done.
(0) dimension subcomponent is not relevant or unable to rate.

APPENDIX C

APPENDIX C

PROBLEM SOLVING DIMENSION DEFINITION

The first step in the problem identification phase of the problem solving model is monitoring the decision environment. Top level candidates will clearly explain the normal, planned and expected actions and mention all of the deviations from them. Average level candidates will explain the expected states, but mention the major deviation and perhaps some of the lesser deviations. The candidates just below the minimally acceptable level will fail to mention one of the normal, planned or expected states, or fail to mention one of the major deviations from them.

The second step in the problem identification phase is to define the emerging problem. This consists of listing the essential details and analyzing them by combining, verifying, classifying and eliminating data in order to assess all of the details of symptoms of the problem. The top level candidate will ask questions and persist in seeking the essential details relating to who was involved, who was effected, what the impact will be on future actions and relations, where and when the problem occurred, what the costs of the problem are and what is the importance of the problem. Average quality candidates will ask questions seeking the problem participants and the impact, but few other details. The information will still be analyzed in a

logical manner and the obtained facts will still be explained. The below average candidates will seek to verify the prior biases or expectations and not seek the details for deviations that may arise after they have made up their minds. The below acceptable candidates may illogically analyze the information or jump to conclusions and may not explain the facts of the problem.

The problem analysis phase begins with the problem solver specifying decision objectives. The top level candidates will state expected outcomes, state the risks and constraints and specify who is to be involved in resolving each identified problem, all explained in practical and realistic terms. The average candidate will state the expected outcomes and risks and constraints, but not always in practical and realistic terms. The average candidate will make sure to mention that the employee must retain control of the problem solving process. The candidate who performs below the minimally acceptable level of performance will state expected outcomes or risks and constraints that are inappropriate, irrelevant or unclear.

The problem analysis phase concludes with a diagnosis of the causes of the problem. The symptoms of the problem were collected and listed out--without any evaluation--in step two, the problem definition phase of the model. In the diagnosis step the top quality candidates will dig beneath the symptoms by asking questions seeking the causes

of the problems. The causes will be addressed from a variety of perspectives: who caused the problem, what caused the problem, where and when did the problem occur, and why and how was the problem caused. In addition the top quality candidates will attempt to seek interrelationships among the causes of the various problems and form a complete picture of the situation. The average quality candidates will ask questions seeking primarily what caused the problem and why, plus some other concerns. The below acceptable candidate will not attempt to resolve the problems and seek only to identify symptoms.

Up to this point the problem solving model looks like a very linear, step by step process. This is not always the case. During a discussion, the candidate, for example, may be developing potential solutions for a particular problem (in subcomponent "C," to be described next) when the problem subordinate presents another issue. In this case, the candidate would need to return to earlier problem solving steps if necessary. This may involve seeking the symptoms of the new problem component, then, based on new information, specify new decision objectives and reassess the causes of the problem. If necessary, a top quality problem solver will return to earlier steps if new information arises.

The decision making cycle begins with the development of potential solutions without judging them or appraising them as they are developed. The above average candidates will develop multiple potential solutions for each identified problem. In addition, these top level performers will seek and persist in seeking assistance from the subordinate in potential solution development. Earlier problem solving steps will be utilized if necessary. The average quality performers will develop multiple solutions for at least one identified problem, but not for all issues. All issues will have at least one solution developed for them. No assistance will be sought in developing the solutions, but the average candidate will return to earlier problem solving steps if necessary. The below acceptable quality candidates will not develop potential solutions consistently for all important issues. This may arise by failing to return to earlier problem solving steps if necessary.

The second component of the decision making loop, is the establishing of a methodology or criteria for evaluation of the potential solutions. The top quality candidates will discuss the criteria development independently of the development of the potential solutions. The discussion will include a variety of factors including whether the problem will be resolved, the subordinate's degree of satisfaction with the solution and the amount of time required to resolve the problem. The average quality

candidates will not discuss the criteria development but will present the material. The criteria developed may fail to evaluate some important component of the problem solution. Since the criteria development step proves to be such a difficult and rare step in the process, the below average candidates will usually fail to explicitly develop any criteria or they may evaluate solutions without stating the criteria.

The third step in the decision making cycle is the actual appraising of the potential solutions. This is done by clearly stating the decision rules and using that same set for appraising all proposed solutions. The average quality candidates will appraise the potential solutions, but their logic may not be stated clearly or they may appraise the solutions as they are developed. The below average candidates will appraise the potential solutions incorrectly or inconsistently

Once the appraisal is done, the candidate will make the decision by choosing the best course of action for each identified problem. The top level candidates will decide on and specify a course of action based on a systematic and thorough evaluation for each identified problem. The average quality candidates will specify courses of action only for the crisis issues, but the other identified problems will be continued to worked on in the future. The candidates performing just below the minimally acceptable level

of performance will fail to specify a course of action for at least one major identified issue, but decisions on some solutions will be reached.

The action subcomponent is directly parallel to the choice step within the decision making loop. Top quality candidates will take specific actions to resolve the identified problems. The average level performers will take specific actions only for the crisis issues and will continue to work on the other issues. The below average candidates will fail to take specific actions to resolve the crisis issues, but some actions will be taken.

Once the actions are started, the effective problem solver will supervise the actions to ensure that the decisions are being carried out according to plan. Top level performers will implement solutions by first, planning and organizing the actions to be taken, then directing the performer of those actions and finally motivating the performers to complete the tasks. These are the basic components of the supervisory process. In the current situation, since the candidates are not supervisors, but only candidates, they are not expected to thoroughly perform the supervisory functions.

Top level performers of these supervisory actions will implement the solutions by clearly stating the goals, tasks, evaluation criteria and time lines. Then the top

performers will organize the tasks and the people to accomplish those tasks, direct the task performers and motivate them. Average level performers will state a plan as a general goal and may do some directing, without organizing or motivating. Since this is a difficult task for candidates for supervisors, the minimally acceptable level of performance is lowered to include providing some directing statements and weak statements of hope with some weak motivating statements or negative incentives used. Negative incentives are veiled threats in the form of "do this and avoid the consequences."

The last subcomponent of the problem solving model completes the loop by once again monitoring the decision environment to check whether the now resolved problem reoccurs. Top level performers specify the actions are to be taken and why. Then, more importantly, the top level performers will set up and initiate a monitoring plan to watch the implementation of the chosen solutions. This step is a relatively rarely performed one without specific training, so even the average level performers, in a selection process, will be found to mention a need for monitoring the solution without initiating it. Since this is a difficult dimension the minimally acceptable level of performance does not require any mentioning of a monitoring plan or the plan that is mentioned may even detract from the accomplishment of the solution.

Dimension Rating scales

B. Problem Analysis.

1. Specify the decision objectives in practical and realistic terms.

(7) for each identified problem states expected outcomes, risks, constraints and the people to be involved, all in practical and realistic terms.
(6) states expected outcomes or risks and constraints but not always in practical and realistic terms; may not mention people to be involved other than that the employee must retain control; may make some statements on risks without a clear problem referent.
(5) states expected outcomes or risks and constraints but not in practical nor realistic terms; does not specify that the employee must retain control of the problem resolution.
(4) states that the expected outcome of the problem solving situation is the absence or correction of a deviation; specifies the risks involved in not correcting the deviations.
(3) states expected outcomes that may be inappropriate, unclear or irrelevant; may specify inappropriate, unclear or irrelevant risks and constraints.
(2) either does not set problem solving objectives or does not mention risks and constraints.
(1) fails to specify problem solving objectives and fails to specify risks and constraints.
(0) dimension subcomponent is not relevant or unable to rate.

2. Defining the Problem: listing the essential details and analyzing them by combining, verifying, classifying and eliminating data in order to assess who was involved, who was affected, what the impact is on future relations and future actions, where and when the deviation occurred, and the costs and importance of the deviation.

   (7) asks questions and persists in seeking the essential details relating to deviations; analyzes the information in a logical manner and explains the essential facts of the problem.
   (6) asks questions and persists in seeking the essential details relating to who was involved and who was affected and the impact of future relations and actions, but does not persist in seeking information on where, when, costs, and future impact details; analyzes the obtained information in a logical manner and explains the obtained facts of the problem
   (5) asks questions but does not always persist in seeking the essential details relating to deviation participants and the impact; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.
   (4) asks questions but does not persist in seeking the essential details relating to deviation participants and the impact; may not explain the facts of the problem; may not explain the analysis of the information obtained and simply state conclusions.
   (3) explains and seeks to verify the details of the pre-work problem definition information without asking questions to obtain the essential details of the issues that may arise during discussion; may jump to conclusions or illogically analyze the information; may not explain the facts of the problem.
   (2) explains the details of the pre-work problem definition information without seeking to verify them or gather the essential details of the issues that arise in discussion.
   (1) does not ask questions on the deviations; does not verify pre-work information; does not explain the problem.
   (0) dimension subcomponent is not relevant or unable to rate.

Dimension Definitions
Problem Solving

A. Problem Identification:

1. Monitoring the Decision Environment: Maintaining a clear understanding of how things should be and recognizing the deviations from normal, planned and expected states.

(7) clearly explains the normal, planned or expected actions and mentions the deviations from them.
(6) clearly explains the normal, planned or expected actions and mentions all but the minor deviations from them.
(5) clearly explains the normal, planned and expected actions and mentions the major deviation and some other secondary ones.
(4) clearly explains the normal, planned and expected actions and mentions only the major deviations.
(3) fails to clearly explain one of the important normal, planned or expected actions or fails to mention one of the major deviations.
(2) fails to clearly explain more than one of the important normal, planned or expected actions or fails to mention more than one of the major deviations.
(1) does not explain the normal, planned or expected actions or does not mention any deviations from these expectations.
(0) dimension subcomponent is not relevant or unable to rate.

D. Challenging : After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. Information Content: challenges are clear, behaviorally centered on observable or reasonably inferred facts and is not personality centered.

   (7) succinct, stated actively ("owned").
   (6) clear, not succinct, stated actively.
   (5) clear, not succinct, not stated actively.
   (4) understandable,not stated actively.
   (3) vague, stated either actively or passively.
   (2) unclear or personality centered or irrelevant, stated passively.
   (1) challenging is not done.
   (0) not relevant; unable to rate.

2. Emotional Content: seeking to create positively motivating feelings and emotions by asking for reactions and not making "hit and run" statements.

   (7) succinct, motivating, soliciting feedback.
   (6) clear, not succinct, motivating, may not solicit feedback.
   (5) clear, not succinct, may not be motivating, usually does not solicit feedback.
   (4) understandable, does not motivate the person to act constructively, may attempt to force correct action, usually does not solicit feedback.
   (3) vague or disruptive, not motivating or negatively motivating, may resist feedback.
   (2) unclear or irrelevant or negatively dominating
   (1) challenging not done.
   (0) not relevant; unable to rate.

3. Delivery: making challenging statements that are tentatively phrased but facilitative and concrete

   (7) tentative, but concrete, gradually increased; friendly.
   (6) tentative, not concrete, gradually increased; friendly.
   (5) tentative, not concrete, gradually increased; not always friendly.
   (4) sometimes blunt or vague, not always tentative or gradually increased; not always friendly.
   (3) sometimes aggressive or passive, may not be tentative or gradually increased; not always friendly.
   (2) disruptive or passive; not gradually increased; not friendly.
   (1) challenging not done.

Dimension Rating scales

C. Probing and Questioning: gathering information from another person.

1. Information Content: asking questions or making probing statements that are clear and understandable.

(7) succinct, stated actively ("owned").
(6) clear, not succinct, stated actively.
(5) clear, not succinct, not stated actively.
(4) understandable,not stated actively.
(3) vague, stated either actively or passively.
(2) unclear or irrelevant, stated passively.
(1) information or point of view is not expressed.
(0) not relevant; unable to rate.

2. Emotional Content: of questions or probes is neutral and controlled by the use of open-ended, not closed (one word answer) questions

(7) succinct, open-ended questions and statements.
(6) clear, not succinct, open-ended questions and statements.
(5) clear, not succinct, open-ended disintegrating into closed (yes/no) questions.
(4) understandable, questions and statements are both open-ended and closed, some +/- emotions present.
(3) vague or repeated closed questions and statements, positive or negative emotions color info. gathering.
(2) unclear or irrelevant or exclusively closed questions and probes, +/- emotions dominate info. gathering.
(1) questions and probing statements not asked.
(0) not relevant; unable to rate.

3. Delivery: asking questions and making probing statements that facilitate and supplement, rather than dominate the discussion.

(7) assertive, facilitates the discussion, gradually increased; friendly.
(6) assertive, facilitates the discussion; friendly .
(5) assertive, may not always facilitate the discussion, may not always be friendly.
(4) may be blunt or vague or may dominate the discussion with questions at times, may not always be friendly.
(3) sometimes aggressive or passive, sometimes controls and dominates discussion with questions, not always friendly.
(2) disruptive or passive; not friendly, does not supplement, but always dominates and controls the discussion.
(1) questioning and probing not done.
(0) not relevant; unable to rate.

B. Responding with Empathy: communicating accurate understanding of the core meaning of another person's words.

1. Information Content: expressing the core meaning of another person's words.

(7) succinct, stated actively ("owned").
(6) clear, not succinct, stated actively .
(5) clear, not succinct, not stated actively.
(4) understandable,not stated actively.
(3) vague, stated either actively or passively.
(2) unclear or irrelevant, stated passively.
(1) information or point of view is not expressed.
(0) not relevant; unable to rate.

2. Emotional Content: expressing the correct type and intensity of another person's feelings and emotions

(7) succinctly states correct type and intensity.
(6) clearly, but not succinctly states correct type and intensity.
(5) correct type and intensity are understood but may not be clearly fed back.
(4) clearly, but not succinctly states correct type, intensity fed back may be a bit off, but not grossly so.
(3) may be a bit off in stating correct type or intensity fed back is clearly off target or may miss opportunity to respond with empathy.
(2) type and intensity of emotion fed back is unclear or irrelevant, may miss opportunity to respond with empathy.
(1) emotional content is not fed back.
(0) not relevant; unable to rate.

3. Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.

(7) clear and succinct, not rambling; assertive but not aggressive; facilitates discussion rather than disrupting; uses metaphor/humor rather than parroting.
(6) clear, not rambling; assertive but not aggressive; facilitates discussion rather than disrupting; sometimes uses metaphor/humor, but does not parroting.
(5) clear, not rambling; assertive but not aggressive; usually facilitates discussion, may be a bit disruptive; some use of metaphor, but may parroting.
(4) clear, but sometimes rambling; assertive but not aggressive; usually facilitates discussion, may be a bit disruptive; weak use of metaphor, may parrot.
(3) may be rambling; may be aggressive; may be a disruptive; may parrot, little use of metaphor/humor.
(2) clearly disruptive or inappropriately.
(1) responding not done.
(0) not relevant; unable to rate.

APPENDIX D

Mock Candidate Pre-Work Material

Memo
To: New Unit Supervisor
From: Mary Johnson, the old Unit Supervisor
Date: August 19, 1985
Subject: Meeting on Chester's Performance Improvement plan

When I found out that I was being rotated to another unit I didn't have time to resolve an employee performance problem. I guess that you'll have to jump into this one cold. I've dragged my feet a little on this, but the situation needs to be fixed. So, I've written this memo to explain the issue a bit. I've also attached the employee's job description, most recent performance appraisal form and a memo from Personnel describing the discipline and termination process. Here's "the bomb:" the meeting to present and discuss this problem with the employee involved had been set up for today and will be held in one-half hour.

Now a little background. The Department Director is really sensitive to complaints from our clientele and their politically powerful and quite vocal advocates. The ultimate problem happened last week. We got chewed out for failing to handle a new case the same day the client attempted to call. Instead, our client ended up on the department head's doorstep at ten o'clock at night last week Tuesday.

Chester, a permanent employee, is the problem child you have inherited who is responsible for the dilemma. It was his client who called. Chester was "sick" that day. No one knew about it until too late.

I can't say we weren't warned. He had been doing a good job in this division for many years, first as a financial worker and a Senior Financial worker, and now as a leadworker, Principal, when performance quality began to fall off. Over the course of the last six weeks he has developed a pattern of not following the County work rules. Chester fails to call the office when sick and is out of the office with colds or other problems too often. When at work spends too much time socializing, was seen having a beer over lunch once, he rarely even ate lunch in the past, and is extremely slow in the, as you know, somewhat technical leadworker's duties of assisting the other workers in determining client eligibility for receiving Public Assistance money. His work is not getting done on time. Chester has requested a transfer to another unit. This request has been denied since he has not been in the unit for the required eighteen month period before becoming eligible to transfer again.

Good luck in your meeting with Chester. Please keep the meeting to about fifteen minutes. Again it will start in one-half hour and will be held in the oral examination room.

PHITEM/PRORAL

CLASS TITLE: Principal Financial Worker

DEFINITION: Under general supervision, performs work of moderate difficulty in a variety of tasks relating to processing requests for public assistance; acts as resource to public and Financial Workers in the areas of responsibility; may function as a lead worker and assist in the training of Financial Workers; performs special tasks and assignments related to program policies and procedures.

EXAMPLES OF DUTIES:

Intake Worker - Performs on a full time basis the tasks related to initial eligibility determination such as: interviews persons requesting public assistance for the purpose of determining initial program eligibility; advises applicants of program eligibility requirements; provides applicant with a list of necessary verifications; reviews verification data and evaluates initial screening assessment; compares personal and financial status of applicants with standards for program eligibility and determines initial eligibility; may participate in process appeals on denied eligibility, amount of benefit, and data assistance begins; generates initial issuance documents; informs applicant of initial eligibility determination and reasons for decision; explains detailed multiprogram information to public, other agencies, or community groups; performs related work as required.

Lead Worker - Officially designated and assigned the responsibility for being the primary backup for the unit supervisor, as such work performed is exampled by the following tasks; acts as the unit resource person for the unit staff; trains or assists in the training of unit staff and instructs on unit procedures; reviews work of unit staff to ensure conformity to established procedures and requirements; observes work of unit staff and assists supervisor with performance evaluations; may assign work to unit staff; carries partial case load which may be more difficult, complex, and/or time consuming cases, performs related work as required.

Support Services Worker - Performs a highly technical and/or multiprogram staff support function directly related to the department's service mission; such activities included are exampled by: Quality control, trainer, advocate, fraud advisor, and the peak period pool.

EMPLOYMENT STANDARDS:

Education and Experience. Two years of experience as a Financial Worker or Senior Financial Worker; or an equivalent combination of training and experience.

Knowledge, Skills and Abilities. Considerable knowledge of interviewing techniques, office procedures and household budget concepts; considerable knowledge of public assistance programs, operations, policies and procedures; good knowledge of community resources.

Considerable ability to follow instructions; considerable ability to organize work and meet deadlines; considerable ability to establish and maintain good relationships with clients; staff and community resources; working ability to express ideas clearly in oral or written form.

Considerable skill in judging the validity of applicant claims; some skills in training Financial Workers.

Licenses and Certificates. Possession of a valid Minnesota driver's license may be required for certain positions in this class.

HENNEPIN COUNTY

6/81
Rev. 6/82
Rev. 3/84
Rev. 5/84
Rev. 5/87

Memo

To: Mary Johnson, Unit Supervisor
From: Bob Zimmerman, Personnel Department
Date: August 14, 1985
Subject: Employee's Performance Improvement plan

As requested, here is the sequence from the Work Rules for employee discipline and termination. Remember that termination must be based on well documented facts that are clearly in violation of a work rule. If you intended to follow this process through to the termination of an employee, give me a call and we can talk in more detail about the process and the documentation necessary.

A) Set up with the employee, in writing, a clearly understood expectation that the employee must follow in order to be considered retainable. The actions you mentioned on the phone--excessive absence and tardiness, alcohol or drug abuse, failure to complete the assigned duties, and the inability to meet consistently one's performance objectives--are all potentially related to the discipline and termination procedure.

B) The first violation, after the employee has been told of the expectation, should lead to a written reprimand. This letter is given to the employee, with copies going to the personnel file and to your department head.

C) The second violation of the expectation will lead to a one-day suspension. Copies of the suspension notice are sent the Personnel file and to your department head.

D) The third violation of the expectation will lead to a five-day suspension. Copies of the suspension notice are sent the Personnel file and to your department head.

E) The fourth violation of the expectation will lead to a termination hearing with the Personnel Board. If the process gets this far, the Personnel Department will be working directly with you to make sure the proper documentation is available.

Let me know about your exact plans. I hope we can resolve this problem.

EMPLOYEE PERFORMANCE REVIEW

Department: Economic Assistance
Employee Name: Chester Marcol
Date of Rating: 2-25-85

(circle one rating per factor)

1. Quality: Performance in meeting standards of work quality.

1	2	(3)	4	5
careless	just gets by	does a good job	errors rare	exception-ally high quality

2. Job knowledge: Understanding in all phases of the work.

1	2	3	4	(5)
inadequate knowledge	improvement necessary-just gets by	knows job fairly well	expert but limited to own job	expert, acts as resource to staff

3. Quantity: Output of satisfactory work.

1	2	(3)	4	5
Slow, work is seldom required amount	Turns out required amount, no more	frequently turns out more than required	usually does more than expected	exceptional, fast output

4. Dependability: Works conscientiously according to instructions.

1	2	3	(4)	5
continuous checking and follow-up required	frequent checking required	follows instructions	very little checking required	dependable, no checking necessary

5. Initiative: Think constructively and originates thought.

1	2	(3)	4	5
requires constant supervision	fair decisions, routine worker	good decisions and actions, but requires some supervision	minimum of supervision	Thinks and acts construct-ive super-vision not required

EMPLOYEE PERFORMANCE REVIEW
PAGE 2

6. Adaptability: Ability to learn and to meet changing conditions.

1	2	(3)	4	5
Prefers old methods, does not remember instructions	learns slowly, reluctant to change	normal ability, routine worker	willing to change, short period for mental adjustments	learns rapidly, adjusts and grasps changes quickly

7. Attitude: Willingness to cooperate and to carry out demands.

1	2	3	4	(5)
poor cooperation argues	passive resistance	limited cooperation	cooperative	good team worker

8. Attendance: amount of absenteeism and tardiness in six months.

1	2	3	4	(5)
more than 4 days absent or tardy in 6 months	Two thru four days absent, 3 days tardy	one day absent, 2 days tardy	no days lost 2 days tardy	no days lost, not tardy

9. Staff Relations: Ability to get along with associates.

1	2	(3)	4	5
disagreeable	difficult to get along with	average or reasonable	well liked and respected	winning personality

10. Public Interpersonal Relations: Ability to get along with the public.

1	2	(3)	4	5
disagreeable	difficult to get along with	average or reasonable	well liked and respected	winning personality

Simulation Exercise Summary
Problem Solving

Problem Solving
Definition and Simulation Exercise Summary

A. Problem Identification

1. Monitoring the Decision Environment: Maintaining a clear understanding of how things should be and recognizing the deviations from normal, planned and expected states by clearly explaining the normal, planned or expected actions and mentioning the deviations from them.

   Potential Deviations from expectations in Chester's work activities to be Mentioned:

   1. a client was not served by Chester and the client caused problems by going to the Director's house at 10 P.M. last week Tuesday.
   2. There have been reports that Chester fails to call when he is sick.
   3. There have been reports that Chester is sick too often, has problems with absenteeism and promptness.
   4. There have been reports that Chester socializes too much when in the office.
   5. Chester's performance has deviated for the worse from his old performance appraisal in February.
   6. There have been reports that Chester has requested a transfer out of the unit.
   7. There has been a report that Chester, who rarely eats lunch at all, had been seen drinking a beer with lunch.
   8. There have been reports of a general increase in problems by Chester.

2. Defining the Problem: listing the essential details and analyzing them by combining, verifying, classifying and eliminating data in order to assess who was involved, who was affected, what the impact is on future relations and future actions, where and when the deviation occurred, and the costs and importance of the deviation by asking questions and persisting in seeking the essential details relating to deviations; analyzing the information in a logical manner and explaining the essential facts of the problem.

   Potential information to be brought out through questioning includes the details within the following categories:

   1. who was involved?: Chester, secretary, client, department head
   2. who was affected?: client, director, Mary, new supervisor, unit staff
   3. what is the impact on future relations?: Chester, director and client upset; unit supervisor is unsure if Chester is reliable

Simulation Exercise Summary
Problem Solving

A. Problem Identification (continued)

2. Defining the Problem (continued)

4. what was the impact on the actions to be done: client not served; Chester is not being productive and is not complying with several "work rules"
5. where did the problem occur?: in the office while Chester was at home
6. when did the problem occur?: specifically, last week Tuesday, but also over the last six weeks
7. what is the cost of the problem?: the dollar cost not specified in the information provided, but, Chester's work is not getting done on time. This is costing money through decreased productivity and lost work time.
8. What is the importance of this problem?: The problem client was not served and several others are probably not being served effectively and Chester has been skipping out of work. These are vary serious problems
9. what happened?: Chester, whose productivity has dropped and who is violating several "work rules" on absenteeism, tardiness, alcohol use during the lunch hour and excessive socializing in the office, failed to call the office when he was sick one day last week and a client who was not given proper service by Chester got so angry that he went to the Director's house at ten o'clock in the evening last week Tuesday.

B. Problem Analysis.

1. Specify the decision objectives in practical and realistic terms for each identified problem by stating expected outcomes, risks, constraints and the people to be involved, all in practical and realistic terms.

The general problem categories and the possible expectations to be mentioned include Chester being asked to:

A. Improve productivity

1. help the client(s) and avoid these service delivery problems in the future
2. Do the job expected of a PFW

B. Follow the "work rules"

1. call in when sick
2. improve attendance/promptness
3. do not drink alcohol on the work site
4. decrease socializing

Simulation Exercise Summary
Problem Solving

B. Problem Analysis. (continued)

1. Specify the decision objectives (continued)

C. Improve job satisfaction
1. improve satisfaction/ work relationships
2. become promotable

The people who may be involved in meeting the expectations may include:

1. Chester should retain control of his own problems and work to resolve them.
2. Chester should tell the Supervisor if he feels overworked

The potential risks that can be mentioned are:

1. First, an oral reprimand
2. Second, a written reprimand
3. Third, a one day suspension
4. Fourth, a five day suspension
5. Finally getting fired after a termination hearing with the Personnel Board
6. A general statement of adverse consequences can be mentioned

2. Diagnosing the causes of the problem: by digging beneath the symptoms: who and what caused the problem; when and where did the problem occur; why and how was the problem caused. The supervisor accomplishes all this by asking questions, seeking the cause of the problem and all the details about who, what, where, when, why and how the problem occurred and seeking interrelationships among the data.

Questions diagnosing the cause of the problem will include the following:

1. Who caused the problem: Chester
2. what caused the problem: Promotion issue with Stan
3. when was the problem caused: six weeks ago
4. where was the problem caused: on the unit, and in the selection processes
5. why was the problem caused: anger, frustration, burnout
6. how was the problem caused: work slow down and failing to follow the work rules, especially not coming into work or phoning in when sick
7. Seeking interrelationships among deviations and among causes (e.g., Chester may be frustrated because Stan was only a Senior when he was promoted; the work rule problems may only be symptoms of the promotion issue)

Simulation Exercise Summary
Problem Solving

C. Decision Making:

1. Developing possible solutions or courses of action that may be used to resolve the problems without judging or appraising them as they are developed by developing at least three possibilities for each identified problem; and seeking and persisting in seeking help from the employee in possibility development. The possibilities are developed over the full range of quality. The supervisor should return to earlier problem solving steps if necessary (e.g., when new information comes up on the promotional issue and "Stan", additional facts are needed, etc.).

   Possibilities that can be developed include those which deal with the following topics:

   A. Improve productivity

      1. Chester should assess his service, accuracy, and productivity.
      2. The new supervisor may review Chester's and the unit's workload.

   B. Follow the "work rules"

      1. Chester should come to work on time and call in when sick. This must be stated.
      2. Chester should cut down on socializing.

   C. Improve job satisfaction (promotability)

      1. Chester should be assisted by the supervisor in developing a plan for promotion.
      2. Chester should reassess his career with the County to determine whether he still wants to work here or if he would rather work some where else.

   D. General supervisory action possibilities

      1. The supervisor should seek assistance from Chester in developing possibilities.
      2. Chester should ask for help from the new supervisor.
      3. Chester and the new supervisor may want to meet again to continue the problem solving discussion.
      4. The new supervisor may choose to initiate a special review processes.

Simulation Exercise Summary
Problem Solving

C. Decision Making: (continued)

2. Establish a methodology or criteria that is the most appropriate for appraisal of the possibilities by discussing the criteria independently of possibility development. The discussion includes whether the problem will be resolved, the degree of satisfaction with the solution, and the amount of time required.

   Potential criteria that can be established to evaluate the utility of possible solutions may include several things within each general problem. The important decisions to be reached include whether the possibility will cause Chester to:

   A. Improve productivity
      1. provide proper service to clients
      2. face discipline or resolve problems

   B. Follow the "work rules"
      1. come to work on time
      2. call in when sick
      3. follow rules on socializing
      4. follow rules on drinking
      5. general following of rules

   C. Improve job satisfaction (promotability)
      1. become promotable again
      2. improve satisfaction, without transfer
      3. take the initiative to improve

3. Appraising possibilities using the same set of criteria by clearly stating the decision rules. The appraisal is systematic and thorough; appraising the possibilities by stating the degree to which the criteria will be met.

   Actions may be taken and possibilities selected in the following areas:

   A. calling in when sick (action is required to be selected)
   B. Improving productivity
   C. Improving satisfaction and promotability

4. Choose the best possible solution or course of action by deciding on and specifying a course of action based on a systematic and thorough evaluation for each identified problem.

   Solutions that are chosen may include the following :

   A. Improve productivity
      1. review workload
      2. improve accuracy/ productivity.

Simulation Exercise Summary
Problem Solving

C. Decision Making: (continued)

4. Choose the best possibility (continued)

B. Follow the "work rules"
1. improve attendance/promptness/ calling in when sick.

C. Improve job satisfaction (promotability)
1. work on resolving promotion problems
2. review career with County

D. General supervisory actions:
1. meet again to work on problems and goals
2. use of the discipline process.
3. use of a special review processes

D. Action:

1. Implementing the best possible solution or course of action by taking specific actions to resolve the identified problems.

A. Improve productivity
1. Chester should work to improve (retain control of resolving the problem with) his accuracy and productivity by following the work rules and returning to his former level of performance.
2. The supervisor may review the workload to see if distribution should be adjusted.

B. Follow the "work rules"
1. Chester should improve his attendance and promptness by calling the supervisor within a half hour of the start of the work day whenever he is going to be sick or out of the office.

C. Improve job satisfaction (promotability)
1. Chester should work to improve his promotional preparedness by working with the supervisor to set up a program to begin to deal with this issue.

D. General supervisory actions:

1. The Supervisor and Chester should meet again to continue working of the solutions.
2. The supervisor may offer assistance in pinpointing the problems and resolving errors.
2. The Supervisor may initiate the disciplinary process if Chester fails to meet expectations on productivity and work rules.

E. Supervision:

1. Ensuring that the decision is carried out according to plan by planning clear goals and tasks, specifying evaluation criteria and time lines; organizing the tasks to be accomplished and directing and motivating the employee through impiementing solutions. This will require clearly stating the goals, tasks, evaluation criteria and time lines; organizing the tasks and people to accomplish the tasks; and directing and motivating the staff to accomplish the goals.

   The supervisor should:
   1. state a plan.
   2. direct Chester's actions.
   3. motivate Chester to continue working on the problem's solution.

F. Problem Identification

1. Monitoring the Decision Environment: Maintaining a clear understanding of how things should be and recognizing the deviations from normal, planned and expected states by specifying what actions are to be taken and why; setting up and initiating a method for monitoring the solutions chosen.

   The supervisor should:
   1. initiate a plan

Interpersonal Relations Summary

A. Self Expression: disclosing or communicating one's point of view or other information (including starting and stopping the conversation) by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. The Information Content of one's expressions should be stated succinctly and actively. Sentences should be concrete and goal directed and should use action verbs stated in the first person (i.e., "owned, by saying, for example, "I think," not "the department understands that," or "one would think" etc.).

2. The Emotional Content of one's expressions (i.e., expressing one's feelings and emotions) should be stated succinctly and actively. Sentences should be concrete and should use action verbs stated in the first person (i.e., "owned, by saying, for example, "I feel" not "the department feels" or "one would feel" etc.). Emotions should be understood to be legitimate parts of behavior and expressed genuinely, constructively and not held back

3. The Delivery of one's point of view (whether disclosing information or feelings) should be appropriate in depth and amount to the listener, the relationship already established, the situation and the goal of the conversation. The delivery should be presented assertively (neither passively nor aggressively), gradually increased in intensity (rather than being disruptive, blunt or "dumped" on the listener or not delivered at all), friendly (rather than unfriendly and negative), and shared (not one sided).

May 2, 1987
PHIRSMRY/PRORAL

B. Responding with Empathy: communicating accurate understanding of the core meaning of another person's words. This will give the other person the feeling of being understood and should be done before any information gathering or challenging takes place. If the other person becomes resistive or defensive, further empathic responding should take place before proceeding.

1. The Information Content in another person's communication should be feed back to that person by expressing the core meaning of the person's words, but not the detail. Empathic responses should be stated succinctly and actively. Sentences should be concise and as brief as possible and should use action verbs stated in the first person without containing questions (i.e., "owned, by saying, for example, "You think", not "you think that" or "one would think that" etc.).

2. Feeding back to another person the correct Emotional Content of their words (i.e., responding to another by mentioning the type and intensity of another's feelings and emotions) should be stated succinctly and stated actively. That is sentences should be concise and as brief as possible and should use action verbs stated in the first person (i.e., "owned, by saying, for example, "you feel angry" not "you feel that something bad is happening" or "one would feel that" etc.). This must happen whenever the other person expresses some significant (whether stated or not stated) emotions.

3. When expressing accurate understanding of another person's thoughts or feelings, The expression should be Delivered in a way that is appropriate in depth and amount to the information and feelings expressed by the other person. The delivery should be facilitative (not disruptive or judgmental), assertive (but not aggressive or passive), gradually increased in intensity (rather than bluntly being "dumped" on the listener or not delivered at all), friendly (rather than unfriendly and negative), and shared (not one sided).

C. Probing and Questioning: gathering information from another person, but not verifying preconceptions.

1. The Information Content of questions and probing statements should be clear and understandable by being stated succinctly and stated actively. Probes and questions should be concise and as brief as possible and should use action verbs stated in the first person.

2. The Emotional Content of probes and questions should be controlled by asking neutral, succinct, open-ended questions and statements, not vague or irrelevant or closed (yes/no, one word answer) questions and statements.

3. The Delivery of questions and probing statements should supplement the discussion by being appropriate in depth and amount to the ongoing nature of the relationship. The probes and questions should by delivered assertively (neither passively nor aggressively disruptive), gradually increased in intensity (rather than bluntly being "dumped" on the listener or not delivered at all), friendly (rather than unfriendly and negative), and shared (not one sided).

D. Challenging : Attempted only after establishing a positive interpersonal relationship (through the expression, listening, responding "loop") providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. The Information Content of challenges should be succinct (concrete and as brief as possible), behaviorally centered on observable or reasonably inferred facts and should use action verbs stated in the first person (i.e., "owned) and not be unclear, personality centered, irrelevant or stated passively.

2. The Emotional Content of challenges should seek to create positive feelings and emotions. The challenges should be succinct and motivating, causing the other person to act in a constructive manner. Negative reactions should be controlled by soliciting feedback from the person being challenged.

3. The challenging statements should be Delivered in a style that is facilitative and appropriate in depth and amount to the problem to be resolved. Challenging statements should be made tentatively (for example "It seems to me" or It might be" rather than disruptively saying "You should," or being vaguely), gradually increased in intensity (not blunt or passive), made in a friendly way and shared (not forced).

APPENDIX E

Background Information
On the "Mock" Supervisor Oral Exam
The Exercise, Chester's Script and Your Goal as a Rater

This simulation exercise for entry level supervisor is situated in the Department of Economic Assistance. It could have been situated anywhere. This Department was chosen simply because an extensive test development job analysis had been completed prior to the start of the research project. Chester, the subordinate with whom the candidate - supervisor must meet, is a "whiny," defensive, top level line worker who has been having some problems. Chester could have been employed in any department. Chester's strong, if obnoxious, defenses do not reflect in any way on the activities of the real Department of Economic Assistance. Chester feels perfectly free to use what ever he can think of at the moment to explain away his bad attitude. If this involves making the entire department's staff look incompetent, Chester will have no qualms about doing so. Needless to say, the actual operation of the Department of Economic Assistance is nothing similar to the way Chester portrays it.

Chester's role is consistent from candidate to candidate. His script is made up of several actual incidents of subordinate problems that a new supervisor would be expected to handle. The intention is to allow the candidate - supervisor to display his or her problem solving and interpersonal relations skills in a simulated "real" exercise, without needing any department specific technical knowledge. As a supervisor, one can expect Chester's style of problem (or similar ones) to arise sooner or later in any department. The better the supervisor's problem solving and interpersonal relations skills, the easier the problem will be to handle.

Chester's scripted role is this: he is a Principal Financial Worker whose assignment is as a lead worker. He has been in the department for nine years. He had been looking forward to being promoted to a unit supervisor position. He is extremely dissatisfied at not having been promoted in what was, he believes, his only chance for the next five years. Several factors add to his frustration. One of his co-workers ("Stan", a Senior Financial Worker, who is actually below Chester's level) was promoted to supervisor. It is quite rare for a Senior to be promoted to Supervisor, skipping the Principal level position. Chester thinks that Stan is unqualified. Chester feels that he had been doing most of Stan's work for him while he had been out "politicking" for the promotion. Both Chester and Stan got on the eligible list for promotion.

Since Stan left the unit, all of his work has been given to Chester, because Chester is the lead worker. Chester has also been given the assignment of training Stan's replacement. Now, since the supervisor is also brand new (i.e., the candidate plays the role of Chester's new supervisor), Chester thinks that he could get stuck with running the unit for an inexperienced newcomer and not get paid for it. Chester sees himself as doing three people's jobs. In fact, however, all of the assignments are in his job description.

Chester has decided to take things easier, since (at least from his point of view) so many other people are doing the same. Chester is not intentionally making mistakes, but will actively resist and get defensive about any suggestions that he is not working as hard as the other employees. He feels tired of doing a thankless job and feels insulted by an impersonal system that is trying to use him without rewarding him. These emotions, however, are not clearly thought through. The long term problem faced by the candidate - supervisor is to resolve these dilemmas being felt by Chester and to get him to be productive and promotable again.

The short term problem is to make sure that Chester comes to work on time, calls the supervisor when he will be ill and improve his attendance. These issues must be handled during the discussion. Chester's phone calls also need to be handled the same day that they are received. Chester's response will be that when a person is at home sick (and the secretary does not know it), then the secretary assumes that the worker is out in the field doing visitations or collecting information from clients. Any messages are left on the worker's desk. Since many employees tend to spend several hours at a time in the field, leaving messages is a common, accepted practice. The worker's phone calls are usually answered when the worker returns to the office either later that day or the next morning.

Chester feels that the problem belongs to the secretary. Chester thinks that the secretary should leave messages in a common location so that other staff members can see when a worker's phone messages are not being answered. If other staff members see the messages, then these types of problem clients can be dealt with quickly. Chester himself has done this "favor" for other workers in the past and he is angry that they had not returned the favor to him this time.

Now, a final word about the videotaped performances. None of the people participating in the videotapes are professional actors or actresses. Because of this, the "acting" is weak. You will quickly notice that the lines being spoken are being read off of "cue cards." Please do not let this affect your ratings of the candidates. At times the nonverbal behaviors of the candidate-supervisors or Chester become stilted or inconsistent with what is being said. When this happens, and at all times in this mock oral exam **please rate the candidates only on what they are saying, not their nonverbal behaviors.** Your goal in this mock oral examination is to **rate the candidates for supervisor as accurately as you can on the two major rating scales.**

APPENDIX F

Script 1
PS=6, IR=6

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	6	clear, not succinct, stated actively, goal directed.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	6	+/- emotion is usually genuine, always expressed constructively, dealt with rather than held back.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	6	appropriate to goals, gradually increased; friendly, not always shared.

Related Paragraphs: 1,2,3,7,8,9,10,11,12,13,14,15,16,17,22,23,24,25

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	6	clear, not succinct, stated actively .

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	6	clearly, but not succinctly states correct type and intensity.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	6	clear, not rambling; assertive but not aggressive; facilitates discussion rather than disrupting; sometimes uses metaphor/humor, but does not parroting.

Related Paragraphs:
3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,19,20,21,22,23,24,25

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	6	clear, not succinct, stated actively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	6	clear, not succinct, open-ended questions and statements.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	6	assertive, facilitates the discussion; friendly.

Related Paragraphs: 2,3,4,6,7,8,10,11,17,19,20,23,24

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	6	clear, not succinct, stated actively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	6	clear, not succinct, motivating, may not solicit feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	6	tentative, not concrete, gradually increased; friendly.

Related Paragraphs: 6,7,8,9,10,11,12,13,15,16,21,22,23,24,25

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 2
PS=6, IR=5

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. Information Content: stating information or point of view

Actual	Target	Description
____	5	clear, not succinct, not stated actively, mostly goal directed.

2. Emotional Content: expressing one's feelings and emotions

Actual	Target	Description
____	5	+/- emotion is nearly always expressed constructively, dealt with rather than held back, sometimes not succinct.

3. Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.

Actual	Target	Description
____	5	usually appropriate to goals, assertive, gradually increased; not always friendly and shared.

Related Paragraphs: 1,2,7,9,10,11,12,24,25,26,27,28,30

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	5	correct type and intensity are understood but may not be clearly fed back.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	5	clear, not rambling; assertive but not aggressive; usually facilitates discussion, may be a bit disruptive; some use of metaphor, but may parroting.

Related Paragraphs:
3,4,5,6,7,8,9,10,11,12,13,14,16,17,18,19,22,23,24,25,26,27,28,29,30,31,32,33

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	5	clear, not succinct, open-ended disintegrating into closed (yes/no) questions.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	5	assertive, may not always facilitate the discussion, may not always be friendly.

Related Paragraphs:
1,3,4,5,6,7,8,10,12,13,14,16,17,19,22,23,24,25,31,32

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	5	clear, not succinct, may not be motivating, usually does not solicit feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	5	tentative, not concrete, gradually increased; not always friendly.

Related Paragraphs: 7,9,10,12,16,17,18,24,25,26,27,28,29,30

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	3	goal (purpose) or info. vague, stated either actively or passively.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	3	+/- emotion is sometimes not expressed constructively by being expressed aggressively or passively.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	3	sometimes inappropriate in depth, amount or duration, or irrelevant to target person, situation or detracting from the ongoing nature of the relationship.

Related Paragraphs: 1,2,3,4,5,6,7,8,9,10,11,13,14,15

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	3	may be a bit off in stating correct type, or intensity fed back is clearly off target, or may miss opportunity to respond with empathy.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	3	may be rambling; may be aggressive; may be a disruptive; may parrot, little use of metaphor/humor.

Related Paragraphs: 2,3,4,5,6,7,8,9,10,12,16,14,15,16

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	3	vague or repeated closed questions and statements, positive or negative emotions color info. gathering.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	3	sometimes aggressive or passive, sometimes controls and dominates discussion with questions, not always friendly.

Related Paragraphs: 1,4,10,11,12,13

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	3	vague or disruptive, not motivating or negatively motivating, may resist feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	3	sometimes aggressive or passive, may not be tentative or gradually increased; not always friendly.

Related Paragraphs: 1,3,4,5,6,7,8,9,10,11,12,13,14,15

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	6	clear, not succinct, stated actively, goal directed.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	6	+/- emotion is usually genuine, always expressed constructively, dealt with rather than held back.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	6	appropriate to goals, gradually increased; friendly, not always shared.

Related Paragraphs:
1,3,4,5,6,7,8,10,11,12,13,14,15,16,17,18,19,20,22,23,24,25

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	6	clear, not succinct, stated actively .

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	6	clearly, but not succinctly states correct type and intensity.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	6	clear, not rambling; assertive but not aggressive; facilitates discussion rather than disrupting; sometimes uses metaphor/humor, but does not parroting.

Related Paragraphs:
3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	6	clear, not succinct, stated actively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	6	clearly, but not succinctly states correct type and intensity.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	6	sometimes aggressive or passive, sometimes controls and dominates discussion with questions, not always friendly.

Related Paragraphs: 1,2,5,6,7,8,10,12,14,15,20

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	6	clear, not succinct, stated actively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	6	clear, not succinct, motivating, may not solicit feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	6	tentative, not concrete, gradually increased; friendly.

Related Paragraphs: 6,8,10,11,12,13,14,17,18,19,20,21,22,23,24,25

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 5
PS=5, IR=5

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	5	clear, not succinct, not stated actively, mostly goal directed.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	5	+/- emotion is nearly always expressed constructively, dealt with rather than held back, sometimes not succinct.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	5	sometimes inappropriate in depth, amount or duration, or irrelevant to target person, situation or detracting from the ongoing nature of the relationship.

Related Paragraphs:
1,2,3,4,5,7,8,10,11,12,13,14,15,16,17,18,19,20,24,25,27,28

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	5	correct type and intensity are understood but may not be clearly fed back.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	5	clear, not rambling; assertive but not aggressive; usually facilitates discussion, may be a bit disruptive; some use of metaphor, but may parroting.

Related Paragraphs:

2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	5	clear, not succinct, open-ended disintegrating into closed (yes/no) questions.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	5	assertive, may not always facilitate the discussion, may not always be friendly.

Related Paragraphs: 1,2,3,4,9,10,12,19

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	5	clear, not succinct, may not be motivating, usually does not solicit feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	5	tentative, not concrete, gradually increased; not always friendly.

Related Paragraphs:
2,3,4,5,7,8,10,11,12,13,14,15,16,17,19,20,24,25,26,27,28,29

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 6
PS=5, IR=3

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	3	goal (purpose or info. vague, stated either actively or passively.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	3	+/- emotion is sometimes not expressed constructively by being expressed aggressively or passively.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	3	sometimes inappropriate in depth, amount or duration, or irrelevant to target person, situation or detracting from the ongoing nature of the relationship.

Related Paragraphs: 1,2,3,5,10,11,12,13,14,19,22,23,24,25,26

Quality Rating Scale	"+" Topic directly and completely covered "0" Topic covered, but weakly or indirectly "-" Topic covered incorrectly " " (blank) Topic not covered

Script 6
PS=5, IR=3

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	3	may be a bit off in stating correct type, or intensity fed back is clearly off target, or may miss opportunity to respond with empathy.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	3	may be rambling; may be aggressive; may be a disruptive; may parrot, little use of metaphor/humor.

Related Paragraphs:
2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,

Quality Rating Scale	"+" Topic directly and completely covered "0" Topic covered, but weakly or indirectly "-" Topic covered incorrectly " " (blank) Topic not covered

Script 6
PS=5, IR=3

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	3	vague or repeated closed questions and statements, positive or negative emotions color info. gathering.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	3	sometimes aggressive or passive, sometimes controls and dominates discussion with questions, not always friendly.

Related Paragraphs: 2,3,4,5,6,7,8,15,16

Quality Rating Scale	"+" Topic directly and completely covered "0" Topic covered, but weakly or indirectly "-" Topic covered incorrectly " " (blank) Topic not covered

Script 6
PS=5, IR=3

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	3	vague or disruptive, not motivating or negatively motivating, may resist feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	3	sometimes aggressive or passive, may not be tentative or gradually increased; not always friendly.

Related Paragraphs: 4,5,6,9,10,12,13,14,17,18,19,20,21,22,23,25,26

Quality Rating Scale
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 7
PS=3, IR=6

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	6	clear, not succinct, stated actively, goal directed.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	6	+/- emotion is usually genuine, always expressed constructively, dealt with rather than held back.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	6	appropriate to goals, gradually increased; friendly, not always shared.

Related Paragraphs: 1,2,3,4,6,7,9,11,14,16

12

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 7
PS=3, IR=6

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	6	clear, not succinct, stated actively .

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	6	clearly, but not succinctly states correct type and intensity.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	6	clear, not rambling; assertive but not aggressive; facilitates discussion rather than disrupting; sometimes uses metaphor/humor, but does not parroting.

Related Paragraphs: 2,3,4,5,6,7,8,9,11,12,13,14,15

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 7
PS=3, IR=6

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	6	clear, not succinct, stated actively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	6	clear, not succinct, open-ended questions and statements.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	6	assertive, facilitates the discussion, friendly.

Related Paragraphs: 1,2,3,4,5,8,9,12,14,

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 7
PS=3, IR=6

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	6	clear, not succinct, stated actively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	6	clear, not succinct, motivating, may not solicit feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	6	tentative, not concrete, gradually increased, friendly.

Related Paragraphs: 1,3,6,9,10,11,13,15

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 8
PS=3, IR=5

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	5	clear, not succinct, not stated actively, mostly goal directed.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	5	+/- emotion is nearly always expressed constructively, dealt with rather than held back, sometimes not succinct.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	5	usually appropriate to goals, assertive, gradually increased; not always friendly and shared.

Related Paragraphs: 1,2,3,8,9,10,12,13,14,15,17,18,22,23,24

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	5	correct type and intensity are understood but may not be clearly fed back.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	5	clear, not rambling; assertive but not aggressive; usually facilitates discussion, may be a bit disruptive; some use of metaphor, but may parroting.

Related Paragraphs:
2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25

Quality Rating Scale	"+" Topic covered directly and completely
	"0" Topic covered, but weakly or indirectly
	"-" Topic not covered
	" " (blank) Topic not covered

Script 8
PS=3, IR=5

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	5	clear, not succinct, open-ended disintegrating into closed (yes/no) questions.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	5	assertive, may not always facilitate the discussion, may not always be friendly.

Related Paragraphs: 1,2,4,5,6,7,10,12,13,14,15,17,18,19,20

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	5	clear, not succinct, not stated actively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	5	clear, not succinct, may not be motivating, usually does not solicit feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	5	tentative, not concrete, gradually increased; not always friendly.

Related Paragraphs: 3,10,11,12,15,22,23,24

15

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 9
PS=3, IR=3

SCRIPT VERIFICATION WORK SHEET
INTERPERSONAL RELATIONS

Rater's Name:_____

A. Self Expression

Disclosing or communicating one's point of view or other information by expressing experiences, behaviors and feelings through words, actions and nonverbal.

1. **Information Content: stating information or point of view**

Actual	Target	Description
____	3	vague, stated either actively or passively, goal directed.

2. **Emotional Content: expressing one's feelings and emotions**

Actual	Target	Description
____	3	+/- emotion is sometimes not expressed constructively by being expressed aggressively or passively.

3. **Delivery: disclosing information or feelings that is appropriate in depth and amount to the goal of the conversation.**

Actual	Target	Description
____	3	sometimes inappropriate in depth, amount or duration, or irrelevant to target person, situation or detracting from the ongoing nature of the relationship.

Related Paragraphs:
1,2,3,5,8,9,10,11,12,14,15,16,18,19,22,23,24,25,26,27,28,29,30,31,32

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

B. Responding with Empathy

Communicating accurate understanding of the core meaning of another person's words.

1. **Information Content: expressing the core meaning of another person's words.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: expressing the correct type of another person's feelings and emotions**

Actual	Target	Description
____	3	may be a bit off in stating correct type, or intensity fed back is clearly off target, or may miss opportunity to respond with empathy.

3. **Delivery: expressing accurate understanding of another person's thoughts or feelings in a way that is appropriate in depth and amount to the information and feelings expressed.**

Actual	Target	Description
____	3	may be rambling; may be aggressive; may be a disruptive; may parrot, little use of metaphor/humor.

Related Paragraphs:
2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

C. Probing and Questioning

Gathering information from another person.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: expressing feelings and emotions when asking questions or making probing statements.**

Actual	Target	Description
____	3	vague or repeated closed questions and statements, positive or negative emotions color info. gathering.

3. **Delivery: asking questions and making probing statements that are facilitative and appropriate in depth and amount to the ongoing nature of the relationship.**

Actual	Target	Description
____	3	sometimes aggressive or passive, sometimes controls and dominates discussion with questions, not always friendly.

Related Paragraphs: 3,4,5,6,7,8,9,10,12,15,18,20,21,25,28,30,31

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

D. Challenging

After establishing a positive interpersonal relationship, providing feedback to a person on past performance and seeking to motivate them to improve future performance by describing the person's strengths, noting actions inconsistent with standards, explaining implications or conclusions that were unrecognized, or confronting the person to examine their actions more closely.

1. **Information Content: asking questions or making probing statements that are clear and understandable.**

Actual	Target	Description
____	3	a bit off target or vague, stated either actively or passively.

2. **Emotional Content: seeking to create feelings and emotions when challenging.**

Actual	Target	Description
____	3	vague or disruptive, not motivating or negatively motivating, may resist feedback.

3. **Delivery: making challenging statements that are facilitative and appropriate in depth and amount to the problem to be resolved.**

Actual	Target	Description
____	3	sometimes aggressive or passive, may not be tentative or gradually increased; not always friendly.

Related Paragraphs:
8,9,10,12,15,16,17,18,19,21,22,23,24,25,26,27,29,30,31,32

Quality Rating Scale	"+" Topic covered directly and completely "0" Topic covered, but weakly or indirectly "-" Topic covered incorectly " " (blank) Topic not covered

Script 1
PS=6, IR=6

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	6	clearly explains the normal, planned or expected actions and mentions all but the minor deviations from them.

Quality	Deviations Mentioned (2,3,9,16,17)	Paragraphs
+	1. problem client	(2,3,17)
+	2. failed to call when sick	(9)
+	3. sick too often/absenteeism	(16)
+	4. socialize too much	(16)
+	5. performance deviation from old appraisal	(16,17)
	6. requested transfer	()
+	7. alcohol at lunch	(16)
0	8. general increase in problems	(16,17)
+	9. tardiness/promptness	(16)

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	6	asks questions and persists in seeking the essential details relating to who was involved and who was affected and the impact of future relations and actions, but does not persist in seeking information on where, when, costs, and future impact details; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Details mentioned (1,2,3,5,7,8,9,10,12,17)	Paragraphs
+	1. who involved: Chester, secretary, director, client	(2,3,5,8,12)
+	2. who affected: client, director, Mary, new supervisor, unit staff	(1,2,3,7,9,10)
+	3. impact on future relations: Chester, director and client upset	(3,7)
+	4. impact on actions: client not served, went to director's house; relates to other problems	(3)
+	5. where: office/ Chester at home	(3,5)
+	6. when: last week Tuesday	(3,17)
	7. cost: dollar cost not specified	()
+	8. importance: high, client not served	(10,17)
+	9. what happened: description of the client/ service problem	(3,5,8)

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	6	states expected outcomes usually in practical and realistic terms; states risks and constraints realistically but may use the risks to convince the employee that the issue is important; mentions that the employee must retain control and the supervisor must be involved in resolving the problem.

Quality	Expectations (9,10,11,13,15,16,17,21,22,24)	Paragraphs
	Problem 1: improve productivity	
+	1. improve productivity	(16,17,21,22,24)
+	2. help the client/avoid problems in future	(9,10)
	3. Do the job expected of a PFW	()
	Problem 2: follow the work rules	
	1. following the rules	()
+	2. call in when sick	(9,10,11,13,21,24)
+	3. improve attendance	(11,13,16,17,21)
0	4. no alcohol on work site	(16)
0	5. decrease socializing	(16)
+	6. improve promptness	(13,16,21)
	Problem 3.: improve job satisfaction	
0	1. improve satisfaction/ work relationships	(16)
+	2. become promotable	(21,22)

Quality	People to be involved (11,15,22,24)	Paragraphs
+	1. Chester retain control of PS	(11,15,22,24)
+	2. tell Sup. if overworked	(24)

Quality	risks (8,13,14,24)	Paragraphs
+	1. oral reprimand	(8,24)
+	2. written reprimand	(8,24)
	3. one day suspension	()
	4. five day suspension	()
	5. fired	()
+	6. general consequences	(13,14)

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

B. Problem Analysis (continued)

2. Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	6	asks questions seeking the cause of the problem and all of the essential details (i.e., who, what, where, when, why and how); may not seek interrelationships among causal details or among deviation facts.

Quality	Diagnosing (6,7,10,16,17,19,20,21)	Paragraphs
+	1. Who caused the problem: Chester	(6,10,16)
+	2. what caused the problem: Promotion issue with Stan	(16,19,21)
+	3. when was the problem caused: six weeks ago	(16,19,20)
+	4. where was the problem caused: on the unit, and in the selection processes	(6,7,10,16,21)
+	5. why was the problem caused: anger, frustration, burnout	(6,7,10,16,19,21)
+	6. how was the problem caused: work slow down	(6,16,17,21)
	7. seeking interrelationships among deviations and among causes	()

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	6	develops multiple possible solutions for most of the identified problem, each problem issue has at least one possible solution mentioned; seeks but does not persist in obtaining help in possibility development; returns to earlier problem solving steps if necessary.

Number	Possible Solutions (10,11,21,22,24)	Paragraphs
	Problem 1: improve productivity	
3 +	1. assess service/accuracy/productivity	(22,24)
	2. review workload	()
	Problem 2: follow the work rules	
4 +	1. come to work on time/call in when sick	(11,21,24)
	2. cut down on socializing	()
	3. follow the rules	()
	Problem 3: improve job satisfaction	
4 +	1. develop plan for promotion	(21,22)
	2. reassess career with County	()
	3. improve job satisfaction	()
	General Supervisory Actions	
2 +	1. seeks assistance from Chester	(10,11)
1 +	2. ask for help from supervisor	(24)
1 +	3. meet again	(22)
	4. special review process	()

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 1
PS=6, IR=6

C. **Decision Making**(continued)

2. Establishing criteria: stating a means for appraisal of possible solutions

Actual	Target	Description
____	6	discusses criteria at the time the possible solutions are developed; discussion includes whether or not the problem will be resolved, and the amount of time required.

Quality	Criteria (21,22,24)	Paragraphs
	Problem 1: improve productivity	
+	1. provide proper service to clients	(21,22,24)
	Problem 2: follow the work rules	
+	1. come to work on time	(21,22,24)
+	2. call in when sick	(21,22,24)
	3. follow rules on socializing	()
	4. follow rules on drinking	()
+	5. face discipline or resolve problems	(24)
	6. general following of rules	()
	Problem 3: improve job satisfaction	
+	1. become promotable	(21,22)
	2. improve satisfaction, without transfer	()
+	3. take the initiative to improve	(24)

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	6	clearly states the decision rules that are used to evaluate possible solutions; appraisal may not be systematic and thorough; appraises only preselected solutions by stating the degree to which they meet the criteria (i.e., the "If this, then this" logic is clear).

Quality	Appraisal (21)	Paragraphs
	Problem 1: improve productivity	
+	1. If we assess needed skills, plan for developing them and practice them, then the goals of being productive and promotable will be reached.	(21)
	Problem 2: follow the work rules	
+	1. It's going to take more than just coming to work on time or calling in sick to resolve this problem.	(21)
	Problem 3: improve job satisfaction	
+	1. If we assess needed skills, plan for developing them and practice them, then the goals of being productive and promotable will be reached.	(21)

Quality Rating Scale: "+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" "

Script 1
PS=6, IR=6

C. Decision Making(continued)

4.choosing:making a decision to implement the best solution

Actual	Target	Description
____	6	decides on and specifies a course of action for the full set of problems but not based on a systematic and thorough evaluation.

Quality	Choosing (17,21,22,24,25)	Paragraphs
	Problem 1: improve productivity	
	1. review workload	()
+	2. improve accuracy/ productivity.	(21,22,24)
	Problem 2: follow the work rules	
+	1. improve attendance/promptness.	(21,22,24)
	Problem 3: improve job satisfaction	
+	1. work on resolving promotion problems	(21,22)
	2. review career with County	()
	General supervisory actions	
+	1. meet again to work on problems and goals	(17,22,25)
+	2. use of the discipline process	(24)
	3. special review processes	()

D. Action

1.Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	6	takes specific actions to resolve the crisis issues and the most important identified problems, discusses but does not necessarily take specific actions for all issues discussed.

Quality	Implementing (22,24,25)	Paragraphs
	Problem 1: improve productivity	
+	1. accuracy and productivity	(22,24)
	2. workload distribution	()
	Problem 2: follow the work rules	
+	1. attendance and promptness	(22,24)
	Problem 3: improve job satisfaction	
+	1. promotion preparedness	(22)
	General supervisory actions	
+	1. meet again	(22,25)
+	2. disciplinary process	(24)
0	3. assist in locating and resolving problems	(24)

Quality Rating Scale: "+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " -

Script 1
PS=6, IR=6

E. Supervision

1.Supervision: planning, organizing and motivating implementation

Actual	Target	Description
____	6	states a general plan and directs action toward the goal, provides some positive motivation.

Quality	Ensuring (16,21,22,23,24,25)	Paragraphs
+	1. states plan	(21,23)
+	2. directs	(22,23,24,25)
+	3. motivates	(16,22,23,24)

F. Problem Identification(second time)

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	6	initiates a monitoring plan, may not insure that the actions to be taken are understood.

Quality	Monitoring (24)	Paragraphs
+	1. initiates plan	(24)

Quality Rating Scale:
"+" Topic Directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " Topic not covered

Script 2
PS=6, IR=5

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	6	clearly explains the normal, planned or expected actions and mentions all but the minor deviations from them.

Quality	Deviations Mentioned (1,7,28,29)	Paragraphs
+	1. problem client	(1)
+	2. fails to call when sick	(29)
+	3. sick too often/absenteeism	(1)
+	4. socialize too much	(7)
+	5. performance deviation from old appraisal	(7,28)
+	6. requested transfer	(7)
+	7. alcohol at lunch	(7)
+	8. general increase in problems	(1)
+	9. tardiness/promptness	(29)

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	6	asks questions and persists in seeking the essential details relating to who was involved and who was affected and the impact of future relations and actions, but does not persist in seeking information on where, when, costs, and future impact details; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Details mentioned (1,2,3,4,5,6,7,8,19,23,26,28,29)	Paragraphs
+	1. who involved: Chester, secretary, director, client	(2,5,6,7,19,23)
+	2. who affected: client, director, Mary, new supervisor, unit staff	(2,5,6,7,19,23,26)
0	3. impact on future relations: Chester, director and client upset	(2)
+	4. impact on actions: client not served, went to director's house; relates to other problems	(5,23,26,28)
+	5. where: office/ Chester at home	(2,3)
+	6. when: last week Tuesday	(1)
	7. cost: dollar cost not specified	()
+	8. importance: high, client not served	(1,2,26,28,29)
+	9. what happened: description of the client/ service problem	(2,3,4,5,8)

Quality Rating Scale: "+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	6	states expected outcomes usually in practical and realistic terms; states risks and constraints realistically but may use the risks to convince the employee that the issue is important; mentions that the employee must retain control and the supervisor must be involved in resolving the problem.

Quality	Expectations (7,9,12,25,26,27,28,29)		Paragraphs
	Problem 1: improve productivity		
+	1.	improve productivity	(7,28,29)
0	2.	help the client/avoid problems in future	(28,29)
	3.	Do the job expected of a PFW	()
	Problem 2: follow the work rules		
	1.	following the rules	()
+	2.	call in when sick	(29)
+	3.	improve attendance	(12,25,26,27,29)
0	4.	no alcohol on work site	(7)
0	5.	decrease socializing	(7)
+	6.	improve promptness	(29)
	Problem 3.: improve job satisfaction		
0	1.	improve satisfaction/ work relationships	(12,27,28,29)
+	2.	become promotable	(9,26,29)

Quality	People to be involved ()	Paragraphs
+	1. Chester retain control of PS	(12,25,26)
+	2. tell Sup. if overworked	(24,26)

Quality	risks (27,28,29)		Paragraphs
0	1.	oral reprimand	(27)
+	2.	written reprimand	(27)
	3.	one day suspension	()
	4.	five day suspension	()
+	5.	fired	(27)
+	6.	general consequences	(27,28,29)

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	6	asks questions seeking the cause of the problem and all of the essential details (i.e., who, what, where, when, why and how); may not seek interrelationships among causal details or among deviation facts.

Quality	Diagnosing (5,7,8,9,13,14,16,17,18,28)	Paragraphs
+	1. Who caused the problem: Chester	(16)
+	2. what caused the problem: Promotion issue with Stan	(9,16,17,18,19)
+	3. when was the problem caused: six weeks ago	(8)
+	4. where was the problem caused: on the unit, and in the selection processes	(14)
+	5. why was the problem caused: anger, frustration, burnout	(13,14,19,28)
+	6. how was the problem caused: work slow down	(5,7,13)
	7. seeking interrelationships among deviations and among causes	()

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	6	develops multiple possible solutions for most of the identified problem, each problem issue has at least one possible solution mentioned; seeks but does not persist in obtaining help in possibility development; returns to earlier problem solving steps if necessary.

Number	Alternatives (10,12,17,18,24,25,26,27,28,29,31)	Paragraphs
	Problem 1: improve productivity	
4 +	1. assess service/accuracy/productivity	(26,27,28,29)
4 +	2. review workload	(26)
	Problem 2: follow the work rules	
3 +	1. come to work on time/call in when sick	(26,29)
	2. cut down on socializing	()
	3. follow the rules	()
	Problem 3: improve job satisfaction	
4 +	1. develop plan for promotion	(10,17,18,26)
3 +	2. reassess career with County	(27,29)
	3. improve job satisfaction	()
	General Supervisory Actions	
6 +	1. seeks assistance from Chester	(12,24,25,27)
2 +	2. ask for help from supervisor	(24,25,27,28)
3 +	3. meet again	(27,31)
	4. special review process	()

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

C. **Decision Making**(continued)

2. **Establishing criteria: stating a means for appraisal of possible solutions**

Actual	Target	Description
____	6	discusses criteria at the time the possible solutions are developed; discussion includes whether or not the problem will be resolved, and the amount of time required.

Quality	Criteria (9,12,25,27,28,29)	Paragraphs
	Problem 1: improve productivity	
+	1. provide proper service to clients	(28,29)
	Problem 2: follow the work rules	
+	1. come to work on time	(29)
+	2. call in when sick	(29)
	3. follow rules on socializing	()
	4. follow rules on drinking	()
+	5. face discipline or resolve problems	(27,29)
+	6. general following of rules	(27)
	Problem 3: improve job satisfaction	
0	1. become promotable	(9)
	2. improve satisfaction, without transfer	()
+	3. take the initiative to improve	(12,25,27)

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

C. **Decision Making**(continued)

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	6	clearly states the decision rules that are used to evaluate possible solutions; appraisal may not be systematic and thorough; appraises only preselected solutions by stating the degree to which they meet the criteria (i.e., the "If this, then this" logic is clear).

Quality	Appraisal (27,29)	Paragraphs
	Problem 1: improve productivity	
+	1. If we develop a contract specifying those things Chester is required to do, the expectations will be met; otherwise the discipline process will be used	(27,29)
	Problem 2: follow the work rules	
+	1. If we develop a contract specifying those things Chester is required to do, the expectations will be met; otherwise the discipline process will be used	(27,29)
	Problem 3: improve job satisfaction	
0	1. If we develop a contract specifying those things Chester is required to do, the expectations will be met; otherwise the discipline process will be used	(27,29)

4. choosing: making a decision to implement the best solution

Actual	Target	Description
____	6	decides on and specifies a course of action for the full set of problems but not based on a systematic and thorough evaluation.

Quality	Choosing (26,27,29,30)	Paragraphs
	Problem 1: improve productivity	
+	1. review workload	(26,30)
+	2. improve accuracy/ productivity.	(26,27,29)
	Problem 2: follow the work rules	
+	1. improve attendance/promptness.	(26,29)
	Problem 3: improve job satisfaction	
0	1. work on resolving promotion problems	(26)
+	2. review career with County	(26)
	General supervisory actions	
+	1. meet again to work on problems and goals	(26,27)
+	2. use of the discipline process	(27)
+	3. special review processes	(27)

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 2
PS=6, IR=5

D. Action

1.Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	6	takes specific actions to resolve the crisis issues and the most important identified problems, discusses but does not necessarily take specific actions for all issues discussed.

Quality	Implementing (26,27,28,30,31)	Paragraphs
	Problem 1: improve productivity	
+	1. accuracy and productivity	(28)
+	2. workload distribution	(27,30)
	Problem 2: follow the work rules	
+	1. attendance and promptness	(29)
	Problem 3: improve job satisfaction	
0	1. promotion preparedness	(26,31)
	General supervisory actions	
+	1. meet again	(27,30,31)
+	2. disciplinary process	(27,28)
0	3. assist in locating and resolving problems	(27)

E. Supervision

1.Supervision: planning, organizing and motivating implementation

Actual	Target	Description
____	6	states a general plan and directs action toward the goal, provides some positive motivation.

Quality	Ensuring (10,12,24,26,28,31,33)	Paragraphs
+	1. states plan	(26,27,31)
+	2. directs	(24,26,28,31)
0	3. motivates	(10,12,28,33)

F. Problem Identification(second time)

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	6	initiates a monitoring plan, may not insure that the actions to be taken are understood.

Quality	Monitoring (26,28)	Paragraphs
+	1. initiates plan	(26,28)

Quality Rating Scale:
"+" Topic Directly and Completely covered
"0+ Topic covered weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	6	clearly explains the normal, planned or expected actions and mentions all but the minor deviations from them.

Quality	Deviations Mentioned (1,3,4,5,7)	Paragraphs
+	1. problem client	(1,3,7)
+	2. fails to call when sick	(1,3)
+	3. sick too often/absenteeism	(1)
	4. socialize too much	()
+	5. performance deviation from old appraisal	(4,5,7)
+	6. requested transfer	(4)
+	7. alcohol at lunch	(1,4)
+	8. general increase in problems	(1)
+	9. tardiness/promptness	(1)

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	6	asks questions and persists in seeking the essential details relating to who was involved and who was affected and the impact of future relations and actions, but does not persist in seeking information on where, when, costs, and future impact details; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Details mentioned (1,3,5,7,15)	Paragraphs
+	1. who involved: Chester, secretary, director, client	(1,3,7)
+	2. who affected: client, director, Mary, new supervisor, unit staff	(1,3,7,)
0	3. impact on future relations: Chester, director and client upset	(3,7)
+	4. impact on actions: client not served, went to director's house; relates to other problems	(3,7)
+	5. where: office/ Chester at home	(1)
+	6. when: last week Tuesday	(1)
	7. cost: dollar cost not specified	()
+	8. importance: high, client not served	(3,5,7,15)
+	9. what happened: description of the client/ service problem	(1,3)

Quality RatingScale: "+" Topic directly and completely covered

Script 3
PS=6, IR=3

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	6	states expected outcomes usually in practical and realistic terms; states risks and constraints realistically but may use the risks to convince the employee that the issue is important; mentions that the employee must retain control and the supervisor must be involved in resolving the problem.

Quality	Expectations (1,3,4,5,6,7,8,9,10,14,15)	Paragraphs
	Problem 1: improve productivity	
+	1. improve productivity	(6,7)
+	2. help the client/avoid problems in future	(5,7)
+	3. Do the job expected of a PFW	(4,5)
	Problem 2: follow the work rules	
+	1. following the rules	(7,10)
+	2. call in when sick	(3,7,8,10,14)
+	3. improve attendance	(6,7,8,14)
0	4. no alcohol on work site	(1)
	5. decrease socializing	()
	6. improve promptness	()
	Problem 3.: improve job satisfaction	
	1. improve satisfaction/ work relationships	()
0	2. become promotable	(5,7)

Quality	People to be involved (3,14,15)	Paragraphs
+	1. Chester retain control of PS	(14,15)
+	2. tell Sup. if overworked	(3,15)

Quality	risks (3,5,6,8,9,10)	Paragraphs
0	1. oral reprimand	(8)
+	2. written reprimand	(8)
+	3. one day suspension	(8)
+	4. five day suspension	(8,9)
+	5. fired	(9,10)
+	6. general consequences	(3,5,6,8)

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	6	asks questions seeking the cause of the problem and all of the essential details (i.e., who, what, where, when, why and how); may not seek interrelationships among causal details or among deviation facts.

Quality	Diagnosing (1,4,5,7,15)	Paragraphs
+	1. Who caused the problem: Chester	(1)
+	2. what caused the problem: Promotion issue with Stan	(1,4,5)
+	3. when was the problem caused: six weeks ago	(1)
+	4. where was the problem caused: on the unit, and in the selection processes	(1,7)
+	5. why was the problem caused: anger, frustration, burnout	(4,5,15)
+	6. how was the problem caused: work slow down	(4,5)
	7. seeking interrelationships among deviations and among causes	()

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	6	develops multiple possible solutions for most of the identified problem, each problem issue has at least one possible solution mentioned; seeks but does not persist in obtaining help in possibility development; returns to earlier problem solving steps if necessary.

Number	Alternatives (3,4,5,7,11,12,13,14,15)	Paragraphs
	Problem 1: improve productivity	
3 +	1. assess service/accuracy/productivity	(4,5,14)
1 +	2. review workload	(15)
	Problem 2: follow the work rules	
8 +	1. come to work on time/call in when sick	(3,7,11,12,13,14)
	2. cut down on socializing	()
1 +	3. follow the rules	(14)
	Problem 3: improve job satisfaction	
2 +	1. develop plan for promotion	(5,7)
	2. reassess career with County	()
	3. improve job satisfaction	()
	General Supervisory Actions	
2 +	1. seeks assistance from Chester	(14,15)
4 +	2. ask for help from supervisor	(4,7,14)

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

3 +	3. meet again	(14)
2 +	4. special review process	(14)

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

C. **Decision Making**(continued)

2. Establishing criteria: stating a means for appraisal of possible solutions

Actual	Target	Description
____	6	discusses criteria at the time the possible solutions are developed; discussion includes whether or not the problem will be resolved, and the amount of time required.

Quality	Criteria (3,5,7,8,9,10,11,12,13,14,15)	Paragraphs
	Problem 1: improve productivity	
+	1. provide proper service to clients	(5,7)
	Problem 2: follow the work rules	
+	1. come to work on time	(7,11,14)
+	2. call in when sick	(3,7,11,12,13,14)
	3. follow rules on socializing	()
	4. follow rules on drinking	()
+	5. face discipline or resolve problems	(8,9,14)
+	6. general following of rules	(10,14)
	Problem 3: improve job satisfaction	
0	1. become promotable	(7)
	2. improve satisfaction, without transfer	()
+	3. take the initiative to improve	(14,15)

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

C. Decision Making(continued)

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	6	clearly states the decision rules that are used to evaluate possible solutions; appraisal may not be systematic and thorough; appraises only preselected solutions by stating the degree to which they meet the criteria (i.e., the "If this, then this" logic is clear).

Quality	Appraisal (6,7,14)	Paragraphs
+	Problem 1: improve productivity	
	1. Chester's work must be done on time or he will face discipline	(6,14)
	Problem 2: follow the work rules	
+	1. Chester must correct his excessive absenteeism, improve use of sick leave and failure to report when sick, or he will face discipline	(6)
	Problem 3: improve job satisfaction	
0	1. Some other time we can get together and discuss your inability to compete successfully for promotions, and that you don't like training or can't work fast enough or find time for it	(7)

4. choosing: making a decision to implement the best solution

Actual	Target	Description
____	6	decides on and specifies a course of action for the full set of problems but not based on a systematic and thorough evaluation.

Quality	Choosing (5,7,9,10,11,12,13,14,15)	Paragraphs
	Problem 1: improve productivity	
+	1. review workload	(15)
+	2. improve accuracy/ productivity.	(5,14,15)
	Problem 2: follow the work rules	
+	1. improve attendance/promptness.	(10,11,12,13,14)
	Problem 3: improve job satisfaction	
0	1. work on resolving promotion problems	(5,7)
	2. review career with County	()
	General supervisory actions	
+	1. meet again to work on problems and goals	(14,15)
+	2. use of the discipline process	(9,10,14)
+	3. special review processes	(14,15)

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 3
PS=6, IR=3

D. Action

1. Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	6	takes specific actions to resolve the crisis issues and the most important identified problems, discusses but does not necessarily take specific actions for all issues discussed.

Quality	Implementing (7,11,12,13,14,15)	Paragraphs
	Problem 1: improve productivity	
+	1. accuracy and productivity	(11,12,13,14)
0	2. workload distribution	(15)
	Problem 2: follow the work rules	
+	1. attendance and promptness	(11,12,13,14)
	Problem 3: improve job satisfaction	
0	1. promotion preparedness	(7)
	General supervisory actions	
+	1. meet again	(14,15)
+	2. disciplinary process	(14,15)
0	3. assist in locating and resolving problems	(14,15)

E. Supervision

1. Supervision: planning, organizing and motivating implementation

Actual	Target	Description
____	6	states a general plan and directs action toward the goal, provides some positive motivation.

Quality	Ensuring (7,8,10,11,12,13,14,15)	Paragraphs
+	1. states plan	(7,14,15)
+	2. directs	(11,12,13,14)
0	3. motivates	(7,8,10,11,12,13,14)

F. Problem Identification (second time)

1. Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	6	initiates a monitoring plan, may not insure that the actions to be taken are understood.

Quality	Monitoring (14,15)	Paragraphs
+	1. initiates plan	(14,15)

Quality RatingScale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	5	explains the normal, planned and expected actions and mentions the major deviation and some other secondary ones.

Quality	Deviations Mentioned(1,3,4,7,8,10,11,14,20,23)	Paragraphs
+	1. problem client	(10,11)
+	2. fails to call when sick	(10,11)
+	3. sick too often/absenteeism	(7,8,10,20)
	4. socialize too much	()
+	5. performance deviation from old appraisal	(3,4,8,18,20)
	6. requested transfer	()
+	7. alcohol at lunch	(14,23)
0	8. general increase in problems	(1)
+	9. tardiness/promptness	(7,20)

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	5	asks questions but does not always persist in seeking the essential details relating to deviation participants and the impact; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Details mentioned (1,3,4,6,9,10,11,13,22,23)	Paragraphs
+	1. who involved: Chester, secretary, director, client	(9,10,11,23)
0	2. who affected: client, director, Mary, new supervisor, unit staff	(3,6,9,10,23)
0	3. impact on future relations: Chester, director and client upset	(8,10)
+	4. impact on actions: client not served, went to director's house; relates to other problems	(10,11)
0	5. where: office/ Chester at home	(10)
0	6. when: last week Tuesday	(10,22)
	7. cost: dollar cost not specified	()
0	8. importance: high, client not served	(1,3,4,13)
+	9. what happened: description of the client/ service problem	(10,11)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	5	states expected outcomes but not typically in practical and realistic terms; states risks and constraints but not practically nor realistically ; mentions that the employee must retain control but does not mention that the supervisor must be kept informed.

Quality	Expectations (6,10,11,12,14,15,16,18,20,22,23,24)		Paragraphs
	Problem 1: improve productivity		
+	1.	improve productivity	(18,20)
0	2.	help the client/avoid problems in future	(10)
0	3.	Do the job expected of a PFW	(6)
	Problem 2: follow the work rules		
+	1.	following the rules	(25)
+	2.	call in when sick	(10,11,12,22)
+	3.	improve attendance	(12,20)
+	4.	no alcohol on work site	(14,15,23)
	5.	decrease socializing	()
+	6.	improve promptness	(12,20)
	Problem 3.: improve job satisfaction		
0	1.	improve satisfaction/ work relationships	(24)
+	2.	become promotable	(12,16,18,22)

Quality	People to be involved (16,17,24)		Paragraphs
	1.	Chester retain control of PS	()
0	2.	tell Sup. if overworked	(16,17,24)

Quality	risks (22)		Paragraphs
	1.	oral reprimand	()
+	2.	written reprimand	(22)
	3.	one day suspension	()
	4.	five day suspension	()
	5.	fired	()
+	6.	general consequences	(22)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	5	asks questions seeking what caused the problem and why the problem occurred, plus some other concerns.

Quality	Diagnosing(3,4,5,6,8,9,10,11,16,17,18,20,22)		Paragraphs
+	1.	Who caused the problem: Chester	(3,4,5,9,11)
+	2.	what caused the problem: Promotion issue with Stan	(5,6,8,9,16,17)
0	3.	when was the problem caused: six weeks ago	(4,16,18,20)
+	4.	where was the problem caused: on the unit, and in the selection processes	(9,10,11)
+	5.	why was the problem caused: anger, frustration, burnout	(6,8,9,10,16,17,22)
0	6.	how was the problem caused: work slow down	(6,8,9,11)
	7.	seeking interrelationships among deviations and among causes	()

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	5	develops multiple possible solutions for at least one identified problem, with all other issues having one possible solution mentioned; no assistance is sought; returns to earlier problem solving steps if necessary.

Number	Alternatives (12,13,17,18,22,24,25,)	Paragraphs
	Problem 1: improve productivity	
2 +	1. assess service/accuracy/productivity	(18,24)
2 +	2. review workload	(24,25)
	Problem 2: follow the work rules	
2 +	1. come to work on time/call in when sick	(12,13,22)
	2. cut down on socializing	()
	3. follow the rules	()
	Problem 3: improve job satisfaction	
2 +	1. develop plan for promotion	(17,18)
	2. reassess career with County	()
	3. improve job satisfaction	()
	General Supervisory Actions	
	1. seeks assistance from Chester	()
	2. ask for help from supervisor	()
2 +	3. meet again	(24,25)
1 +	4. special review process	(24)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

C. **Decision Making**(continued)

2. Establishing criteria: stating a means for appraisal of possible solutions

Actual	Target	Description
____	5	develops and presents criteria, but does not discuss; criteria may fail to evaluate an important aspect of a possible solution or may be unable to evaluate a possible solution; process is linked with development of possible solutions.

Quality	Criteria (11,12,14,15,17,18,20,22,23,24,25)	Paragraphs
	Problem 1: improve productivity	
+	1. provide proper service to clients	(18,20,24,25)
	Problem 2: follow the work rules	
+	1. come to work on time	(11,12)
+	2. call in when sick	(11,12)
	3. follow rules on socializing	()
+	4. follow rules on drinking	(14,15,23)
+	5. face discipline or resolve problems	(22)
+	6. general following of rules	(24,25)
	Problem 3: improve job satisfaction	
0	1. become promotable	(17,18)
	2. improve satisfaction, without transfer	()
	3. take the initiative to improve	()

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	5	Appraises possible solutions by evaluating them as they are developed; states decision rules when pressed. ("If this, then this" logic is not clear).

Quality	Appraisal (12,24,25)	Paragraphs
	Problem 1: improve productivity	
+	1. If after a month Chester's case load accuracy is back up to par, then the work load will be dealt with and some of the other problems discussed in the meeting can be resolved.	(24,25)
	Problem 2: follow the work rules	
0	1. Follow the rules on attendance and punctuality. Drinking is discouraged.	(12,14)
	Problem 3: improve job satisfaction	
0	1. If after a month Chester's case load accuracy is back up to par, then the work load will be dealt with and some of the other problems discussed in the meeting can be resolved.	(24,25)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

C. Decision Making(concluded)

4.choosing:making a decision to implement the best solution

Actual	Target	Description
____	5	decides on and specifies a course of action for the major identified problems.

Quality	Choosing (1,12,20,21,22,24,25)	Paragraphs
	Problem 1: improve productivity	
+	1. review workload	(21,24,25)
+	2. improve accuracy/ productivity.	(20,24)
	Problem 2: follow the work rules	
+	1. improve attendance/promptness.	(12,24)
	Problem 3: improve job satisfaction	
0	1. work on resolving promotion problems	(12,22)
	2. review career with County	()
	General supervisory actions	
0	1. meet again to work on problems and goals	(1,24)
+	2. use of the discipline process	(22)
0	3. special review processes	(24,25)

D. Action

1.Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	5	takes specific actions to resolve the crisis issues; discusses and takes some actions for the other issues discussed, but does not necessarily take specific actions for all issues discussed.

Quality	Implementing (20,21,22,24,25)	Paragraphs
	Problem 1: improve productivity	
0	1. accuracy and productivity	(20,24)
+	2. workload distribution	(24,25)
	Problem 2: follow the work rules	
+	1. attendance and promptness	(21,24,25)
	Problem 3: improve job satisfaction	
	1. promotion preparedness	()
	General supervisory actions	
0	1. meet again	(24)
+	2. disciplinary process	(22)
0	3. assist in locating and resolving problems	(24,25)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 4
PS=5, IR=6

E. Supervision

1.Supervision: planning, organizing and motivating implementation

Actual	Target	Description
____	5	states the plan as a goal, may do some directing, without organizing or motivating.

Quality	Ensuring (1,17,22,23,24,25,26)	Paragraphs
+	1. states plan	(1,24,25)
0	2. directs	(22,23,25)
0	3. motivates	(17,26)

F. Problem Identification(second time)

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	5	mentions a monitoring plan but does not necessarily initiate it.

Quality	Monitoring (24,25)	Paragraphs
0	1. initiates plan	(24,25)

Quality Rating Scale: "+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly

Script 5
PS=5, IR=5

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	5	explains the normal, planned and expected actions and mentions the major deviation and some other secondary ones.

Quality	Deviations Mentioned (2,3,5,7,8,9,10,11,12,15,19,24,27)	Paragraphs
+	1. problem client	(2,5,7,10)
+	2. fails to call when sick	(7,10,15,19)
+	3. sick too often/absenteeism	(7)
+	4. socialize too much	(10,11,12)
+	5. performance deviation from old appraisal	(2,3,8,12,24,27)
+	6. requested transfer	(9)
	7. alcohol at lunch	()
	8. general increase in problems	()
0	9. tardiness/promptness	(7)

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	5	asks questions but does not always persist in seeking the essential details relating to deviation participants and the impact; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Details mentioned (1,2,3,4,5,7,8,10,11)	Paragraphs
+	1. who involved: Chester, secretary, director, client	(1,2,5)
+	2. who affected: client, director, Mary, new supervisor, unit staff	(1,2,5,8)
0	3. impact on future relations: Chester, director and client upset	(5,7)
0	4. impact on actions: client not served, went to director's house; relates to other problems	(2,7)
+	5. where: office/ Chester at home	(2,3,4)
0	6. when: last week Tuesday	(2,3)
	7. cost: dollar cost not specified	()
0	8. importance: high, client not served	(2,5,7)
+	9. what happened: description of the client/ service problem	(1,5,7,10,11)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	5	states expected outcomes but not typically in practical and realistic terms; states risks and constraints but not practically nor realistically ; mentions that the employee must retain control but does not mention that the supervisor must be kept informed.

Quality	Expectations (4,5,10,11,12,13,14,15,16,19,20,24,25,26,27,28,29)	Paragraphs
	Problem 1: improve productivity	
+	1. improve productivity	(10,24,27,29)
	2. help the client/avoid problems in future	()
+	3. Do the job expected of a PFW	(10,24,27)
	Problem 2: follow the work rules	
0	1. following the rules	(28,29)
+	2. call in when sick	(4,5,14,15,16,19,20,26)
+	3. improve attendance	(25,27,29)
	4. no alcohol on work site	()
+	5. decrease socializing	(10,11)
	6. improve promptness	()
	Problem 3.: improve job satisfaction	
+	1. improve satisfaction/ work relationships	(12,13)
0	2. become promotable	(13)

Quality	People to be involved (27)	Paragraphs
	1. Chester retain control of PS	()
+	2. tell Sup. if overworked	(27)

Quality	risks (17)	Paragraphs
	1. oral reprimand	()
	2. written reprimand	()
	3. one day suspension	()
	4. five day suspension	()
	5. fired	()
+	6. general consequences	(17)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	5	asks questions seeking what caused the problem and why the problem occurred, plus some other concerns.

Quality	Diagnosing (3,9,12)		Paragraphs
+	1.	Who caused the problem: Chester	(3,9)
+	2.	what caused the problem: Promotion issue with Stan	(9,12)
	3.	when was the problem caused: six weeks ago	()
0	4.	where was the problem caused: on the unit, and in the selection processes	(3,9)
+	5.	why was the problem caused: anger, frustration, burnout	(9,12)
	6.	how was the problem caused: work slow down	()
	7.	seeking interrelationships among deviations and among causes	()

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	5	develops multiple possible solutions for at least one identified problem, with all other issues having one possible solution mentioned; no assistance is sought; returns to earlier problem solving steps if necessary.

Number		Alternatives	Paragraphs
		(4,5,7,10,11,12,13,14,15,24,25,26,27)	
		Problem 1: improve productivity	
4	0	1. assess service/accuracy/productivity	(12,24,26,27)
		2. review workload	()
		Problem 2: follow the work rules	
5	+	1. come to work on time/call in when sick	(4,5,7,13,14,15,25,26)
2	+	2. cut down on socializing	(10,11)
		3. follow the rules	()
		Problem 3: improve job satisfaction	
1	0	1. develop plan for promotion	(27)
		2. reassess career with County	()
3	+	3. improve job satisfaction	(12,13)
		General Supervisory Actions	
1	0	1. seeks assistance from Chester	(27)
1	+	2. ask for help from supervisor	(27)
		3. meet again	()
		4. special review process	()

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

C. Decision Making(continued)

2. Establishing criteria: stating a means for appraisal of possible solutions

Actual	Target	Description
____	5	develops and presents criteria, but does not discuss; criteria may fail to evaluate an important aspect of a possible solution or may be unable to evaluate a possible solution; process is linked with development of possible solutions.

Quality	Criteria (10,11,12,13,14,15,22,24,26,27)	Paragraphs
	Problem 1: improve productivity	
+	1. provide proper service to clients	(12,24,27)
	Problem 2: follow the work rules	
+	1. come to work on time	(27)
+	2. call in when sick	(14,15,22,26)
0	3. follow rules on socializing	(10,11)
	4. follow rules on drinking	()
0	5. face discipline or resolve problems	(14)
0	6. general following of rules	(12)
	Problem 3: improve job satisfaction	
	1. become promotable	()
0	2. improve satisfaction, without transfer	(13,14,27)
	3. take the initiative to improve	()

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	5	Appraises possible solutions by evaluating them as they are developed; states decision rules when pressed. ("If this, then this" logic is not clear).

Quality	Appraisal (12,13,14,15,17,26,27)	Paragraphs
	Problem 1: improve productivity	
+	1. It would be good to see you turn your performance around to where it was nine months ago; you'll need to keep it up. We'd like to see you work as a team member. We'll be monitoring your work accuracy and promptness.	(12,27)
	Problem 2: follow the work rules	
+	1. You are expected to call in if you are not going to be here. It would be good to get your attendance and . . . Let's not get to the disciplinary process.	(14,15,17,26,27)
	Problem 3: improve job satisfaction	
0	1. It would be good to see you be more satisfied with your job.	(13)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

C. Decision Making(concluded)

4.choosing:making a decision to implement the best solution

Actual	Target	Description
____	5	decides on and specifies a course of action for the major identified problems.

Quality	Choosing (13,14,16,24,26,27)	Paragraphs
	Problem 1: improve productivity	
0	1. review workload	(27)
+	2. improve accuracy/ productivity.	(24,26,27)
	Problem 2: follow the work rules	
+	1. improve attendance/promptness.	(13,14,16,26)
	Problem 3: improve job satisfaction	
0	1. work on resolving promotion problems	(27)
	2. review career with County	()
	General supervisory actions	
0	1. meet again to work on problems and goals	(27)
+	2. use of the discipline process	(14)
	3. special review processes	()

D. Action

1.Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	5	takes specific actions to resolve the crisis issues; discusses and takes some actions for the other issues discussed, but does not necessarily take specific actions for all issues discussed.

Quality	Implementing (14,17,24,26,27)	Paragraphs
	Problem 1: improve productivity	
+	1. accuracy and productivity	(24,26,27)
	2. workload distribution	()
	Problem 2: follow the work rules	
+	1. attendance and promptness	(14,27)
	Problem 3: improve job satisfaction	
	1. promotion preparedness	()
	General supervisory actions	
0	1. meet again	(27)
+	2. disciplinary process	(17)
0	3. assist in locating and resolving problems	(27)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 5
PS=5, IR=5

E. Supervision

1.Supervision: planning, organizing and motivating implementation

Actual	Target	Description
____	5	states the plan as a goal, may do some directing, without organizing or motivating.

Quality	Ensuring (12,13,24,26,27,28)	Paragraphs
0	1. states plan	(27)
+	2. directs	(12,24,26,27,28)
0	3. motivates	(13)

F. Problem Identification(second time)

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	5	mentions a monitoring plan but does not necessarily initiate it.

Quality	Monitoring (27)	Paragraphs
0	1. initiates plan	(27)

Quality Rating Scale:
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) topic not covered

Script 6
PS=5, IR=3

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	5	explains the normal, planned and expected actions and mentions the major deviation and some other secondary ones.

Quality	Deviations Mentioned (2,4,5,9,10,11,13,24)	Paragraphs
+	1. problem client	(5)
+	2. fails to call when sick	(5,24)
+	3. sick too often/absenteeism	(5,11,13)
	4. socialize too much	()
+	5. performance deviation from old appraisal	(2,4,5,9,10,11,13)
	6. requested transfer	()
	7. alcohol at lunch	()
+	8. general increase in problems	(5)
+	9. tardiness/promptness	(11,13)

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	5	asks questions but does not always persist in seeking the essential details relating to deviation participants and the impact; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Details mentioned (2,3,4,5,10,11,14,24)	Paragraphs
+	1. who involved: Chester, secretary, director, client	(5,10,11)
+	2. who affected: client, director, Mary, new supervisor, unit staff	(2,3,5,10,11,14,24)
0	3. impact on future relations: Chester, director and client upset	(4,5,6)
+	4. impact on actions: client not served, went to director's house; relates to other problems	(4,5,6)
+	5. where: office/ Chester at home	(5)
+	6. when: last week Tuesday	(5)
	7. cost: dollar cost not specified	()
+	8. importance: high, client not served	(2,4,5)
+	9. what happened: description of the client/ service problem	(5,6)

Quality Rating Scale
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 6
PS=5, IR=3

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	5	asks questions but does not always persist in seeking the essential details relating to deviation participants and the impact; analyzes the obtained information in a logical manner and explains the obtained facts of the problem.

Quality	Expectations (8,9,11,13,16,17,18,19,21,22,24)	Paragraphs
	Problem 1: improve productivity	
+	1. improve productivity	(13,18,22,24)
	2. help the client/avoid problems in future	()
+	3. Do the job expected of a PFW	(9,13,18,19)
	Problem 2: follow the work rules	
	1. following the rules	()
+	2. call in when sick	(9,24)
+	3. improve attendance	(9,22)
	4. no alcohol on work site	()
	5. decrease socializing	()
+	6. improve promptness	(9,22,24)
	Problem 3.: improve job satisfaction	
+	1. improve satisfaction/ work relationships	(16,22,24)
+	2. become promotable	(8,11,13,16,17,21,22)

Quality	People to be involved (13,18,22)	Paragraphs
+	1. Chester retain control of PS	(13,22)
0	2. tell Sup. if overworked	(18)

Quality	risks (13,24)	Paragraphs
	1. oral reprimand	()
	2. written reprimand	()
	3. one day suspension	()
	4. five day suspension	()
	5. fired	()
+	6. general consequences	(13,24)

Quality Rating Scale
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 6
PS=5, IR=3

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	5	asks questions seeking what caused the problem and why the problem occurred, plus some other concerns.

Quality	Diagnosing (5,6,11,12,13)	Paragraphs
+	1. Who caused the problem: Chester	(5)
+	2. what caused the problem: Promotion issue with Stan	(5,6,11,12,13)
	3. when was the problem caused: six weeks ago	()
+	4. where was the problem caused: on the unit, and in the selection processes	(11,12)
+	5. why was the problem caused: anger, frustration, burnout	(13)
0	6. how was the problem caused: work slow down	(6,11,12)
	7. seeking interrelationships among deviations and among causes	()

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	5	develops multiple possible solutions for at least one identified problem, with all other issues having one possible solution mentioned; no assistance is sought; returns to earlier problem solving steps if necessary.

Number	Alternatives (13,16,18,21,22,23,24)	Paragraphs
	Problem 1: improve productivity	
2 +	1. assess service/accuracy/productivity	(18,24)
2 +	2. review workload	(18)
	Problem 2: follow the work rules	
2 +	1. come to work on time/call in when sick	(13,24)
	2. cut down on socializing	()
	3. follow the rules	()
	Problem 3: improve job satisfaction	
4 +	1. develop plan for promotion	(13,16,21,22,23)
	2. reassess career with County	()
	3. improve job satisfaction	()
	General Supervisory Actions	
1 0	1. seeks assistance from Chester	(21)
2 +	2. ask for help from supervisor	(18)
	3. meet again	()
	4. special review process	()

Quality Rating Scale
"+" Topic directly and completely covered
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorrectly
" " (blank) Topic not covered

Script 8
PS=3, IR=5

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	3	fails to clearly explain one of the important normal, planned or expected actions or fails to mention one of the major deviations.

Quality	Deviations Mentioned (1,2,7,9,10,12,15,19)	Paragraphs
+	1. problem client	(1,2,7,9,10)
	2. fails to call when sick	()
	3. sick too often/absenteeism	()
+	4. socialize too much	(10)
+	5. performance deviation from old appraisal	(10,12,15,19)
+	6. requested transfer	(10)
+	7. alcohol at lunch	(10)
+	8. general increase in problems	(10)
	9. tardiness/promptness	()

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	3	explains and seeks to verify the details of the pre-work problem definition information without asking questions to obtain the essential details of the issues that may arise during discussion; may jump to conclusions or illogically analyze the information; may not explain the facts of the problem.

Quality	Details mentioned (1,2,3,4,5,6,7,8,9,10,12,15,19)	Paragraphs
-	1. who involved: Chester, secretary, director, client	(1,2,5,6,7,8,9,12,19)
+	2. who affected: client, director, Mary, new supervisor, unit staff	(1,5,6,7,8,9,19)
+	3. impact on future relations: Chester, director and client upset	(19)
+	4. impact on actions: client not served, went to director's house; relates to other problems	(7,8,9)
+	5. where: office/ Chester at home	(1)
+	6. when: last week Tuesday	(1,2,3)
	7. cost: dollar cost not specified	()
+	8. importance: high, client not served	(1,3,8,9,10,15)
+	9. what happened: description of the client/ service problem	(2,4,5,6,7,8,9,10)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	3	states expected outcomes that may be inappropriate, unclear or irrelevant; may specify inappropriate, unclear or irrelevant risks and constraints.

Quality	Expectations (9,10,15,17,19,22,24)	Paragraphs
	Problem 1: improve productivity	
0	1. improve productivity	(10,15)
-	2. help the client/avoid problems in future	(9,10,17)
	3. Do the job expected of a PFW	()
	Problem 2: follow the work rules	
0	1. following the rules	(10)
-	2. call in when sick	(9)
	3. improve attendance	()
0	4. no alcohol on work site	(10)
0	5. decrease socializing	(10)
	6. improve promptness	()
	Problem 3.: improve job satisfaction	
0	1. improve satisfaction/ work relationships	(15,19,24)
+	2. become promotable	(17,22)

Quality	People to be involved (22,24)	Paragraphs
	1. Chester retain control of PS	()
+	2. tell Sup. if overworked	(22,24)

Quality	risks (3,10)	Paragraphs
	1. oral reprimand	()
	2. written reprimand	()
	3. one day suspension	()
	4. five day suspension	()
	5. fired	()
+	6. general consequences	(3,10)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	3	does not ask questions that are intended to identify the cause of the problem; seeks only symptoms (i.e., who was involved and affected, what was the impact on relations and actions, where and when did the problem occur and what were the costs).

Quality	Diagnosing (5,6,7,9,10,12,15,17,18,19,20)	Paragraphs
-	1. Who caused the problem: Chester	(5,6,7,10)
+	2. what caused the problem: Promotion issue with Stan	(9,15,17,18,19,20)
	3. when was the problem caused: six weeks ago	()
0	4. where was the problem caused: on the unit, and in the selection processes	(12)
+	5. why was the problem caused: anger, frustration, burnout	(7,15,17,18,19,20)
0	6. how was the problem caused: work slow down	(15,19)
	7. seeking interrelationships among deviations and among causes	()

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	3	possible solutions are not consistently developed for all identified problems; may not seek help when developing possible solutions; may not return to previous problem solving steps if a new problem arises.

Number	Alternatives (15,22,23,24)	Paragraphs
	Problem 1: improve productivity	
1 -	1. assess service/accuracy/productivity	(15)
	2. review workload	()
	Problem 2: follow the work rules	
	1. come to work on time/call in when sick	()
	2. cut down on socializing	()
1 0	3. follow the rules	(23)
	Problem 3: improve job satisfaction	
1 +	1. develop plan for promotion	(22)
	2. reassess career with County	()
	3. improve job satisfaction	()
	General Supervisory Actions	
3 0	1. seeks assistance from Chester	(22,23,24)
2 +	2. ask for help from supervisor	(22,24)
2 +	3. meet again	(22,23)
	4. special review process	()

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

C. **Decision Making**(continued)

2. Establishing criteria: stating a means for appraisal of possible solutions

Actual	Target	Description
____	3	criteria not discussed or evaluations are made without objective standards being stated or criteria are not developed.

Quality	Criteria (15,17,22,23,24)	Paragraphs
	Problem 1: improve productivity	
-	1. provide proper service to clients	(15,17)
	Problem 2: follow the work rules	
	1. come to work on time	()
	2. call in when sick	()
	3. follow rules on socializing	()
	4. follow rules on drinking	()
	5. face discipline or resolve problems	()
0	6. general following of rules	(23,24)
	Problem 3: improve job satisfaction	
+	1. become promotable	(22,23,24)
	2. improve satisfaction, without transfer	()
+	3. take the initiative to improve	(15,22,23,24)

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	3	appraises possible solutions incorrectly or inconsistently.

Quality	Appraisal (22,23)	Paragraphs
	Problem 1: improve productivity	
0	1. Even though this problem client issue cannot be blamed on you, we've had a chance to acknowledge that we'll start fresh with each other.	(23)
	Problem 2: follow the work rules	
0	1. Even though this problem client issue cannot be blamed on you, we've had a chance to acknowledge that we'll start fresh with each other.	(23)
	Problem 3: improve job satisfaction	
+	1. Even though this problem client issue cannot be blamed on you, I'd like you to help you in polishing your skills for promotion.	(22)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

C. Decision Making(concluded)

4.choosing:making a decision to implement the best solution

Actual	Target	Description
____	3	decides on and specifies a course of action, but a decision is not reached on at least one important or crisis issue.

Quality	Choosing (9,15,22,23,24)	Paragraphs
	Problem 1: improve productivity	
	1. review workload	()
0	2. improve accuracy/ productivity.	(15,23)
	Problem 2: follow the work rules	
-	1. improve attendance/promptness.	(9)
	Problem 3: improve job satisfaction	
+	1. work on resolving promotion problems	(22,23)
	2. review career with County	()
	General supervisory actions	
+	1. meet again to work on problems and goals	(9,22,24)
	2. use of the discipline process	()
	3. special review processes	()

D. Action

1.Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	3	does not take specific actions to resolve the crisis issues, but some actions are taken.

Quality	Implementing (22,23)	Paragraphs
	Problem 1: improve productivity	
	1. accuracy and productivity	()
	2. workload distribution	()
	Problem 2: follow the work rules	
0	1. attendance and promptness	(23)
	Problem 3: improve job satisfaction	
+	1. promotion preparedness	(22,23)
	General supervisory actions	
+	1. meet again	(22,23)
	2. disciplinary process	()
+	3. assist in locating and resolving problems	(22,23)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 8
PS=3, IR=5

E. Supervision

1.**Supervision: planning, organizing and motivating implementation**

Actual	Target	Description
____	3	directing statements and statements of hope are expressed, some weak motivating or negative incentives with few supervisory actions mentioned

Quality	Ensuring (21,22,23,24)	Paragraphs
0	1. states plan	(22)
+	2. directs	(22,23,24)
0	3. motivates	(21,22,23,24)

F. Problem Identification(second time)

1.**Monitoring: understanding expected actions and recognizing deviations**

Actual	Target	Description
____	3	monitoring is not mentioned or may be done in a way that detracts from the solution.

Quality	Monitoring ()	Paragraphs
	1. initiates plan	()

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic not covered
" " (blank) Topic not covered

Script 9
PS=3, IR=3

SCRIPT VERIFICATION WORK SHEET
PROBLEM SOLVING

Rater's Name:_____

A. Problem Identification

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	3	fails to clearly explain one of the important normal, planned or expected actions or fails to mention one of the major deviations.

Quality	Deviations Mentioned (2,3,5,12,14,15,25,28,31)	Paragraphs
-	1. problem client	(14,28)
+	2. fails to call when sick	(2)
+	3. sick too often/absenteeism	(2)
+	4. socialize too much	(2)
+	5. performance deviation from old appraisal	(2,3,5,12,15,31)
	6. requested transfer	()
+	7. alcohol at lunch	(25)
	8. general increase in problems	()
	9. tardiness/promptness	()

2.Defining the problem: seeking and analyzing problem related facts

Actual	Target	Description
____	3	explains and seeks to verify the details of the pre-work problem definition information without asking questions to obtain the essential details of the issues that may arise during discussion; may jump to conclusions or illogically analyze the information; may not explain the facts of the problem.

Quality	Details mentioned (14,18,19,28,31)	Paragraphs
-	1. who involved: Chester, secretary, director, client	(14,19,28)
-	2. who affected: client, director, Mary, new supervisor, unit staff	(19,28)
	3. impact on future relations: Chester, director and client upset	()
-	4. impact on actions: client not served, went to director's house; relates to other problems	(14)
	5. where: office/ Chester at home	()
	6. when: last week Tuesday	()
	7. cost: dollar cost not specified	()
+	8. importance: high, client not served	(31)
-	9. what happened: description of the client/ service problem	(18,19)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

B. Problem Analysis

1.Specifying decision objectives: listing risks, and expectations

Actual	Target	Description
____	3	states expected outcomes that may be inappropriate, unclear or irrelevant; may specify inappropriate, unclear or irrelevant risks and constraints.

Quality	Expectations (10,12,13,16,17,18,21,32)	Paragraphs
	Problem 1: improve productivity	
+	1. improve productivity	(10,12,13,16,17,18,32)
	2. help the client/avoid problems in future	()
-	3. Do the job expected of a PFW	(10,21)
	Problem 2: follow the work rules	
+	1. following the rules	(18)
	2. call in when sick	()
	3. improve attendance	()
	4. no alcohol on work site	()
	5. decrease socializing	()
	6. improve promptness	()
	Problem 3.: improve job satisfaction	
	1. improve satisfaction/ work relationships	()
+	2. become promotable	(17)

Quality	People to be involved (10,17,18,19,21,31,32)	Paragraphs
+	1. Chester retain control of PS	(19,31,32)
+	2. tell Sup. if overworked	(10,17,18,21)

Quality	risks (12,21,22,23,24,25,26,27,29,30,31)	Paragraphs
	1. oral reprimand	()
+	2. written reprimand	(12,21,22)
+	3. one day suspension	(23,24,25)
+	4. five day suspension	(29)
+	5. fired	(29,30)
+	6. general consequences	(12,23,26,27,30,31)

7

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

B. Problem Analysis(continued)

2.Diagnosing: digging beneath symptoms to identify problem causes

Actual	Target	Description
____	3	does not ask questions that are intended to identify the cause of the problem; seeks only symptoms (i.e., who was involved and affected, what was the impact on relations and actions, where and when did the problem occur and what were the costs).

Quality	Diagnosing (7,8,9,10,11,18,21)	Paragraphs
+	1. Who caused the problem: Chester	(8,9,10)
-	2. what caused the problem: Promotion issue with Stan	(7,8,10,11,18,21)
	3. when was the problem caused: six weeks ago	()
0	4. where was the problem caused: on the unit, and in the selection processes	(7)
	5. why was the problem caused: anger, frustration, burnout	()
+	6. how was the problem caused: work slow down	(7,8,11)
	7. seeking interrelationships among deviations and among causes	()

C. Decision Making

1.Developing possible solutions

Actual	Target	Description
____	3	possible solutions are not consistently developed for all identified problems; may not seek help when developing possible solutions; may not return to previous problem solving steps if a new problem arises.

Number	Alternatives (10,16,17,18,21,31,32)	Paragraphs
	Problem 1: improve productivity	
4 +	1. assess service/accuracy/productivity	(10,16,17,21)
2 -	2. review workload	(10,17)
	Problem 2: follow the work rules	
	1. come to work on time/call in when sick	()
	2. cut down on socializing	()
	3. follow the rules	()
	Problem 3: improve job satisfaction	
1 +	1. develop plan for promotion	(17)
	2. reassess career with County	()
	3. improve job satisfaction	()
	General Supervisory Actions	
3 +	1. seeks assistance from Chester	(18,31,32)
2 0	2. ask for help from supervisor	(17,18)
1 +	3. meet again	(32)
1 +	4. special review process	(32)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

C. **Decision Making**(continued)

2. Establishing criteria: stating a means for appraisal of possible solutions

Actual	Target	Description
____	3	criteria not discussed or evaluations are made without objective standards being stated or criteria are not developed.

Quality	Criteria (10,17,21,23)	Paragraphs
	Problem 1: improve productivity	
+	1. provide proper service to clients	(10,17)
	Problem 2: follow the work rules	
	1. come to work on time	()
	2. call in when sick	()
	3. follow rules on socializing	()
	4. follow rules on drinking	()
+	5. face discipline or resolve problems	(21)
+	6. general following of rules	(23)
	Problem 3: improve job satisfaction	
+	1. become promotable	(17)
	2. improve satisfaction, without transfer	()
	3. take the initiative to improve	()

C. **Decision Making**(continued)

3. Appraising Possible Solutions: using the criteria to assess possible solutions

Actual	Target	Description
____	3	appraises possible solutions incorrectly or inconsistently.

Quality	Appraisal (10,17,32)	Paragraphs
	Problem 1: improve productivity	
-	1. Don't do other people's jobs, just do your case load in a timely fashion, then you can look at getting promoted.	(10,17)
	Problem 2: follow the work rules	
0	1. Make a copy of this memo, go over it and strategize how to keep working and look at your work load	(32)
	Problem 3: improve job satisfaction	
0	1. Don't do other people's jobs, just do your case load in a timely fashion, then you can look at getting promoted.	(10,17)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

C. Decision Making(concluded)

4.choosing:making a decision to implement the best solution

Actual	Target	Description
____	3	decides on and specifies a course of action, but a decision is not reached on at least one important or crisis issue.

Quality	Choosing (10,12,16,17,18,21,32)	Paragraphs
	Problem 1: improve productivity	
	1. review workload	()
+	2. improve accuracy/ productivity.	(10,16,17)
	Problem 2: follow the work rules	
	1. improve attendance/promptness.	()
	Problem 3: improve job satisfaction	
+	1. work on resolving promotion problems	(17)
	2. review career with County	()
	General supervisory actions	
+	1. meet again to work on problems and goals	(18,32)
+	2. use of the discipline process	(12,21)
+	3. special review processes	(18)

D. Action

1.Implementing: acting on the chosen solution for problem solution

Actual	Target	Description
____	3	does not take specific actions to resolve the crisis issues, but some actions are taken.

Quality	Implementing (17,18,21,30,31,32)	Paragraphs
	Problem 1: improve productivity	
-	1. accuracy and productivity	(18,21)
+	2. workload distribution	(17,32)
	Problem 2: follow the work rules	
	1. attendance and promptness	()
	Problem 3: improve job satisfaction	
0	1. promotion preparedness	(17)
	General supervisory actions	
0	1. meet again	(18,32)
+	2. disciplinary process	(30,31,32)
+	3. assist in locating and resolving problems	(18)

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

Script 9
PS=3, IR=3

E. Supervision

1.Supervision: planning, organizing and motivating implementation

Actual	Target	Description
____	3	directing statements and statements of hope are expressed, some weak motivating or negative incentives with few supervisory actions mentioned

Quality	Ensuring (17,18,31,32)	Paragraphs
+	1. states plan	(18,32)
+	2. directs	(17,18,32)
-	3. motivates	(17,31)

F. Problem Identification(second time)

1.Monitoring: understanding expected actions and recognizing deviations

Actual	Target	Description
____	3	monitoring is not mentioned or may be done in a way that detracts from the solution.

Quality	Monitoring ()	Paragraphs
	1. initiates plan	()

Quality Rating Scale
"+" Topic covered directly and completely
"0" Topic covered, but weakly or indirectly
"-" Topic covered incorectly
" " (blank) Topic not covered

APPENDIX G

FILE:PS6IR6.PH Target Script 1:
Problem Solving = 6, Interpersonal Relations = 6

1 S: I'm glad that we could meet today Chester. I'm your new Supervisor. It's nice to meet you.

C: Hi.

2 S: As your new supervisor I've been given some information about a situation that occurred involving a client and the department head . . .

C: (MUMBLED TO HIMSELF) Ughhh.

3 S: . . . and ___ a series of other points, which may or may not be relevant to the matter. . . Umm, what I'd like to do is briefly go through these issues and get your opinions on them and see what we need to do about it, if anything. Apparently you were out sick on Tuesday, last week. One of your clients had an appointment with you. When he came in, no one in the office knew that you were sick. So the client went untreated that day and must have been angry enough to wind up on our director's door at ten o'clock at night.

C: Yea, I heard about that. Umm, I don't think it was my fault, I . . . I was sick that day.

4 S: Uh huh, please go on.

C: Umm, and if . . . the problem is the secretary. The secretary leaves messages on peoples desks, and ahh . . . you know the messages pile up and no one sees if they're there. If there was some central location to put the messages down people could have seen that my messages weren't being answered and followed up on them. Um, then they would have known the guy had come in. I didn't remem . . . It must have been a ringer, you know, somebody from Channel 5 or something; I don't really think it was my fault . . . it was just one of those things.

5 S: So, the problem is that a client went to the reception desk, asked for you, and since the receptionist couldn't get ahold of you personally, she put the message on your desk.

C: That's the problem.

6 S: That leaves me puzzled, Chester. You were sick and it's the established pattern that notes are left on people's desks. I'm unclear about why you feel this issue may not be important.

C: Well look . . . more than once I've seen . . . even Stan, the guy who got my promotion . . . he, he ah, gets messages and stuff left on his desk and I go over, take his messages and

cover his work assignments when he's out screwing around . . . you know I handle his problems and stuff and . . . the favor hasn't been returned to me, you know.

7 S: Ah, you're frustrated because you've handled messages for fellow workers in the past and you had expected them to return the favor for you this time. And that didn't happen. It sounds like your fellow workers were unaware that they needed to help you.

C: Well, it's part of the . . . if I didn't do that, we'd have all this crap piled up all the time. I covered Stan's butt so much that he got the promotion that I deserved.

8 S: Well . . . You're a bit put out after you lost a promotion because you've covered other peoples' work for them. That's a serious issue. Let's talk about that in a minute, but first I'd like to go back to the specific issue. Did you call in and arrange for somebody to cover your client load or inform your clients that you would be unavailable that day?

C: I was sick that day. I'm just doing what everybody does . . .

9 S: Chester, the rules require you, when you are sick, to call in and inform your supervisor. You didn't call in nor arrange for someone to cover your load. That's why this is important now.

C: I assumed that they'd realize I wasn't gonna be there. More than once I've covered for people who ended up being sick that day. And no one noticed. Nothing ever seems to be done about it until I personally cover the phone messages or meet a client out in the lobby . . . who's been wandering around for 45 minutes getting angry.

10 S: Chester, you've been a benefit to the unit for a long time, apparently without your fellow workers being fully aware of it. That really points out the issue here. It almost sounds like you knew that you are the only person who will help a fellow worker by handling their messages when they are not in the office. Even so, you didn't let your supervisor know that you would be out so your own work load could be handled. How can we handle these situations in the future to avoid these types of problems?

C: I do it for everybody else. Why don't they do it for me? It always been my policy.

11 S: Yes, you've really been doing some good work as a lead worker. You also have the responsibility of coming in or calling in, one of the two, to assure that your workload is met. In this case you failed to do that. Chester, can you give me any suggestions for how we can avoid this issue of

your not telling us when you are ill and, as a result, your work load doesn't get covered?

C: (Exhales loudly) I didn't think it was any big deal, it was just this ringer that shows up and goes marching over to the Director's doorstep at ten o'clock at night.

12 S: I have . . . I have no data to say that this is a ringer. The only thing we have is a client who's not been served, not in our established format. You didn't take any action to get someone to relieve you.

C: I'm doing whatever everybody else is doing. People are always blowing off around here.

13 S: Well, if people in my unit are doing it in that way, then they will get this same conversation. I want to . . . to establish one thing very clearly. This is what I will expect in the future: if you are sick or cannot make work or will be late for work that you will call me and inform me, or at least leave a message, if I am not here, with my secretary. If you fail to do that in the future it will be a cause for reprimand. The more frequent and serious the infractions, the more stringent the discipline. This conversation is just my way of letting you know how serious this is.

C: Oh man, what . . . one problem and now I'm getting stuck for it.

14 S: Yes.

C: Oh boy, how many other times has this happened and I . . . and I covered for people. . . Even Stan who gets my promotion. I've helped him hundreds of times while he's out politicking to get a job that I should have gotten. Ah gee, now I get stuck for it, huh?

15 S: Yes. It's not always pleasant to be held responsible for your actions. You are still held responsible.

C: Well, that's the last time I'm gonna ever help somebody It just hasn't paid off.

16 S: Chester, I disagree. You're too good of a worker. You could be supervisor material. It sounds like we've had something happen in the past few weeks that shook your confidence in your promotability and caused you to slack off of a work pace that . . . that has lead to very good performance ratings in the past. There's a number of problems issues that seem to relate back to this confidence problem: you've been absent or tardy a bit too often, you've been spending a little too much time socializing rather than working while at the office, so your work has been late and there has even been a report of alcohol use during the work day. These have all been noted

Script 1
PS=6, IR=6

by Mary. Your high quality of performance evidently hasn't paid off as quickly as you had anticipated, so you've slowed down. But good performance does not go unrecognized. I will notice, my boss will notice.

C: Yea? Well, it didn't happen this time. I think I got screwed.

17 S: Chester we need to talk about these other issues that have become apparent in the past six weeks. These issues go beyond not phoning in sick one day. You and I won't have sufficient time today to give these topics the full attention that they deserve, but they concern potentially very serious matters leading to low productivity. These are important issues that seem to relate to the issue we have been discussing. Describe for me, if you would, what's been causing these problems as of late?

C: I don't know if you can give me my promotion, but it sure would be nice. . . I've been here for nine years and I've been really doing a good job . . . Stan gets a promotion, all of a sudden there's a brand new person in his job, and I'm supposed to train that person. We have a brand new supervisor, Mary takes off. And, you know, I don't know you very well, but, you know, there's a lot of stuff that people come to me for um, you know, covering their little problems and stuff, and ah, I end up doing all of that and I'm not getting rewarded for it, so screw it, I'm not going to do it. Why should I? I'll just do what everybody else is doing, you know, how many people in the office, when the tracks were open the first year, blow off on Friday afternoons and head down to Canterbury Downs or, how many people in the office take off for a nice long lunch, I've never done that stuff, and I'm tired of it, you know, holding everybody up and not getting any rewards for it.

19 S: It sound like you've been pretty busy lately Chester. You feel overworked, like you were doing the job of a new supervisor and not getting noticed for it. What . . . I'm unclear about how this relates to the situation of the last six weeks, where you're out sick and not calling. Has something happened? It sounds like your work problems might be a reaction to Stan's getting the job you wanted.

C: Well . . .(PAUSE)

20 S: When did Stan get the job? Six weeks ago? Is that when you got passed over?

C: Yea, something like that.

21 S: As a result you feel discouraged and burned out over not being promoted, yet you also see your workload is going up. So your work output has diminished. You seem most upset at

losing this promotional opportunity. Even if Stan was out there doing some politicking, it may also be that he has a set of skills that were as good or better than yours in the eyes of the individual supervisor that selected him. It's going to take more than just coming to work on time or calling in sick to resolve this problem. These are a must, but you'll also need to become productive and promotable again. What I'd like to do is to help you to figure the skills you need to polish to be promoted, work up a plan for you to develop these skills, and the practice them. I think this will meet both our goals. What do you think?

C: So that's what I'm supposed to do is go out and try to snow the supervisors now.

22 S: No. Performance is the way to do it, not non-performance. Chester, I want to set up another meeting with you, let's say next week Wednesday at 10:00. In that meeting we'll talk about how you're doing on your work accuracy and sick leave usage and if you are sick, on calling in. But more importantly, before that meeting I'd like you to spend a little time and develop a short list of what you think an ideal supervisor's skills would be. I'm quite serious. Just write a short list of the skills you feel a supervisor should have. If you want to become promotable you are going to have to have a solid plan for getting there. The first step will be to get a clear picture of the goal. Our next meeting will be an effective way of turning this situation into something positive; into a starting point for your advancement. O.K.?

C: Can you promise me my promotion?

23 S: Chester, I can't promise that. I can't promise you that you will get a promotion in the immediate future. But if we work together . . . if you work diligently on your goal, then this will have an effect. Can I get your agreement on resolving these problems and working toward the goals?

C: Yea, I guess.

24 S: Good. You don't sound too enthused about this, but it's a start. Right? Now, since I've gotten your agreement to call me whenever you are going to be sick and to work to improve your accuracy and promptness on work assignments, then in this case I choose to do nothing in terms of documenting this. For right now since I'm a new supervisor, we'll have this be our first talk. If you need help, if you think you are doing an unfair share of the work, come and talk to me. If the negative behavior we've talked about persists, if you do not personally take the initiative to improve, then we'll have another conversation and, unfortunately, it would result in the first step in the discipline process, a written reprimand. Do we have an agreement? Can we overcome these problems?

Script 1
PS=6, IR=6

C: Well, o.k.

S: Ok, then let's start afresh and I'll see you next week with your thought on supervision.

C: All right.

FILE:PS6IR5 Script 2
Problem Solving = 6 Interpersonal Relations = 5

1 D: Well Chester, Ah an issue that concerns me has come up and we need to talk about it. As you're aware I'm new to the unit and, ah, I think that what we'd like to do here is, ah, . . . is, ah get some idea of your view and your perception of a concern. The concerns are two. Essentially one that you've been . . . you're sick and missing a lot of time in the last few weeks and there seems to be a pattern of problems over the last six weeks. And most recently last Tuesday. This resulted in a client not being served in a timely fashion . . . There may even be other clients not being served.

C: (Breathes heavily)

2 D: This client, in turn, camped out on the Department Director's doorstep and I heard a lot about that, that problem. You weren't here that day, and when you're not here that does create that kind of a problem.

C: I don't think that was my fault at all.

3 D: Tell me about that.

C: Well, it's the secretary. She leaves phone messages pile up on your desk. She took a message from the guy and left it on my desk. I saw it when I got in the next day. I mean, you know, a lot of people are out of the office a lot.

4 D: Umm hmm.

C: But, all she does is she takes the message and lays them down on your desk and they just sit there. Umm, that's not my fault, I mean, you know people are sick a lot. I've been covering for a lot of people on their phone messages.

5 D: How can it happen that a client like this doesn't get any attention? The secretary leaves messages for people who will be out of the office for the day. Is that the only reason?

C: Will yeah, If . . . if someone would have seen the message on the desk they could have helped him. I mean, my god, I didn't know the guy was a ringer.

6 D: Well Chester, since I'm new to the unit, I'd like to find out whether Mary had . . . had any discussions with you about your sick leave usage and your work output?

C: No, Mary was most interested in just getting out of the unit. I've always been a good employee here. Mary knew that.

7 D: Your performance appraisals show that . . . that you've been a good employee. Ah, the concern is, concerns are about some

Script 2
PS=6, IR=5

problems we seem to be having in the last six weeks or so. Your actions are having an impact on more than just yourself and one client. You've been talking too much, sometimes drinking over the lunch hour and in general not being very productive. You've asked for a transfer. You have not been meeting expectations. Are you sure that this is the first you're hearing about this? Mary never mentioned this?

C: I don't think that I'm the one having the problem. I think it's the unit that's having the problem. You know, here I was up for promotion and all of a sudden, you know, that Stan he gets promoted over me. And Stan couldn't do his job worth anything. Well, he spent the last six, eight, ten weeks politicking for the promotion. He gets the promotion that I deserve, and then I get dumped with his work. And, you know, you're a brand new supervisor, and ah, and a lot of the stuff, just the organizing and stuff that's going on, I'm ending up doing that myself. And I think, that I'm tired of it, nobody else works; so why should I?

8 D: Tell me about the promotion, all these details that have happened in the past few weeks.

C: Well . . . I've been working here for nine years. I know this job better than anybody working on the floor. And, here we go through this whole promotion process, it takes forever, and I pass the test and get on the list and I don't get the promotion. I mean it's clear that they didn't want me to and they picked Stan, a guy who doesn't even know his job. I've been helping him out for all this time.

9 D: Well, let's get back to the issue, the thing of concern now. I understand your feelings and concerns about the promotional issue. It would be nice to see you successful and have an opportunity to be promoted. I think ah, the . . . the . . .

C: Is it too late for me to get the promotion then?

10 D: It's too late to get that promotion. But there are possibilities in the future. What about the, ah, next one?

C: Well that might not be for five years.

11 D: Well that's, ah, that can't be controlled. Tell me . . . it would be helpful to hear about how we can solve the problem, what you think you're willing to do to reduce that sick leave problem. It will be important to have you here more. Your expertise has been valued. You have a chance now with a new supervisor to start fresh and to work to once again have an effective working relationship with me during the time you are in the unit. Ah, I think you have a chance you may not have had in the past. What can I do to help you?

Script 2
PS=6, IR=5

C: Well you can get somebody to handle all of Stan's work that knows what's going on. I'm still doing all his work and my work. And ah, I figured, well, I'm gonna do what, oh, the unit's doing. I mean, you know, I'm not getting any benefit from bustin my butt over here. I'm getting real tired of it. I've been here for nine years doing this. And after nine years of doing all this work and really working with these people I don't get any recognition for it. I mean, why should I continue?

13 D: So in the past month or so you've gotten tired and frustrated and as a result, slowed down on completing your job. You've been baffled because you don't get the recognition you feel you deserve?

C: I didn't get my promotion and my workload doubled. Gimme a break.

14 D: Why. . . okay, let's talk about that. You lost a chance at a promotion and this has caused you to start having problems on the job. Why don't you think you got the promotion or recognition? What's your view?

C: I think its because I was doing Stan's work while he was out politicking with the supervisors who were doing the hiring. They picked who they wanted to pick.

16 D: Well, let me ask you this. What . . . what is it that Stan is more effective at or you think is more effective at?

C: Buttering up the people who are hiring.

17 D: So you're talking about his interpersonal skill . . . His ability to work with management.

C: I don't know if I'd call them skills, but he's certainly sleazy enough to pull that stuff off.

18 D: Now you sound pretty mad about this, but "sleazy" is a value judgment. What you're saying is that there is some difference in his ability to work with and influence people.

C: I was doing his job for him. I mean I was answering a lot of his phone calls, a lot of his cases while he's out screwing around in the afternoon, pretending to see the clients.

19 D: Did you talk to anyone about that? I need to get all the facts on why this problem is occurring. Did you bring it to anyone's attention, to Mary?

C: When most of his work comes out with my handwriting on it? I suppose they could figure that out.

23 D: Chester, please tell me. Did you tell Mary?

Script 2
PS=6, IR=5

C: I don't think that it was my place to . . .

24 D: She needs to know about it. The concern is it would be good if you run into anything like that to be direct and let me hear about it first. Come in and tell me when you're concerned about something that's occurring on the job. Mary didn't get told so, what can we . . . what would we have to do to solve the problem of your being overworked.

C: Well, find somebody to handle the rest of that job. Find somebody that knows what's going on enough so that they're not going to screw it up.

25 D: OK, OK. Let's not get any more angry. I'd like to see us work well together. I'd like to see you take some responsibility. I want to know, specifically, what you're going to do to deal with some . . . some of these problems, and what kind of help you want specifically from me. It, you know, takes time to staff some of these positions. We do have to depend on experienced people here. We need you here, Chester.

C: Does that mean I'm going to end up hold down every . . . three people's jobs?

26 D: It probably means . . . that ah, what we can do there is . . . is look at those additional tasks and distribute those among some of your fellow workers here. We'll take a look at the workload situation, maybe everyone has to help out a little bit more. We can solve that. The long term issue is the concern here. The concern is about you. Umm, it would be nice to ah, to see that we use your expertise usefully here. That would be good for both of us. The concern remains about a pattern that seems to be developing . . . um, you seem to be linking it into the promotional issue. That should be checked on, get some reasons, we'll talk about it when we get some further info. You're still going to have to be doing a bit more. But I want you to tell someone when you feel you are getting dumped on. The workload and the attendance patterns will be monitored to try to make sure that it stays as even as possible. Everyone should pull their own weight.

C: I'd settle for that.

27 D: That . . . That's agreed on. Umm, you should . . . you should probably think about your career here. Think about ah, whether you want to continue working here. Ah, it would be nice to get together, maybe in a couple of days ah, we'll set up something that is convenient for you, and we should . . . we should set up some expectations . . . some performance criteria to meet. We should have a contract with you, a formal understanding that if the good actions occur, then the problems will be solved and you can focus on getting

promoted, but if the expectations are not met, the consequences will occur. We should put something in writing ah, and we should deal with the issue of sick leave usage. You should know that if you do what is expected, what we put in writing, then you will be successful. The consequences that can happen should also be understood by you, if we have people missing . . . missing work. The steps in the disciplinary process indicate that the next step for an infraction is a written reprimand. You only get a few more serious warnings. After that, then you could even get fired. Now ah, why don't you come back with a plan, or at least some thoughts ah, on what you're willing to do. If you want us . . . if you want to continue to work here ah, it's gotta . . . it's gotta be a two-way street. We have to expect something from each of us. It would be nice if you would identify what you'd like to see from me, so that we have an effective working relationship. So we'd like you to identify what you're willing to do to make that effective.

C: Holy smokes. I miss a ah, a day and I already have to set up expectations for if I want to keep working here?

28 D: Chester, please realize how serious this is. The pattern that's started is something that ah, maybe you don't feel is serious, but we do. We're very concerned here about serving the client needs. We're sensitive to . . . to what employees need to do to accomplish that. We don't want you to forget why we are here. You're an experienced person, a valued employee ah, it sounds like you probably have some . . . you may have some strong feelings . . . some strong reasons for . . . for feeling discouraged right now. It is important for us to work through those, if we can. You should also know that your productivity, your performance accuracy is . . . is a serious enough issue, that whether your former supervisor said anything to you or not, . . . serious consequences will be put behind it, if it is necessary, if . . . if we don't, if you're not able to resolve this. We'll need to be watching closely for any recurrences. You're a good employee, . . . you're an excellent ah, teacher to the other people in the unit, your experience really makes a difference. We value that a lot. I'm gonna need you here, and I . . . yet I think that . . .

C: I'm hearing a lot of different messages then.

29 D: There are two messages. I think one is your attendance: coming to work on time, not being tardy and phoning to let us know if you are ill and will be out of the office. This is a serious problem. This is the immediate problem. There is a longer term issue here, and that's . . . I want . . . I'm concerned about . . . about your effectiveness, your career here in this organization. I want to make you successful. I want the unit to be successful in serving our clients. And . . . and I'm not sure whether you fit into our long term plan.

Script 2
PS=6, IR=5

You're going to have to prove that you do by eliminating these attendance problems and working to improve your performance. We'll work on the promotional issues, but only you can work on following the attendance policies. Let's eliminate these problems and avoid the consequences.

C: I will, but I don't wanna go back to holding up everybody else's work for them.

30 D: That is understood. What we'll do is we're going to look at that workload situation. We're going to check on the workloads of the other people, ah, one thing that can be said is that we intend to distribute the workload as evenly as possible, as fairly as possible. Everyone's going to have to do a little bit more for awhile. Until we're through this.

C: Okay.

31 D: Okay, so can we meet again on Monday at 8:00? We'll talk some more about your career desires and your thoughts about this conversation.

C: Yeah, okay.

33 D: Alright, good.

FILE: PS6IR3 Script 3
Problem Solving = 6 Interpersonal Relations = 3

1 S: OK Chester. Thank you for coming in today. As you know I'm Mary Johnson's replacement and I am your new supervisor. And this meeting was set up, I did not schedule it but since we have an opportunity to meet . . . we, aah, had some problems on your performance and your adherence to work rules. And I have some background from Mary's records, but I need to get all the facts and then decide what I am going to do to resolve the problems that have been presented. So, first I'd like to tell you what I can surmise are the problems from reviewing some material I have been given. First of all, you've always . . . in reading over the past performance appraisals they've always listed you as an acceptable employee to this division, your work has been timely, but I've been given information that over the last, oh, six to eight weeks that your performance quality has fallen off, not only in the case work you are doing in your interactions with clients, but also in your attendance and punctuality and your adherence to work rules related to consumption of alcohol during the work day. You're aware that your client, on the day that you were sick, that you had made some assurances to, that you would meet, ah, showed up on the department director's doorstep at ten o'clock on Tuesday night? Are you aware of that? Umm, what has happened over the last six weeks? I need an explanation from you as to why you think this has been occurring.

C: I don't really think that there is a problem. (Exhales loudly) I still don't consider that my fault. I mean, I was sick that day and it's not my fault that the secretary doesn't . . . you know . . . she just piles messages up on my desk without letting anybody know that they're just kinda piling up there. I mean someone else coulda handled that. I don't think that I'm the only one that works here. I don't think that it should have been a big deal. I've been doing good work for a long time and I still think that I'm doing good work. I think that I'm even doing better work now than I did in the past, I think the amount of work that is getting shoved down my throat is unbelievable and I think I'm doing pretty good at it.

2 S: Well I think . . .

C: I think that this guy was a ringer anyway.

3 S: Well, that's irrelevant. The issue is that you didn't let us know in time that ah, you were going to be out that day, so we couldn't, and didn't know that that client was going to be coming in and that he was a fairly demanding client and now you're in trouble for it. The fact that the secretary did what was expected of her is not related to the problem. Next time warn someone the day before or call the client or call

Script 3
PS=6, IR=3

the supervisor and say you are going to be late. You should have advised Mary that you had asked this client to come that day so that the paperwork could have been finished.

C: Well, no one else but me bothers to do that. I mean this happens to me all the time. I'm always running around to check other people's desks because they're sick on Monday or they're blowing out halfway through Friday. Umm, I just assumed someone would take care of it for me. I seem to be the only one working in the unit.

4 S: No, that's not true. A PFW is expected to cover for other people, but other people have also covered for you. We need to know that you have made some assurances to clients that you can't keep. And as you know, we're here to serve the client. The issue . . . a second issue, is your job performance, your work is not getting done on time. You hate your job; you have requested a transfer. We know how this incident occurred. You've slowed down and become less accurate in your work. But the question is why. Is there something about the work that's been bothering you? Do you feel like you haven't gotten proper training? Or you haven't been given the tools you need to do the job? You've got to tell me what you think the problem is? Why you've been skipping work? Why you've been drinking on the job?

C: I don't hate my job. I've been doing this job for nine years. But, my supervisor takes off on me, Stan takes off on me and gets my promotion. I'm doing three people's work back there and I'm real tired of it. And now I'm getting my throat sliced for being sick on one day. It's really maddening.

5 S: Ok, Mary didn't mention that you are being assigned more work than the other employees. And it might be hard on an employee when he missed a promotion or when his supervisor's been replaced, and the lead workers are expected to help in the transition. You have been here a long time. You should know a lot. However, the work has to get done on time, clients at least deserve your respect. This is the standard you must meet. People who don't do their work do not get promotions. I consider it a serious problem that your performance has slipped to this level.

C: (Quick loud exhalation) I don't think that my performance has slipped. I think I'm doing three people's jobs. Umm, I'm training someone who, ah, I have no business training, just because I know the work better than anybody else. I get real tired of it. They don't listen. They don't follow up right away. Umm, and then I've got all my own work to do. And when I'm out the secretary piles up messages on my desk and doesn't even bother to tell anybody that I'm not there that day. And then Stan gets the promotion that I deserve more than he did. I don't think that I should have to hold up the

Script 3
PS=6, IR=3

work for the unit any more, I mean I'm just not getting any payoff for doing good work. So, what's the point?

6 S: The point is you're work is poor, it's below standards. The problems of your not getting work done on time and your excessive absenteeism must be corrected or you will face the consequences.

C: Well, I think what the problem should have been is that I should have got the promotion and I didn't. And you know, you give it to Stan and he didn't deserve it. He just went around and politicked for the entire six weeks before the test and I covered his job. I think that's what the real problem is.

7 S: We're not going to discuss that today. Some other time, if you'd like to get together with me when I've had an opportunity to completely review your total personnel file and talk about your inability to compete successfully in promotional opportunities or that you don't like training, I would be available to do that. But, I don't have any knowledge of these situations and that's not the purpose we're here to resolve today. The two issues today are your absenteeism and poor performance. What I am concerned about is your coming to work on time and not being absent, your one-on-one relationship with the clients and the accuracy of the eligibility you are determining. These must be corrected.

C: Boy it doesn't look like I'm going to get any service out of this anyway. I mean is that all you're here to do is lynch me? Is that the problem?

8 S: No, because you're apparently a good employee. But the problems are serious enough that, if you don't correct them, you will receive serious consequences. So, let's review first of all, what the consequences are for improper usage of sick leave and failing to report when you are sick. Today this is a written warning, ah, well a warning for you anyway - ah, I'll just be putting them in writing so that you can consult those notes. However, if these rules are violated, the next step will be, you will get a written reprimand, and that will go in your personnel file. Should you make another violation, after the time of the written reprimand, you will have a one-day suspension. A third violation of these work rules would lead to a five-day suspension.

C: Ohhh, my God, you're going to hang me for not calling in sick.

9 S: No, but you should understand what the rules are. I don't think this will occur, but I want to make sure you understand what would occur to any employee, not just you, should a violation of rules occur. A five-day suspension for the third violation. The fourth violation of these rules . . .

Script 3
PS=6, IR=3

C: Oh, my God, you're out to hang me.

10 S: . . . would lead to a termination with the personnel board. Like I said, you've apparently got some good abilities, and I don't think that it will ever get to that point. But, it's important that you understand. You understand what I've told you about calling in when sick and, also, the consequences of not following the rules. Do you understand?

C: Yeah.

11 S: You'll do what I say? You won't be absent and will be coming to work on time?

C: I get caught and I get killed. Wow, thanks.

12 S: That's not my question. Will you be calling me to let me know that you will be out of the office or not?

C: I'm not sick all that often.

13 S: Will you call me when you are home sick? Will you or will you not?

C: It's no big deal, yeah. I don't think my performance is any different from anybody else.

14 S: It is different from other people. And other people are not really the issue. If other people are having performance problems, I'll need to talk to them also. It's your performance that we need to fix. You've got to either come to work on time or call me personally with a legitimate excuse when you are ill. Now, we don't have enough time to finish the task now, so I want to set up another meeting with you as soon as possible so I can develop a corrective action plan. What I intend to do is use a special review process. During the review period I'll be closely monitoring your performance. At the end of the review period I'll review your progress. I'm not going to take any specific disciplinary action at this time, because, as I said, we'll get over the problems. However, at the next meeting some objectives and time lines will be set. And at the special review in a couple of months if there is no improvement in your attendance and these other issues, that will result in discipline. So,can you meet on Monday afternoon, after lunch? By then you should be able to think seriously about the steps that you personally can take to get your poor performance up to the standards we have discussed today.

C: I kind a feel like I'm getting picked on here. I mean so many people taking off, the workload just kind of piles up. You know, those other folks, they don't know what's going on. I end up doing their work for them. And, um, now I've got all

Script 3
PS=6, IR=3

these threats at me. You know, do this by now, or else get your throat sliced.

15 S: This isn't really a threat. We know what the problems are. We'll find out when, how and why your problems typically occur and then corrective action plans will be built. If that involves how the workload is distributed, then we'll deal with that. And your grabbing control of and actively following the objectives will help you do your cases better. That should have always been a goal of yours. These are not threats, but I do want you to consider the seriousness of your performance and your violations. So count on seeing me on Monday, after lunch.

C: All right.

16 S: Thanks.

FILE:PS5IR6 Script 4
Problem solving = 5, Interpersonal Relations = 6

1 B: Chester, I'd like to start off with a brief explanation of the purpose of why we're getting together today. I have some information provided to me by Mary Johnson relative to your work performance and some concerns that she had, that she had noted for me prior to her being reassigned to another unit. I know that this meeting was set up beforehand. I'd like to spend some time with you going over some of her concerns, if that's alright?

C: Fine, yea. We were going to meet.

2 B: Did Mary share with you some of the concerns she had, prior to this?

C: Not really, no.

3 B: OK. Um. A couple of the issues that Mary brought up, that she feels need to be worked with and which I'm going to work with you on have to do with your performance, and there are some concerns with how you are doing your work and the accuracy of some of your work and . . .

C: You're kidding?

4 B: No, I'm not kidding. And we'll go over them in a second. But in looking over your past performance report, completed last February, it's quite favorable and I guess I'm quite surprised to have these issues come up and let's talk about them for a few minutes. One of the things that Mary has identified that is a real cause of concern for me is that you are failing to complete your assigned duties and ah. . .

C: Oh, I think I'm doing the same as everyone else.

5 B: Could there be things with your work or your case load that are giving you problems?

C: Well yeah, my workload got tripled. After I didn't get my promotion and Stan got my promotion, now I got all of Stan's work, I'm training the new guy and now I've got a new supervisor. Quite frankly, the amount of work I'm doing here is tripled.

6 B: OK. You feel greatly overworked. Now, you mentioned that your work is tripled, and being new to the unit, how does this compare to the rest of the co-workers? Ordinarily I would think it compares favorably with them.

C: I wish it did. I think I'm doing a lot more than most of them. I'm the lead worker in the unit. Everybody comes to me with their concerns and I'm trying to get my own work done.

Script 4
PS=5, IR=6

Everybody else is screwing off. I'm just getting tired of it. I can't handle all that stuff.

7 B: Umm, part of getting tired of it, I guess, is another concern I have. It appears that you're missing more time than you have in the past, and you're also tardy on more frequent occasions.

C: I don't think I'm more tardy then anyone else.

8 B: Well, in looking over the records that I have, that fact is held out as being true. There's a change in your attendance pattern and a change in the way that your work has been going. What I'm hearing is that because you didn't get promoted, it's affecting how you're going to do your job and whether or not you're going to be here.

C: (Pause) Well, it's certainly affected the amount of work I'm required to do. It's tripled. Well, I'm doing as much as I can, but a guy burns out after awhile.

9 B: OK, so while the other staff members are screwing off, you're overworked and you're burned out and frustrated because of it.

C: Absolutely, absolutely.

10 B: One factor that may have an effect on feeling overworked is attendance. Mary has mentioned that when you are out sick you don't let us know. You've got cases to cover. You've got clients to see. We had a person who was evidently your case, that was so upset that he showed up at the director's doorstep at 10 in the evening one night.

C: Oh, is that where this is coming from? Oh, I should have known that. Some ringer. You know it's . . . the problem is the secretary, when she takes messages and stuff she just piles them up on people's desks. Yea, I'm out seeing clients and stuff in the afternoon and a lot of times I've seen, you know, ah, even Stan, the guy who got the promotion, more than once I've gone and followed up on his phone messages because he didn't make it back to the office, even though he said he was. He gets the promotion and I'm handling his phone messages for him. The secretary shouldn't let phone messages pile up on my desk if I'm not going to be there.

11 B: OK. The secretary is putting messages on peoples' desks when they're away. You feel dumped on for having problems now when the secretary could have handled it before. But I think that's a bit off the issue. The issue as I see it is that you're failing to call in when you're sick. So I don't know whether or not I'm going to have somebody to return those

Script 4
PS=5, IR=6

calls or whether I should assign someone to provide coverage on your case load during those days.

C: Well, If I'm not here by nine o'clock, I think that you could figure out that I'm going to be sick that day.

12 B: Well, that's . . . the rule governing you as a County employee is to let me know within one-half hour of the start of the workday that you're not going to be there that day, or that you're going to be late. That's one of our work rules. I don't want to put another employee in the same position you've just described with Stan.

C: I didn't think that it was a problem. But . . .

13 B: That <u>is</u> one of the problems, Chester.

C: That's no big deal.

14 B: Good. Thank you. Another thing that Mary has mentioned is that she has some suspicions that you are drinking over your lunch hour. As you know, this is contrary to the County Personnel policy.

C: Oh, name me three people in the office who don't go out and have a couple of beers over lunch on Fridays.

15 B: Now that isn't the point, we are talking about you. We're dealing with your situation as it affects your work performance.

C: It's not affecting my work performance, I don't do it all the time. But you know the whole staff goes out to lunch together and I've never done that and now I've decided to be one of the gang. I'm doing what everybody else is doing. I don't think it's affecting my performance. I don't think it's anybody's business.

16 B: OK. The next thing that I think you and I need to talk out is your feelings about that promotion. Could you tell me what happened? Is this real recent? Why do you feel that Stan got the promotion on you?

C: Well, I handled his workload for a couple of months while he was out brown nosing the hiring authority and stuff. I mean he's out there going out to lunch with them and joking with them during the day and I'm doing his work for him. He just politicked until he got the promotion. I ended up doing his work.

17 B: I can appreciate your feelings on that, I really can. You are overworked and hurt because of the lost opportunity. But I would say that since you've been passed over once, that won't mean that you won't have another opportunity. The

promotions are competitive, but supervisors get selected a couple of times a year.

C: But the next opportunity's not going to be for five years.

18 B: Now, we don't know that and I guess I think that you need to continue to provide the kind of performance that you were providing back in February.

C: You mean you want me to go back to doing everybody else's work for them. I don't want to do that. I'm tired of doing that.

19 B: I'm not asking you to do somebody else's work.

C: Well, someone's got to go out there and help those people who can't add numbers correctly and that kind of stuff.

20 B: That's my job. I guess I'd like to get things back on track with yourself bringing your accuracy back up and improving your performance again. Back in February, Mary's performance report on you indicates that you had no days lost and that you weren't tardy. And now it appears, on the contrary, that you are missing days and that you are tardy on a regular basis. That's quite a difference.

C: I get real tired watching everybody else blows off on Friday afternoons. Boy, since horse racing track opened up half the staff blows out of the office at noon on Friday, or Wednesdays, even.

21 B: Well, that's another issue for me to be aware of then. Something that I can take a look at and monitor as the new Unit Supervisor.

C: Oh, if you could handle that . . . I just get real tired of watching these people take off and screw off all afternoon, and I sit here holding the bag. I've been here for nine years, I've been doing a good job like you said; I've been training people.

22 B: Yes, that's evident and I want you to continue in that capacity. You feel disheartened after not having your hard work pay off. But, losing the promotion is most likely a result, not a cause here. Now um, as I've just told you regarding calling in, I need to hear from you. If I don't hear from you, next time I'll have to put it in writing for you. I don't want to do that, but I will. And then further disciplinary action could be taken. I expect you to be here.

C: Wow, I don't like it. Because I missed this one guy's message. He's probably a ringer. He's probably from channel five.

Script 4
PS=5, IR=6

23 B: I'm not talking about the one client. I'm talking about a series of events. The other area that I would really discourage you from participating in is having a few bumps over lunch. I know that there are some other individuals who do this, but when it's noted here by Mary, I don't know how frequently it was, but she's noted it, I'd like you to refrain from that.

C: That's no big deal, I'm just trying to enjoy myself. I don't want to burn out here; I'm getting real tired.

24 B: No, I don't want you to burn out here either. Let's resolve these isaues first. We should probably get back together at some future date and talk about how you're doing. I'll have a chance to look at some more cases that you've got and I'll see how the accuracy is then. I'll have more time to see how you're doing. If everything's back up to par, then I'll be able to find any workload problems.

C: I don't want to have to do everybody else's work, though.

25 B: The additional time will give me an opportunity to look at the Unit's statistics and the comparable jobs that people are doing, the volumes, what's going in, what's going out. I'll take a look at it. I'm not going to guarantee that I'm going to reduce your case load. But I want it to be comparable with the rest of the people who are doing similar jobs to you. That's the only thing that I can make in that respect. But I'll make a commitment that I'll look into these things and that we'll work to try to resolve some of these problems, provided that you follow the rules we've discussed today.

C: Alright.

26 B: Good, Thanks a lot for your time.

File PS5IR5 Script 5
Problem Solving = 5 Interpersonal Relations = 5

1 G: Thanks for coming in today Chester. Umm, I'm not . . . Thank you for taking the time away from your desk to come in and meet about something that . . . has just been given to me, and that we feel we need to talk to you about. Umm, a memo came from your previous supervisor regarding some potential problems that she was aware of, but it's really uncertain if she addressed these particular issues with you. Ah, do you know what's being referred to at all?

C: Well, I suppose . . . no, Mary never mentioned anything.

2 G: Okay umm, the . . . I guess the reason we should talk and . . . about it now is that a . . . a client complaint did come to our attention. Umm, apparently one of your ah, clients did not get the benefits that you were umm, working on determining and brought the complaint on up the ladder all the way to the director. And we're concerned as to why you didn't act on this particular case. But, well, umm, we should go back just a little bit further. Parts of your file were given to me, and I've looked at your previous evaluation, which was done about six months ago. And, at that point in time, it looked like you were really doing some pretty good work, you were rated as . . .

C: I was . . .

3 G: . . . fully capable or above in every category. There weren't any improvement umm, improvement needed categories pointed out at that point. And we're a little bit concerned now because your performance at this time, six months ago and what's coming from your previous supervisor now, seem to indicate that there may be a decline in your performance.

C: Yeah, I heard about that. But geez, I was sick that day. Umm, you know the secretary when she gets these phone messages from people, all she does is she comes and she sticks 'em on your desk. Well, you know, if I wasn't there that day, the messages just kinda pile up. And oh, you know, I just assumed somebody would be able to see that. I do that for other people, you don't see their messages piling up when I know they're not going to make it back in the afternoon, and stuff. And I'll handle that for them and nobody handled my message for me. I just feel I got a bad deal.

4 G: You're, are you aware of the County . . . well, you should be aware of the County rules regarding calling in if you're not going to be reporting that day.

C: Well, you know, if I'm sick, and I'm oversleeping and stuff like that, a lot of people do that, you know. If I get sick, I get sick. I can't help that.

Script 5
PS=5, IR=5

5 G: Well, that happens to be one of the County rules. Wouldn't the messages pile up on your desk because the staff was expecting you to be doing your job that day? And with this, your particular job it's very important that your supervisor and co-workers in addition to the secretary know if you are or are not coming in that day so that adequate coverage can be planned.

C: Well, I get the feeling that the office comes to a grinding halt when I'm not here. I mean I'm . . .

6 G: Well . . .

C: doing a lot of other people's work for them. Even Stan, who got my promotion on me, he, ah, I'm covering, still covering a lot of his work. I'm covering a lot for the new guy that got Stan's old job. I'm covering a lot of his job. Ah, even you, you're brand new and I don't mean any offense but, you know, I really feel like I'm doing a lot of extra people's work around here. One day I, a couple days I get sick and don't make it in to work, then, ah, it's all of a sudden jump all over Chester's back.

7 G: But, without you calling, we were uncertain if you were going to come in within the next hour or two, or what exactly was happening. So, that, that's one particular issue, that of the excessive absenteeism, the fact that you don't call to let us know, so that we can plan . . .

C: I don't think I'm absent all that much. Geez, I'm doing what everybody else is doing.

8 G: Well, let's not compare your performance to others. We're looking at just your performance com . . . now, compared to your performance in the past, and you had a very good attendance record six months ago.

C: I was really bustin' my butt for these people around here and, and umm, I got on the, you know, I was working real hard to get my promotion and, and ah, I think I'd make a good supervisor and ah, could be, you know, people see me as the expert around here. I'm constantly handling problems and, and you know, working extra and all that kind of stuff, and then all of a sudden, you know, Stan gets my promotion on me, then Mary, the supervisor, she, she takes off and now I'm stuck, you know, doing everybody's job for them and I, I get real tired of that.

9 G: Well, do you think that some of these things that are umm, occurring now that didn't occur in the past, like your troubles in getting a promotion, had something to do with why you're having troubles now, with why you requested a transfer, your feeling of frustration and dissatisfaction and your not getting that particular job?

Script 5
PS=5, IR=5

C: Well . . . Here I've been doing all this . . . I've been doing this great job for, for nine years I've been with this department, and, and ah, now I get my shot at the promotion and they give it to a guy who I was covering for him. I was handling his phone messages, you know, he, they do little tricks, you know, sneaking out of the office during their horse racing season and stuff like that. That's no surprise to anybody.

10 G: Well, I think that's a little bit off the issue, umm. Let's just stick to what's happening with your immediate job right now as a PFW, as a lead worker . . . Just going back a little bit again. O.K. We talked about attendance, we talked about you calling in when you were sick. There are a couple of other issues here, too. Well, a, one is, or an additional one would be spending a, quite a bit of time talking with your co-workers, socializing when you should probably be focusing in on your workload.

C: Oh man, you want me to go back to, to being the only person in the office that's working (exhales loudly). I'm getting real tired of that. I mean, you know, shoot, a guy's got to say "Hi" to people just to keep from burning out. I'm getting real tired of it.

11 G: Well, a little bit of "Hello" and socializing is fine, but it seems as if, from what we're getting from your previous supervisor, that was getting a little bit out of hand even if you do feel angry about it.

C: Oh, I don't think it's any different from what anybody else is doing. I'm doing my work . . . It gets done. Except for this one character who (laughs) . . .

12 G: Well, it's not really a humorous situation, and umm, this may have just generated a, an occasion for me to start mentioning these things to you. Is there anything else that should contribute to this, these factors that we've mentioned, other than what you've been explaining about a, your feelings about not getting that job? Anything else that would, that I should know about that may have an effect on your performance and on your satisfaction with your job? It would be good to see you be able to turn your performance back around to where it was six months ago. I think you, you've been with us a long time, and you're really, obviously have done your job in the past and, you'll certainly need to keep that up. To be fair to yourself, you've given us a lot of good years and put out a lot of good work.

C: I hope this ringer doesn't kill me off, because that's really what it sounds like is going on . . .

13 G: Well, o.k., let's look at it from this poi . . . from this point forward. These issues need to be brought up to you.

Script 5
PS=5, IR=5

It would be good to not let this one incident cloud things. It would be good to see you be more satisfied with your job. I'd, let's start off . . .

C: Good.

14 G: on a clean slate. O.K., we've looked at these various issues. Now, you . . . you are expected to call in if you are not going to be here. It's important that we know.

C: Well, I just got sick. It's not a big deal.

15 G: Well, that's just a basic County rule that shouldn't be a problem. If you're not going to be in, tell us . . . And that . . .

C: Umm.

16 G: that shouldn't be that difficult of a thing . . . umm.

C: Oh, man.

17 G: So let's not let it get to the discipline process. There's several stages that can be used.

C: One slip up . . . and now I'm being threatened by heavy discipline, or what?

18 G: No, I . . .

C: Holy smokes. I had no idea . . .

Simultaneously

19 G: This rule applies to everybody. You're surprised by a rule you've known about since you were hired? Everybody's expected to call. Everybody else does call.

C: Boy, I wish that were true . . .

20 G: So you are expected to tell us, too.

C: I've covered enough people's messages during the day when they . . .

21 G: They probably . . .

Simultaneously

C: take off, and don't do anything.

22 G: No, you don't really know that for sure. If that were the case, then that . . . those individuals should have been talked to, also.

Script 5
PS=5, IR=5

C: I'm the guy being picked on, so I can't do what everybody else is doing. O.K. That I, that I see.

23 G: Alright. Umm . . .

C: (sarcastically) Thanks.

24 G: And, you know, please try to focus in on your work, umm, maybe trying to process your cases a little bit faster. You're capable of doing that, you have done quite well as a Principal Financial Worker. I can see that from the past. Get back to your former level of quality.

C: So go back to . . . grinding my butt again and watching everybody else horse off, well. Aw, man . . .

25 G: So, just, you should just be concerned about coming in to work and what's happening in our unit . . .

C: Then I don't have to go . . . continue on doing all this stuff that I'm doing extra, like training the new folks, and handling a lot of the internal problems because . . .

26 G: We, well . . .at a minimum it would be appropriate to get your accuracy and case processing speed back to the level you were doing in February and especially call the office when you are going to be out sick. That's just common courtesy.

C: You're not really up to par yet on . . .

27 G: We, well, we'll appreciate any assistance you can give. You really certainly do have the background, and we'd like to see everybody work as a team. If you don't feel that you want to do that, we're certainly not going to force you to, but we will be monitoring how your case load is going and reviewing your accuracy and promptness on completing assignments. But we'd like you to feel free to, umm, talk to us about anything that's giving you any problems, even your desire for promotion. Come in and talk if you like. I would hope it, hope this incident doesn't happen again. Let's not make it into a pattern. So it would be good to get your attendance and . . .

C: Well, if I see bozo's files again, I . . . I'm certainly going to jump on them.

28 G: Alright, well, thank you for talking about it and let's . . .

C: O.K.

29 G: get on the right track now.

C: O.K.

File PS5IR3 Script 6
Problem Solving = 5 Interpersonal Relations = 3

1 S: Chester as you may know I've been just recently assigned to this unit and one of the first things to try to do is to get to know all of the people in the unit. This meeting was started with you simply because the meeting was already set up before I started this position so I'm not picking on you in particular. The same process should be followed with all of the employees who are having problems, and I would like to take some time today to discuss the situation that I have been lead to understand from the previous supervisor, get your input and your feelings as to how I can solve the situation as it stands today.

C: Alright.

2 S: Mary indicated in a memo that she left to me that she had been prepared to address what she found to be a performance problem . .

C: What!

3 S: which she did not do just because of her transfer to the new position.

C: Oh?

4 S: You're surprised? Come on. How could you believe that your performance wasn't perceived as negative in any way?

C: Well, I . . . I think I'm doing a good job.

5 S: Specifically, Mary has indicated and your performance evaluation has indicated that you had been doing pretty good job. But, your performance has been deteriorating so badly that we even had a client of yours get so angry at your failure to keep an appointment that he caused a scene at our director's house last week Tuesday night. Now you and I and the unit are involved. You've started to skip work entirely some days and not even bothered to let us know. This issue of service deterioration has become quite important to the department. What's been going on?

C: Well other than the fact that I didn't get my promotion but ah, I mean geez, I mean look what happened; Stan and I were both ah, going for that promotion and quite frankly Stan didn't deserve it at all. He got it and ah, I had been doing his work for him, I mean you . . . after awhile you get real mad about that, I mean geez, the guys out politicking for his promotion and stuff and talking to folks and downtown and ah, I'm handling his messages for him and stuff and ah, he gets the promotion and I'm here stuck.

Script 6
PS=5, IR=3

6 S: One of the other people in the unit got the promotion you wanted, so you got mad and started causing all these problems I mentioned. But I'm not sure what promotion was involved. Was he promoted to supervisor?

C: The new Unit Supervisor, yea.

7 S: So he was promoted by the department?

C: Well we both got on the list and Stan was the one who was able to ah, talk his way into the job. So, I mean, I don't . . . I don't know why I wasn't promoted; I've been doing a good job here for nine years. I've been helping people out and stuff, I've been helping Stan out and all of a sudden look what's happened. No offense intended, but you're a brand new supervisor, Mary takes off on me, Stan takes off on me who was . . . and now I'm . . . I'm not only training the new guy, his replacement, but I'm handling my own work load and a lot of the stuff that people went to Mary about or Mary had, you know, me taking care of, now I'm doing that too. I feel like I'm doing three people's jobs. I'm tired of it, I don't want to do that any more. I don't think I'm getting any the reward that I deserve. They give Stan the reward while I stay around doing everybody's work for them.

8 S: The fact that you haven't received the . . . what you perceive as your, as the . . . the reward or the recognition for doing your job doesn't mean that the next time there's a position available you won't be given every consideration that you deserve.

C: Oh . . . the position won't open up for another five years.

9 S: Any future promotions will assume that you come to work on time and not call in sick alot without even bothering to call us. You are a Principal Financial Worker. The good ones are also expected to be a leader; to assist other workers, to . . . to be a trainer ah, in effect to act almost as a supervisor of the unit.

C: And I've been a good one of those and look at the reward I get for that; my workload triples.

10 S: The situation that I read about didn't give me any information about Stan. So I can't sit down and compare the two performance records since his went with him when he was promoted. Based on what you're saying and what I read, you were doing a good job based on this prior performance evaluation. There doesn't seem to . . .

C: I think what I'm doing is a pretty good job.

11 S: . . . be any question about that. How that compares to Stan and how the promotion decision was made is more your concern than mine. But promotions are not based just on what you

Script 6
PS=5, IR=3

have been doing in your current job, that's a big factor in it, but the hiring people have to try to predict whose going to be the best at the higher level job. And I think, if in fact the . . . the performance problem that Mary has reported is true--that you're not coming to work on time and skipping days, and your productivity has tailed off--that indicates that the decision that was made was a good decision. That they did . . .

C: Oh, man.

12 S: . . . decide that for probably justifiable attitude reasons you are not ready to be a supervisor. And what I would like . . .

C: I've been doing it, I've been doing it for ten months here.

13 S: You're not a supervisor. You're . . . you're a lead worker which is kind of an assistant supervisor, while you were doing some of those functions that were delegated to you the unit had a real supervisor and she was responsible. So yes, you were supposed to be assisting her in doing some of those duties, but you shouldn't have been functioning as a supervisor. But you don't even come to work on time, some days you don't even bother to come to work at all. Couldn't you just forget about that and put that behind you that a . . . in that instance of not being promoted? Don't let your feelings of frustration get in the way of your doing your job. You've got to take control of this problem and resolve it yourself. Clearly you're capable of performing at a good level based on your prior performance evaluation. You must understand that if this current performance continues, it potentially could lead to disciplinary action on the job. . . Come to work on time.

C: Oh man, I don't think I'm doing bad. I'm just doing the same as everybody else. I don't . . . I'm just tired of doing everybody else's work for them.

14 S: You say you're doing more than people in your job class. You'll have to prove that.

C: People are coming to me all the time, I got all the different staff members asking me questions all the time.

15 S: The people in this unit?

C: Yea, people are coming to me. I'm . . . you know, I handle the difficult cases . . . I'm training a new employee, some of your stuff is coming to me. I don't know, I mean it's just cause Mary's gone or in the process of going, but ah, shoot . . . I used to handle Stan's all the time. The secretary leave messages and stuff on the desks and I'd see Stan would have a stack of them piled up and I'd take care of

Script 6
PS=5, IR=3

his messages or deal with his clients out . . . out in the . . . in the lobby when he was out politicking for his promotion even. So, you know I was doing a good job. I still think I'm doing a good job but I just don't wanna have to kill myself anymore. I'm tired of it. What's the reward?

16 S: The reward is being satisfied with your job, number one; and number two, demonstrating your potentials so that you can get promoted.

C: Well that didn't work this last time.

17 S: Well, there's more than one opportunity.

C: I doubt it.

18 S: As a lead worker you are expected to do the lead worker tasks and to assist others. That's your job. That's why you're being paid at the higher level. If you do . . . if you've got too many job duties to be able to function, I will determine that as I become more familiar with the unit and verify that . . . that your workload is acceptable. If you are doing too much, don't kill yourself, but what we should try to gauge the level of duties that you're doing, the number of duties that you're doing and if it does seem to be unnecessarily high umm, then it should be corrected. But based on the complexity of the duties alone, any lead worker should be able to do it. So don't expect too much of a change. If you feel put upon in having to do these additional jobs, that's part of what we're paying you to do, because of your higher level classification.

C: So in other words go back . . .

19 S: Not if it takes you ten hours a day to get your job done, that shouldn't be the case. You should be faster than that. But you should be doing the more difficult tasks. And should be assisting me, because nobody knows the job better than you.

C: In other words, go back to what I was doing.

20 S: Yes, because you were doing a good job before.

C: But I wasn't getting the recognition, why didn't I get the promotion? Geez. I mean that's real frustrating to watch some goof-off get my promotion.

21 S: I can't comment on that. You might want to find out . . . you should try to get a reason . . . if there was a . . . a particular indication as to why you were not promoted. If you can do it without whining, the person you should talk to is the manager that chose your co-worker. It may well have been umm, something that you're not even aware of and so I

Script 6
PS=5, IR=3

think as a . . . as a, for development the County owes you an explanation as to what criteria was used in this decision. We'll talk about this again after you've gotten more information.

C: Alright.

22 S: So for two reasons: 1) so that you can feel better that a good decision was made and 2) so that you know how to promote to a future position when a similar vacancy comes available, you know that you have a shortcoming in one area, you can take control of the problem and work on that area prior to the next promotional opportunity. One area of improvement is certainly your attendance and promptness. Another is to do the job that you now know is expected of you.

C: I wish it was that rational.

23 S: Well, I hope . . . it should be and this should be pursued to get an explanation on the basis for the decision.

C: OK.

24 S: I appreciate the opportunity to sit down with you, but I just want to remind you that I intend to go through this type of session with all of the employees in the unit and the only thing that has made you first on the list is the fact that Mary had already set this up before she left because of your problem. And umm, while it is a serious problem that needs attention, and serious disciplinary consequences could happen if its not corrected, you should be able, based on your prior performance, you should be able to do the job in this division. You are expected to get your performance back up to standards. And for sure you've got to come to work on time or call me if you're ill. Your following of the attendance rules will be watched closely. Don't let your bad attitude get in the way of following the rules. And that you look at it from a positive standpoint and put that one behind you and work for the future promotions for the good of the unit.

C: OK.

25 S: Thank you.

File PS3IR6 Script 7
Problem Solving = 3, Interpersonal Relations = 6

1 B: Good afternoon Chester. As you know I've just been newly assigned as supervisor of this unit and ah, this meeting has been set up at the impetus of your former supervisor Mary, to discuss some problems that I'd like to go over. I think you and I are gonna be able to, you know, we'll be able to work the things out. So let's go over some of the background . . . Now, ah, things don't always go on a straight and favorable path all the time. Every now and then little things happen. And when I have a good employee, and you certainly are a good employee--I saw several high scores on your performance appraisal report from last February--if certain little problems develop or if there's some explanation for performance falling off or something, I like to talk about it, because if I've got a good employee, I want to make sure that he stays to be a good employee. If there's something I can do to help him out then that's what I want to do. I'm not making any, any definite statements or charging you with doing anything bad or anything like that, I just want to see whether I can figure out something that's happened that's come to our attention. Now, just to make things a little clearer, the information that's come to me is this: that over the last five or six weeks or so, ah, there seems to have been an increase in the number of problems occurring and a general decrease in your performance quality. That's what I wanted to talk to you about a little bit.

C: Oh man. Is this the "Go back to the grindstone" speech. Umm, I been working my butt off in this unit for, for what, about ten months now, and I've been in the department over here, for, ah, for nine years. And, ah, I've always been working my butt off over here, and now I've loosened up a little bit and do like everybody else does and all of a sudden they get on my case. I'm getting a reprimand from the brand new supervisor, part of whose job I think I'm handling right now, you know. I'm doing so much stuff around here. It's really getting to me. I'm getting tired. Geez, I've been doing everybody's work for them.

2 B: You've been working hard and decided to ease off a bit because you got tired of it. As we know, you earn the time, you take the time off; but, over the last five or six months, the information is that this seems to be getting more of a pattern. I seem to have more, ah, more social life, more use of alcohol at lunchtime, some sick leave issues, things like that. Are you feeling overworked? You say that now you're doing a lot of work?

C: Oh, I'm training the new guy from, from Stan that got promoted, brand new supervisor, and I should, I feel like I should have the supervisor's job. I took that test, I got qualified, and I wasn't hired, you know, and it, it gets real

Script 7
PS=3, IR=6

maddening where, where Stan spent his time ah, he was ah, umm, floating around and politicking for these promotions and stuff, and ah, and ah, ya, I end up doing his work for him, he gets promoted, now I'm still doing his work, now we have a brand new supervisor, no ah, no offense intended, you know, I feel like I'm really holding this, holding the thing together, and I'm getting really burned out, watching people float off on Friday afternoons, and late in the summer and the fall, going down to the race track and having a good time, and I'm sitting here trying to get caught up on stuff. I'm, I'm real tired of it. I think I should, you know . . . I'm just going to do what everybody else is doing, you know, I, I don't think, well I should have to bend over backwards to keep everybody else going.

3 B: Now as far as I've been able to tell, in ah, getting some information from your previous supervisor Mary, she didn't seem to be having the same type of problems for the last five or six weeks with other workers. The evidence that you alluded to, that other people are taking off, Mary and I haven't been able to verify those sorts of things. But we do know . . .

C: Boy, it's happening all over the place. Take a look.

4 B: I would feel frustrated too, if others were not working while I was. Well, I'll look into that. But I do have information that says your pattern of work has changed enough for Mary to notice it over the last five or six weeks. And what I'd like to know about is whether there are any things you think that you need. How have you been feeling lately?

C: I'm real mad, I mean, for nine years I've been humpin' my butt off going through the ranks, I've been promoted a couple times, and now all of a sudden I see that, you know, really what it is, is umm, you know a, Chester becomes an expert and you just dump all, all the nasty stuff on Chester's desk and everybody else is just screwing around. Well, I'm tired of it. I don't think, why I should have to deal with that, you know. I'm not paid any more than anybody else out there to do any more work. You know, I took the promotional exam. I deserved it as much as anyone else, and I get stuck because of it, you know? I, I don't see why I should have to work any harder than anybody else.

5 B: So, you feel cheated that they used your, your qualities as an expert too much, and then you didn't get credit for it on the promotion.

C: Well, yea. My workload doubled after, ah, Stan got the promotion. And umm, you know there's nothing, you know. First you get insulted by not getting hired, and then you get double insulted by having your workload double, training the new folks, and covering for them and that kind of stuff. You

Script 7
PS=3, IR=6

know, it's, you know, if I could maybe see some appreciation, and get my promotion or something, then I would.

6 B: Of course, I can examine the workload. I'll be doing that. I can look at that and see whether I need to equalize that. Also, ah, let's review the results of the promotionals to find out whether or not there were some reasons for your scores. Something might have occurred during the interviews, or maybe the questions weren't answered to the satisfaction of the, of the panel. You'll need to get some data on that from personnel. Are there some reasons for that, maybe you can . . .

C: So Personnel kinda had it in for me?

7 B: Well, I don't mean that. I mean that Personnel will have information within their records, because they get the, they store the tests and they will be able to give you some advice as to how to go about better answering questions, or how to better present yourself in the interview or whatever it is. You'll want to talk in general with personnel about that. That may be of some help to you, to help you out. Because, apparently, you have the knowledge base and your, your background seems to be fine. I doubt that you missed the promotion for lack of technical knowledge. But maybe there's something else that's keeping you from being promoted. What I'm saying to you is that there may be a problem with answering questions, whether you hesitate or whether you grope for words or things like that. These are the kinds of things that Personnel could probably tell you. Following up on that would improve your chances of promotion. Because, based upon just looking at your record and all your ratings, you seem to have no problem.

C: But, you know, I did pass the oral, and I got on the list. I just wasn't hired, so . . .

8 B: You were eligible, but didn't get hired. Not knowing why must leave you wondering. Was there an interview with the person who was hiring the supervisor then?

C: I've, I've talked to a couple of folks. I don't know if it was an interview or not. They said come out and talk. So I did. I'm not sure what positions were all open, but I went around. 'Course, you know there's so much work to do that I, I, I kinda had to cut those a little bit short. You know, I'm covering for ah, all the folks who are out, who were actively going out to lunch with the supervisors and stuff like that. You know, there's a lot of work around here and it's . . .

9 B: Well, when you go for the formal interview for selection, those are the ones that you don't have to worry about getting special permission to attend. Informal discussions with

Script 7
PS=3, IR=6

people, where it might help your case, for those you need to check with me. Those things cannot interfere with your work. We're going to be talking about that in our unit meetings. I'm going to be covering the use of time. So count on that. And I think these can be some of these things, if they're wrong, that can be corrected. But ah, what I was trying to get at today was whether or not there was anything else that I might be able to help you with?

C: Well, give me my promotion, you know? . . . (Laughs)

10 B: (Laughs) I mean other than that.

C: I don't think that, that I should have to do other people's work for them. You know, take care of that, that might be real helpful.

11 B: Yeah, I can do that. And, I'm sure that in the unit meetings I'll be going over that. But, ah, I think that, umm, I think that you are aware of the policies. You'll have to follow them. That's just the policy.

C: Ugh.

12 B: Now, if you're really interested in finding a promotion--and it certainly sounds like you are--then you're going to have to take a serious look at the reasons why your actions are being viewed as being below that level. Chester, you can blame Stan for the problem or you can modify your below supervisory level behaviors. Ah, as for the increase in problems, is there anything else, other than what we talked about, that could be causing these problems? I'll need to assess the reasons in order to help you resolve the situation.

C: (Breathes loudly) Well, (shrugs) I'm just, I'm enjoying it more. You know, I'm, I'm, umm, you know, everybody else is, you know, how many hour lunches do you, have you seen me take around the department. I stick pretty close to that. But, you know, I'm going out to lunch with the folks, rather than grabbing a quick sandwich and jumping right back into the paperwork and, and umm, you know, people are screwing off and stuff and, and umm, I'm not breaking any rules, but umm, I don't think I'm breaking any rules, and I'm just tired of grinding my butt so much. So that's, that's really what it is. I'm, I'm, I was really insulted when they, you know, I've been doing such good work for so long, and then all of a sudden, you know, I don't get my promotion, and it's dump on me, more work, more work, I'm just tired of it. I don't know, you know, I, I'll try to deal with it.

13 B: Good. I think what I'll probably do is, well in my beginning meeting, I'll be examining workload and ah, the habits of workers and I'll set some unit procedures, and see whether I can get a line on case load size, duties and that sort of

thing, and get that spread out. And then I think if we can do that, you'll feel more comfortable knowing that the work is all shared. That everybody's doing the same thing. . .

C: That would be good. I'd buy that.

14 B: Well, I don't want to prolong this a long time. Is there anything you'd like to add?

C: No, I, I, thought I was doing reasonably well. I just was not going to, you know, do overkill like I've been doing. I really do think that I deserve that promotion and ah . . .

15 B: Well, I think you'll work toward that. You need to work toward that, and work as hard as you did in the past. If we can get you in the habit of following the rules and doing the work expected of you--we cannot have any recurrences of the, ah, the use of alcohol on the lunch hour--if we can, we can avoid those sorts of things, then I can look at how the unit is operating. When we get the right atmosphere here, then we can get you in a stance where you can be looked at as more promotable. I don't know why you didn't get promoted the last time, but ah, if we can correct some of these things, then I'm sure it will be better the next time around.

C: O.K.

16 B: Oh, well, thanks for coming in.

FILE:PS3IR5 Script 8
Problem Solving = 3 Interpersonal Relations = 5

1 S: Ah, good afternoon Chester have a seat please. Ah Chester, unfortunately we've been left with a . . . with a task here to. . . to accomplish that ah, ah, is not a pleasant one. Umm, what's happened, Chester, is that the former supervisor, as you know, has been rotated out of the unit ah, on quite, quite short notice. But we've been informed of a very serious problem that ah, that has occurred just recently and it had to do with ah (PAUSE), it has to do with a client complaint.

C: (Exhales) Oh yeah, that story.

2 S: Yes, yes the day that the client came in and ah,

C: Ended up on the Director's doorstep at ten o'clock at night, I heard about that.

3 S: That's right, that's right. Obviously this was a very upsetting situation, both for the supervisor and for the Director and now for both of us. There was even mention of possible consequences for you.

C: Yea.

4 S: Since I don't really have any background as to what did happen would you tell me in your . . . from your perspective what went wrong.

C: Well, I don't really think it was my fault at all. I . . . I was sick that day. Umm, you know, people get sick. And . . . and then you know, finding out about it afterwards . . . it's . . . it's the receptionist. She, ah . . . she takes phone messages and client messages and she just sticks them on your desk. Well, if you're not there they just kind of pile up on your desk. And . . . and so that the messages pile up on my desk and it was sitting there the next morning.

5 S: You mean you weren't here?

C: Yea, I was home, I was sick.

6 S: You weren't here at work?

C: No, I mean I was sick that day, and . . .

7 S: Apparently the client wasn't directed to somebody who was on the premises who could help. I wonder why?

C: Well the receptionist takes messages, so the message came back, stuck on my desk. Usually (exhales) I don't know if he was here physically or if he called up or what, but I know he

Script 8
PS=3, IR=5

was real concerned about getting his check. And when I got back the next day the message was sitting on my desk. You know we go out in the field and stuff and that's . . . that's not an unusual occurrence to come back and have a stack of messages on your desk, but I'm . . . I've . . . you know being the Principal Financial Worker here for . . . I've been in the department for nine years, I've been in the habit of really helping out people and you know you see Stan, the guy who got my promotion, he . . . gets messages on his desk all the time and . . . and ah, I'd help him out, check them out and no one did it to me. I think the problem is that no one checked the messages for me.

8 S: (Chuckles) Sounds awfully reasonable.

C: I really think so. I mean, it's . . . I don't really think it was my fault.

9 S: Well, we would have to agree with you then, ah, I don't . . . being a new supervisor here in the unit, the circumstances and the procedures are unfamiliar to me, but I don't see either -- if you weren't here on the premises -- how blame could be placed on you for not seeing a client . . . I don't know. But, if you actually were not here at work ah, how you can be blamed for this incident?

C: Yea.

10 S: But, beyond the fact that this individual, ah, was not ah . . . was not properly taken care of, it seems that there have been other incidents. Umm, your former supervisor has stated that ah, that you haven't been able to accomplish work within deadlines that have been agreed upon; that you have requested to get out of this unit; that you've been socializing a bit too often, and you've even been noticed using alcohol over lunch. There appears to be a problem here. Your former supervisor has even sought out the details of the discipline process. She was quite concerned about those issues, even though perhaps this incident can't be laid onto you. Do you want to talk about some of that?

C: I don't think I'm doing anything different than anybody else. Umm, for nine years I've been a . . . been a good worker in this department and . . . and ah, you know the promotional exam was held awhile ago and . . . and Stan get's the promotion. I've been . . . I've been helping him out for years and . . . and ah,all the while he was out politicking for his promotion and stuff, you know, talking to the new supervisors and all that stuff. I was over doing the work . . .

11 S: (INTERRUPTING) Alright . . .

C: . . . and all of a sudden Stan get's the promotion . . .

Script 8
PS=3, IR=5

12 S: alright . . . let me interrupt you . . . let me interrupt you, because we're not here to talk about Stan or anyone else in this department. Stan is an unknown to me. Let's talk about you. Ah, your performance, relative to no one else's, has changed it appears. Ah, your last performance evaluation in February of this year looks to be very good. It's ah . . . it's an above average type of rating. The information that has been left me however, relative to your performance recently, is something other than that. Now whether it's . . .

C: My (exhales) . . .

13 S: . . . whether you . . .

C: My work . . .

14 S: work with you.

C: My workload's doubled, you know, Stan left and ah, we've got, you know, a brand new person in here and cause I'm the Principal Financial Worker, I'm . . . I'm helping this new person get started and, no offense intended, but you're brand new in this department and whenever little glitches come up the staff come up to me, you know, and I handle the problem. So I really feel that I'm almost doing three peoples work, you know I do get the more sensitive case load. Umm, and, you know, all of a sudden there's this increase in there. I'm real tired of . . . of doing all this extra work and not getting any reward for it. I mean it's . . . geez I've been doing this for nine years and my . . . my chance for promotion comes up and . . . and they hire the guy who's . . . who's . . . who I've been helping. I mean that's real frustrating um . . .

15 S: Well ah, obviously, Chester, no one could sit here on their first morning in the office and make judgments about the quality of your performance or anyone else's. Ah, . . . and we certainly should be willing to start off on ah, a clean slate with you . . . I have got some, you know, I've got some concerns though, Chester that if there had been any problems, ah personal problems, health problems, anything that has occurred that preclude, has precluded you from attending to your job that, that you must give some thought to addressing that situation.

C: Oh, I think the only problem is that Stan got my promotion I mean, geez, what am I supposed to do?

16 S: Well,

C: . . . I'm going to do like everybody else does, you know.

17 S: Umm, as I say, Stan's not the person that should be talked

Script 8
PS=3, IR=5

about, um, but Chester, you appear to be totally crushed about having lost this promotion. I see your face turning red, in spite of the fact that I'm . . . trying to suggest that that's a fact that's occurred . . .

C: I know.

18 S: . . . you do seem to have a real problem with that.

C: I'm really hurt by that I mean, it was hard for me to watch someone whose, who was incompetent get that . . . get the promotion I deserve. I mean, I got on the eligible list. I went through the you know, the interview process and stuff, and they give it to the guy whose work I've been helping out. I mean, Stan didn't really know what was going on, I mean, he got by but I helped him out a lot.

19 S: You feel terrible. Stan's being promoted ah, is an important issue for you . . . there's been . . . there's been a change it appears from the time you were last evaluated and today . . . in your job satisfaction. How does Stan play into this situation?

C: Stan? . . . Well, I suppose that might be. it's just that, . . . I mean if . . . why should I? I should do what Stan was doing cause that's evidently what you have to do to get a promotion around here.

20 S: Losing a promotion is always a real disappointment. You sound like you haven't ever had a rejection of this nature before.

C: Well, . . . I just all of a sudden realized what you've got to do to . . . to get promoted here. Here I am doing all this extra work and all . . . the people that get promoted; they're the ones who . . . who kind of used me to get all the work done while they're out politicking to get their promotion, you know. Why . . . should I do all that stuff if I'm not going to get rewarded for it?

21 S: I'm sure . . . I'm sure that you were helping out. And I'm sure that it's a compliment to recognize the skill and . . . the technical ability that . . . that you do have.

C: I wish it was. I was training the new guy and handling staff questions and stuff like that. I mean you know I'm always the one that had to stick around and . . . and ah, resolve the issues. When everyone else was going out to lunch and ah, I'm always the one that gets started with a new problem, you know, right in the morning, you know and other kinds of stuff. And it's, you know, I . . . I really felt like I was helping out, and all of a sudden I hear like that's not what's being rewarded around here.

Script 8
PS=3, IR=5

22 S: Ah, It's not the only thing that is considered at the time that a promotion is being made. Ah, supervision and . . . and getting work done through others rather than doing it yourself is a . . . is a bit of a different technique too . . . I'm sure that your skills were being recognized by having you work with new people and by training them in . . . But let me suggest that that wasn't the only thing that was being considered at the time that the decision on promotions were made. Chester, I'd certainly like to help you in polishing the skills you need to be promoted. I'd like to talk to you about it and assist you in your development. Perhaps at another, more agreeable time in a few weeks, when you're ready, get back to me and we'll set up a time. I think together we could work this out.

C: Yea, I suppose.

23 S: Well (coughs) Chester, I'm ah . . . Now that we've cleared the air, (coughs) at least I . . . your feelings are clearly understood and we've had a chance to acknowledge at least that we're going to try to start fresh with each other,. For that I'm grateful. I'm just sorry that it had to begin on as negative a note as it did.

C: Yea.

24 S: I ah, I hope from this point it'll be a . . . a much more positive situation that exists between us. I want (coughs) to tell you though that if there is anything that . . . that ah, . . . that can be done at any time to resolve any concerns that you have about the work unit or me that you should feel free to try to do so.

C: OK, thanks.

25 S:. OK.

File PS3IR3 Script 9
Problem Solving = 3 Interpersonal Relations = 3

1 S: Hello . . . I'm, I'm . . . supposed to be a new supervisor.

C: Yea, hi, I'm aware that you are.

2 S: O.K. I'm supposed to first talk to you about the performance ah . . . last performance evaluation . . . And the poor performance issues that were put in the memo. Stuff like, ah, (READING FROM THE MEMO) "Chester fails to call the office when sick and is out of the office with colds." He, ah, spends too much time socializing, having beers over lunch and is slow in . . .doing his work. (PLACE MEMO ON DESK)

C: Huh?

3 S: Umm, the last performance evaluation was on 2/25/85. Has Chester's performance been reviewed, listing . . . any performance concerns, with Mary since that time?

C: No, Mary just took off like everybody else. No.

4 S: OK. There has been no directives from the former supervisor since that time or any feedback on performance?

C: No, I had my performance review back in ah, February.

5 S: OK. That was in the test material. Everything at that time was, I think it was at least meeting, ah, the supervisor's expected . . . expectations in all areas. The memo lists several problems that I am supposed to fix. Umm, how would Chester evaluate at this point umm, the case load, um, being able to meet um, time lines, and meet client's expectations?

C: I think you should know, or I hope you know, exactly what my case load is. It's doubled since Stan got my promotion. Gee, you know, we've got a brand new supervisor now, no offense intended but, holy smokes, you know, Stan takes off and I've been saving his butt forever and all of a sudden he gets his promotion. Now I'm training the person taking his slot and, you know, the people don't exactly know you yet so I've been handling all the questions and stuff and, geez, my workload's a lot heavier than it's ever been.

6 S: Is that creating problems in meeting time lines?

C: Yeah, of course it is. I'm tired of, you know, doing everybody else's work for them and not getting any rewards.

7 S: OK . . . But, where is it being a problem of direct client services?

Script 9
PS=3, IR=3

C: Where is it? Well, I'm doing everybody's work for 'em, I mean there's just not enough time. Stan was out politicking for my promotion and I had to cover his work for him. Answering his phone calls, and meeting his clients, and stuff and, geez, you know the work kind of drops off from there.

8 S: OK. This is a bit tricky. It wasn't in my memo . . . O.K., you already know your workload not getting done, because if you're doing Stan's and Stan's is getting done by you, but it sounds like . . .

C: Well . . .

9 S: . . . your job isn't getting done . . .

C: The new guy over there now, you know, he's catching on but, you know, geez, they don't send us people that are all that smart so it'll take awhile to get him up to snuff but I'm training the guy. I've always been the trainer so, you know, it's just I'm getting tired of not getting rewarded for any of this stuff.

10 S: Ok, then it's not really your problem. There's a number of things that will have to be looked into. I think that one solution is to do your job in a timely fashion. You mentioned some of the things that seem to be getting in the way as far as other people's tasks but I think that's irrelevant. Don't do other people's work. Just do your case load.

C: Yah, you're right. Umm, and I've always been a good worker. I've been here for nine years and I've been really busting my butt trying to be the good guy and always be the tops and stuff, and all of a sudden then this promotion comes up and, geez, you know, I get on the list and then they give my promotion to Stan. I've been covering his butt for years and, oh man, that's hard on a guy. Why should I be working any harder than anybody else if I'm not gonna get rewarded for it, huh?

11 S: The memo didn't mention this promotion stuff, but let's talk about your goals. Do you have any career objectives?

C: Well, my career objective is to get my promotion as the supervisor.

12 S: OK. Well, I mentioned that I wasn't told about any of that. But, you were doing well for nine months, I do know you were doing well last spring and that's evident on here. (MOTIONS TOWARD THE MATERIAL) However, according to the memo, it's this current performance, your ratings are going to be unsatisfactory on the next one.

C: What?

Script 9
PS=3, IR=3

13 S: If you're not doing your job . . .

C: Ah geez. This has got something to do with that ringer, doesn't it?

14 S: I don't know what you mean by a ringer, all I've got to work with is the memo.

C: Oh, man (breathes loudly). You don't even know what I'm doing. How can you tell me I'm gonna be unsatisfactory?

15 S: Well, you already told me you can't really get your tasks done on time didn't you? So I think . . .

C: I'm doing . . . I'm doing three people's work.

16 S: Well, that sounds like the problem. Just do the stuff you're supposed to do and finishing it quickly and accurately. That's something to do.

C: Well, geez, no one else's working that hard. Why should I have to do it all?

17 S: No, no . . . your task is to do your job. If there are problems with that, tell someone and they can look at it. If you feel like other stuff is being dumped in your lap you feel there's other problems like that, let someone know, but I think for right now just do your job. Especially if you're looking at a career development where you can be promoted to a supervisor position, you'll never get there without doing your job.

C: I'm tired of being the only one that has to do all this stuff.

18 S: Well, the work is going to have to happen; you're going to have to do stuff. Are there other problems? Like I said, let someone know what it is and they can look at those and address some of the side issues. But initially that . . . these are the things that are going to have to happen. What should be done at this point, umm, we might . . . what you need to be able to do at this time is have you have some pretty clear expectations. To clarify you know, the expectations, what these expectations are they were put in writing in the material I have and you can let me know about any comments that you may have when you read it.

C: Alright.

19 S: OK, as far as some of these . . . really you told me what the problems are. Umm, it's not Stan's problems or anybody else's, you know, let me . . .

C: Stan hasn't got any problems anymore, he's got my promotion.

Script 9
PS=3, IR=3

20 S: Well the new guy, what's his name?

C: I don't even know.

21 S: You're complaining about spending too much time training, yet you don't know his name? Let me . . . if he's having problems getting his stuff done, and that's relates back to you . . . that's becoming his performance problem. If his performance is affecting you, then we'll handle that. But I think you need to concentrate on your job, not on training. That needs to happen. The memo says that you could be getting a written reprimand on the matter.

C: Oh, man.

22 S: With a copy going to the personnel file and going to the department head.

C: Oh geez, gimme a break.

23 S: This is just the policy. Those are the rules that are followed when you don't do your job. You could also be getting . . .

C: (Exhales loudly)

24 S: . . . a one day suspension, again this is . . .

C: Holy smokes, gimme a break

25 S: . . . sufficient notice . . . Are you aware of this information?

C: No way.

26 S: OK, let me go over . . .

C: People get suspended because of . . .

27 S: Work rules violations.

C: Oh, man. I've been working here for nine years. I've been . . . I got on the promotion list. Oh boy, now I'm getting threatened by being fired, just because I . . . suspended just because I missed a phone call.

28 S: Well, I read about that. I'm not sure what the problems are though ah, what the process is.

C: Oh boy.

29 S: You might even get a five day suspension again with copies going into the personnel file, and to the department head, or you could even get a termination hearing.

Script 9
PS=3, IR=3

C: Holy smokes.

30 S: Umm, yea, the rules are pretty serious. You might get fired for this. I'm not sure how Mary handled it, I'm supposed to . . .

C: (Exhales)

31 S: to . . . to follow these procedures. So this is a serious matter. If you're performance has been satisfactory in the past, umm, and there's some reason it's not occurring now, then let's look at some of those reasons. Do you have any reasons? If there's things you can address appropriately, you should address them.

C: Oh, boy. Yes, I suppose.

32 S: OK, I'll . . . I will get you . . . why don't you make a copy of this stuff. Go over it and strategize how to keep working and look at your workload, then another meeting will occur where your ideas will be reviewed. OK? I guess that's all.

C: OK. Bye.

APPENDIX H

Mock Oral Exam Rater Training Procedure

1. Consent Letter

2. Background Information form on the exercise, Chester's script and the participant's accuracy of rating goal.

3. The Background Questions form.

4. The procedure of the oral examination research project. Breaks after training and candidate number five (if time).

5. The mock "candidate's" test Pre-work material.

   A. The memo to the "New Supervisor."
   B. The Principal Financial Worker job description.
   C. The memo from "Bob Zimmerman" of the Personnel Department.
   D. The Performance Review form.

6. Summary of the "answer" to the question.

   A. Attendance crisis issue.
   B. Long term performance deterioration issue.
   C. Long term job satisfaction issue.

7. The Rating Sheets used during the mock oral exam.

8. The Problem Solving dimension and Rating Scales definitions form.

9. The Interpersonal Relations dimension and Rating Scales definitions form.

10. The procedure for rating the mock candidates.

    A. Complete the candidate sequence information on the form.
    B. Watch the videotape without commenting on the performance.
    C. Quietly rate the Mock candidate using the sequence on the form.
       1. complete the individual item ratings.
       2. complete the individual subcomponent overall rating.
       2. complete the overall dimension and confidence ratings.
       3. complete the ratings in same order for the other dimension.
       4. complete the overall assessment ratings.
       5. complete the causal influence factor ratings.

    D. After rating all nine mock candidates, complete the Follow up questions form .

    E. Debriefing discussion.

08/24/87
PHTRAIN/PRORAL

Consent Letter

Dear Participant:

Thank you for agreeing to participate in this study of the oral examination process. The Hennepin County Department of Personnel is using this study to attempt to seek out methods and techniques for improving the selection oral examination process.

Proper professional practice for the conduct of research studies requires written documentation that participants are freely cooperating in this study and that they understand that they are free to stop participation at any time. The test material you see in these videotapes will not be used in any future selection devices. The data is strictly being used to improve the overall quality of the oral examination process. In addition, the study will not trick or deceive the participants at any time. Finally, all responses are strictly confidential.

The Personnel Department strongly urges you to participate in the study through to completion. If you wish not to participate, please feel free to take this option now. It would be most helpful if you stayed for the full length of the study. If you leave early, your data cannot be utilized. The information you provide will serve to improve our understanding of the selection oral examination process. Thank you for your cooperation.

Sincerely,

Bob Holmgren,
Project Director

I have read the above paragraphs and agree to participate in this study.

signed_____ date_____

August 24, 1987, PHCHESTR/PRORAL

Background Questions

1. Current Position: What is your current job title?

_____2. Management Experience: Do you now or have you ever supervised at least two subordinate supervisors of professional staff in an organization? (yes, no)

_____ If yes, how many months have you performed as a manager?

_____3. Supervision Experience: Do you now or have you ever supervised at least two subordinate staff in an organization? (yes, no)

_____ If yes, how many months have you performed as a supervisor?

_____4. Leadworker Experience: Do you now or have you ever acted as a leadworker over at least two other staff in an organization? (yes, no).

If yes,

_____ Did this leadworker experience include conducting formal performance reviews with subordinates? (yes, no)

_____ How many months have you performed as a leadworker who conducts formal performance reviews?

_____5. Counseling Experience: Do you now or have you ever been trained and or acted as a counselor of some sort (e.g., social worker, clinical or counseling psychologist, etc.)? (yes, no).

_____ If yes, how many months have you performed as a counselor or trainer?

_____6. Performance Appraisal Experience: Have you ever conducted a performance problem correction meeting with a person under your direct supervision? (yes, no)

_____ If yes, approximately how many of these performance problem correcting sessions have you conducted?

Candidate Name:_____ Sequence Number: _____ Session Code:_____

INTERPERSONAL RELATIONS

A. SELF EXPRESSION: disclosing or communicating one's point of view by expressing experiences, behaviors and feelings

_____1. Information Content

_____2. Emotional Content

_____3. Delivery

_____4. Overall Self Expression

B. RESPONDING WITH EMPATHY: communicating accurate understanding of the core meaning of another person's word.

_____1. Information Content

_____2. Emotional Content

_____3. Delivery

_____4. Overall Responding with Empathy

C. PROBING AND QUESTIONING: gathering information.

_____1. Information Content

_____2. Emotional Content

_____3. Delivery

_____4. Overall Probing and Questioning:

D. Challenging: providing feedback to a person on past performance and seeking to motivate them to improve future performance.

_____1. Information Content

_____2. Emotional Content

_____3. Delivery

_____4. Overall Challenging

E._____Overall Interpersonal Relations

F._____ How confident are you of this rating? (1=not 4=moderately 7=highly confident)

Problem Solving

A. Problem Identification: understanding normal states, noticing deviations, and listing and analyzing details of deviations.

_____1.Monitoring the Environment.

_____2.Defining the Problem.

_____3.Overall Problem Identification

B. Problem Analysis: Specify the decision objectives and diagnosing the causes of the problem.

_____1.Specify decision objectives.

_____2.Diagnosing the causes.

_____3. Overall Problem Analysis.

C. Decision Making: Developing potential solutions, establishing criteria for appraisal and choosing a solution.

_____1.Developing potential solutions.

_____2.Establish a methodology/criteria.

_____3.Appraising possible solutions.

_____4.Choosing the best solution.

_____5.Overall Decision Making.

D. Action: Implementing the solution.

_____1.Implementing the solution.

E. Supervision: Ensuring decisions are carried out according to plan.

_____1.Ensuring decisions are carried out.

F. Problem identification: Matching for deviations from expectations.

_____1.Monitoring the decision environment.

G._____Overall Problem Solving

H._____ How confident are you of this rating? (1=not 4=moderately 7=highly confident)

Overall Assessment

_____1. Overall Assessment rating (1=low, 4=minimum passing, 7=high)

_____2. How successful would this candidate be as an entry level supervisor? (1=definitely not 4=moderate 7=high success)

_____3. How confident are you of this rating? (1=not 4=moderately 7=highly confident)

How much influence did the following potential causes/factors have on this candidate's performance in this exercise?

(1=extremely little 4=undecided 7=extremely big)

_____4.1. Candidate's hard work in the test

_____4.2. Candidate's Skill in Interpersonal Relations

_____4.3. Candidate's Skill in Prob. Solving

_____4.4. Candidate's Good (or bad) mood

_____4.5. Candidate's (good or bad) luck

_____4.6. The typical assistance (or problems) provided by Chester

_____4.7. The unusual assistance (or problems) provided by Chester

_____4.8. Ease (or difficulty) of the exercise

_____4.9. Poor acting by role players on tape

_____4.10.Nonverbals inconsistent with words

_____4.11.Other reasons (please list on back)

_____5. Could you see and hear everything on this videotape? (yes, no)

If not, what was the problem and how did it affect your ratings?

_____

_____

Follow-up Questions

_____1. Did you fill out the rating scales in the order specified in the directions, that is, only making each "overall" rating after making the related subcomponent ratings? (yes, no)

_____2. Did you know any of the actor's or actresses in the videotapes? (yes, no)

If yes, what effect did this have on your ratings? Please explain.

_____3. Did you notice any pattern develop in the ratings you made during the experiment? (yes, no)

If yes, what was the pattern and what effect did it have on your ratings?

_____4. Did you understand the problem solving dimension and it's various subcomponents and rating scales?
(1=definitely not, 4=undecided, 7=definitely yes)

If you had difficulties, which subcomponents did you find problematic?

_____5. Did the problem solving dimension definition, subcomponents and ratings scales agree with your personal understanding of this category of performance?

(1=definitely not, 4=undecided, 7=definitely yes)

If the definition was different, how was it different?

_____6. Did you understand the Interpersonal Relations dimension and it's various subcomponents and rating scales?
(1=definitely not, 4=undecided, 7=definitely yes)

If you had difficulties, which subcomponents did you find problematic?

_____6. Did the interpersonal relations dimension definition, subcomponents and ratings scales agree with your personal understanding of this category of performance?

(1=definitely not, 4=undecided, 7=definitely yes)

If the definition was different, how was it different?

APPROVAL SHEET

The dissertation submitted by Robert L. Holmgren has been read and approved by the following committee:

Dr. John D. Edwards, Director
Associate Professor of Psychology, Loyola

Dr. R. Scott Tindale
Assistant Professor of Psychology, Loyola

Dr. Fred Bryant
Associate Professor of Psychology, Loyola

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

November 27, 1989
Date

John D. Edwards
Director's Signature