1978

The Effects of Halo Reduction Training on Individuals Varying in Cognitive Complexity

Kanchan S. Dandage
University of Central Florida

THE EFFECTS OF HALO REDUCTION TRAINING ON INDIVIDUALS VARYING IN COGNITIVE COMPLEXITY

BY

KANCHAN S. DANDAGE

B.S., University of Wisconsin - Milwaukee, 1975

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science: Industrial Psychology
in the Graduate Studies Program of the College of
Social Sciences
of Florida Technological University

Orlando, Florida

1978

Abstract

The effects of training in halo reduction on cognitively complex and noncomplex individuals were studied. Three main hypotheses were tested: 1) There would be a significant negative relationship between cognitive complexity and halo. 2) Training would significantly reduce halo. 3) Training would significantly reduce the amount of halo in the ratings of cognitively noncomplex individuals, but not in the ratings of complex individuals. Forty undergraduate students were given a cognitive complexity test, and high and low complexity groups were identified. Subjects were randomly assigned to either the Training or the No-Training condition. The training groups participated in a lecture-discussion session aimed at rating improvement. The No-Training groups worked on a Sentence-Completion exercise. All subjects viewed a videotaped discussion and rated two discussion participants. The relationship between halo and complexity was not significant for either of the ratees; however, for one ratee, the results were in the expected negative direction (the product moment correlation coefficient between the rating variance and complexity was .3987, $.05 < p < .09$). Training significantly reduced halo for both of the ratees. No significant interaction effect between training and complexity was found. Possible explanations for not receiving support for two of the hypotheses were suggested.

Table of Contents

List of Tables

List of Figures

Introduction

The judgmental process has been studied for many years. One aspect of this process that has been noted throughout the literature is the halo effect. The halo effect was first named by Thorndike (1920). It has been defined as the tendency of a rater to allow a general impression to greatly influence the evaluation of specific traits (Stalnaker & Remmers, 1928). This phenomenon is said to occur when a particular rater tends to rate a particular ratee similarly on all traits (Guilford, 1959).

The halo effect has been statistically defined in four different ways throughout the literature: 1) as an inflated intercorrelation among traits (Gilinsky, 1947), 2) as a general bias factor derived through matrix and factor analysis (Grant, 1955), 3) as the rater-ratee interaction, as stated in Guilford's analysis of variance model (Guilford, 1954), and 4) as a variance score, where halo is inversely proportional to the variance of ratings given by a rater to a ratee across several different traits (Brown, 1968).

This effect was shown to be a phenomenon of judgment, rather than the effect of objective correlation of traits, by Johnson and Vidulich (1956). These investigators compared the variances of ratings made under two conditions. In one condition, subjects rated all of the individuals on one trait per day. In the second condition, subjects rated one individual per day on all traits. Significantly less halo was found in the first condition -- thus supporting the hypothesis that the effect resides in the judging process rather than in objective reality. The term objective reality refers to specific job behaviors. Further, Johnson and Vidulich conceived of halo as an error in the judgmental process.

Indeed much of the literature on halo has treated this phenomenon as an error which should be minimized. Many studies have been devoted to the task of identifying the conditions which would reduce this effect. Several suggestions have been mentioned in the literature (Allport, 1947; Brown, 1968)

1) Rate one trait at a time across all ratees, rather than rating all traits at once for a particular individual.
2) Use raters who are very familiar with the ratee's performance.
3) Carefully define the variables to be rated.
4) Use behaviorally anchored ratings.
5) Train raters in the pitfalls of rating errors.
6) Vary the presentation of the high/low ends of the scale.
7) Use more than one rater.

Some of these suggestions and studies utilizing them are presented below.

One means of halo reduction is suggested by Symonds (1925). According to this method, one trait is rated at a time across all ratees, rather than rating one individual on all traits at once. Several studies have dealt with this topic.

As noted earlier, Johnson and Vidulich (1956) found that the halo effect was significantly reduced when all individuals were rated on one trait at a time than when they were rated on all traits at once; however, several recent studies have failed to confirm these findings.

Blumberg, DeSoto and Kuethe (1966) investigated the halo effect under two conditions. In the "Name" condition, subjects rated each name on seven traits before proceeding to the next trait. No differences between the two conditions in halo was found. Other investigations by Johnson (1963), Taylor and Hastman (1956), Brown (1968) and Fontaine (1977) have reported similar results. That is, the method of rating all individuals on one trait at a time yielded approximately

the same degree of halo as the method of rating each individual on all traits at a time.

A second method of reducing halo mentioned in the literature deals with the rater-ratee interaction. Increased knowledge of the ratee by the rater should decrease the amount of halo in the ratings. Bittner (1948) suggests that the person or persons in closest contact with the ratee's work should do the rating. Generally, studies support the hypothesis that increased acquaintance between the rater and the ratee is related to decreased bias in judgment (Bare, 1954; Koltuv, 1962). Brown (1968) studied the ratings made by students of their peers. The rating form consisted of six bipolar traits on 10-point scales. It was found that the ratings in which the ratee was very well known by the rater exhibited significantly less halo than the ratings in which the ratee was little known by the rater.

A third technique of halo reduction involves the careful definition of all of the variables to be rated (Allport, 1947). Taylor and Manson (1951) used this and other suggestions in a study of rating formats. Their scales were related to the job--not to personality. Scale divisions were carefully defined. Each rater worked under the supervision of investigators using the format suggested by Symonds (1925). None of the usual rater errors--including halo and leniency--were found in the ratings.

Another method of reducing halo suggested in the literature is the use of behaviorally anchored rating scales. Barrett, Taylor, Parker, and Martens (1958) used a format involving complete verbal descriptions of the behavior appropriate to each scale division. This format was found to be superior in the reduction of halo. Buckalow (1960) found that when items relate to observable behaviors, ratings improve--even though halo is not eliminated completely. Several authors have investigated the effects of the behaviorally anchored rating scales developed by Smith and

Kendall (1963) on halo reduction. Hakel (1971) and Landy and Guion (1970) have both found the Smith-Kendall scales to be susceptible to the halo error. Campbell, Dunnette, Arvey and Hellervik (1973) found that behaviorally anchored scales yield less halo error than do Likert-type summated rating scales.

The effects of these behaviorally anchored rating scales were systematically investigated by Burnaska and Hollman (1974). These authors studied the effects of rater response bias on three rating scale formats--the Smith-Kendall type behaviorally anchored scales for derived performance dimensions, scales for the same dimensions but without the behavioral anchors, and scales for dimensions selected on an a-priori basis. The behavioral expectation scales were not found to be superior to the other rating formats in the reduction of halo. Another study (Keaveny & McGann, 1975) compared ratings derived from behavioral expectation scales with ratings on graphic rating scales. In this investigation college students were asked to evaluate their professors. The behaviorally anchored scales resulted in less halo error in ratings of different performance dimensions.

In regard to the use of behaviorally anchored rating scales, as well as other techniques to reduce halo, some researchers (Zedeck, Kafry, & Jacobs, 1976) feel that the time and effort spent on these studies is not called for. It should just be assumed that some raters and formats are more prone to such biases. Given this assumption, then, the emphasis should be on training of all evaluators to be as perceptive and as objective as possible. These authors, then, recommend another strategy in halo reduction--that of training the raters.

Many authors have recommended this strategy. Bittner in 1948 reported on studies demonstrating the need for appraiser training in organizations. In one study, it was found that

84% of army officers felt that more and better training would improve ratings. In another study reported by Bittner, it was found that 95% of the personnel directors in the Owens-Illinois Glass Company indicated that better training of raters would lead to greatly improved ratings.

A current survey of 293 U.S. corporations (Lazer & Wikstrom, 1977) shows the actual incidence of training in managerial performance appraisal. In 43% of the companies, all of the appraisers of lower level managerial performance are trained. At the middle management level, 45% of the companies train all of the appraisers. Finally, 50% of the companies train all appraisers of upper level managerial performance.

Many researchers have been concerned with what constitutes an effective training program in halo reduction. One study (Follman, Wiley, Geiger, & Lavely, 1974) investigated the effect of mere instructions on halo reductions. It was found that mere instructions to consider each item independently of all other items did not significantly reduce the halo effect. Brown (1968) conducted a survey of the literature pertaining to the training of raters. It was found that a training effort is beneficial to the rater when it includes practice with the specific rating scale (Wakeley, 1961), a discussion of rating errors by raters (Levine & Butler, 1952) and an emphasis on the importance of the independence of traits (Taylor & Hastman, 1956).

All of the above three elements were incorporated in the training program used in Brown's study. The effect of training was a significant reduction in halo. All three strategies suggested by Brown were also included in a study by Bernadin and Walter (1977). These authors utilized behavior expectation scales, developed according to the procedures recommended by Bernadin, LaShells, Smith, and Alvarez (1976), rather than trait rating scales. Training

resulted in a significant reduction of halo. Another study (Latham, Wexley, & Pursell, 1975) verified the second and third conclusions of Brown. That is, the training program included a discussion of rating errors by raters and an emphasis on the importance of trait independence. It was found that training significantly reduced halo.

As can be seen, literature on halo has primarily treated this phenomenon as an error that should be minimized or eliminated. Some theorists feel that these response tendencies reflect genuine systematic differences in personality which should be maximized or understood (Wright & Richardson, 1977). An adequate study of interpersonal perception requires a systematic characterization of the rater (Crockett, 1965).

The variable cognitive complexity deals with the characteristics of the perceiver. According to Bieri (1955), the degree of differentiation in construct systems used by people in describing others reflects the cognitive complexity-simplicity of the system. The term cognitively complex refers to a system of constructs which differentiates highly among persons. A construct system which provides poor differentiation among people is said to be cognitively simple.

Research in cognitive complexity has identified many correlates of this variable in the area of interpersonal judgment (Crockett, 1965). Some of these correlates are predictive accuracy and differences in prediction about others, degree of awareness of both positive and negative attributes about others, the level of organization of impressions and the type of task to be judged. The studies dealing with these and other correlates are discussed below.

Several studies have attempted to investigate the relationship between cognitive complexity and predictive accuracy. Bieri (1955) asked subjects to predict the responses to a questionnaire of two classmates who were known

to them. A low, positive statistically significant relationship between the two variables was found. Upon further examining the results, Bieri concluded that the relationship resulted from a superiority among complex subjects in predicting when the other person differed from the subject. Thus, persons in low complexity showed a greater expectation of similarity in attitudes of self and others. Another investigation (Leventhal, 1957) failed to find a statistically significant relationship between complexity and accuracy of prediction; however, Bieri's finding that subjects low in complexity predicted significantly greater similarity between themselves and others than did the subjects high in complexity was confirmed. A recent study (LeCann, 1969) also investigated the relationship between cognitive complexity and accuracy of person perception. The results were consistent with those found by Leventhal. That is, no statistically significant relationship between these two variables was found.

Another study dealing with accuracy investigated the effects of feedback on the accuracy with which clinical judges of different cognitive styles predicted behavior (Gibbs, 1973). It was found that cognitively complex subjects did not improve in accuracy over trials; however, low complexity subjects improved in the accurate feedback and no feedback conditions--but not in the inaccurate feedback condition.

In addition to accuracy, another correlate of complexity is the degree of awareness of both the positive and the negative attributes in others. Because complex people have more dimensions available to them, they should be less likely than noncomplex persons to divide mankind into two groups on the basis of a good-bad dichotomy. This hypothesis has been supported by several studies reported by Crockett (1965).

A recent study (Halverson, 1970) dealt with the above topic of awareness of both the positive and the negative

traits. Specifically, the topic of trait consistency was studied. Results showed that for equally desirable traits, low complexity subjects saw them as going together more often than did high complexity subjects. When the traits were of unequal desirability, low complexity subjects saw them as going together less often than did high complexity subjects. A related study (Menasco, 1976) dealt with the degree of conflict experienced in decision making by persons varying in cognitive complexity. It was found that cognitively complex people are more likely to rate decisions as more difficult and with greater feelings of discomfort than are less cognitively complex persons.

Another correlate of cognitive complexity is the level of organization of impressions. Several studies have been concerned with this topic. Bliese (1974) found that high complexity subjects differentiated more and organized their impressions more than low complexity subjects. Fertig and Mayo (1969) studied the relationships between organization, trait consistency and cognitive complexity. Trait consistency was defined as the ratio of positive to negative traits. It was found that cognitively complex subjects wrote more organized impressions than cognitively simple subjects in the moderate, as compared to the high and low, consistency conditions; however, the interaction effect was not significant.

In addition to the level of organization of impressions, the type of task to be judged is another variable related to complexity. Shepherd (1972) showed that on evaluative traits, the perception of negative figures was more complex than the perception of positive figures. No such differences occurred on non-evaluative traits. Crockett, Mahood and Press (1975) measured differentiation and the level of organization of impressions of complex and noncomplex subjects. It was found that the complex subjects' scores were higher in the understanding set than in the evaluation set; however, for

noncomplex subjects, set did not significantly affect performance.

Thus far, the variables dealing with the area of impression formation as they relate to cognitive complexity were studied. Cognitive complexity is also correlated with variables relating to the quantification of these impressions into the form of a rating. People vary with respect to the particular way they respond to these ratings--thus demonstrating different response styles.

One such response style is the tendency to make extreme judgments. Nidorf and Argabrite (1970) found that cognitively complex subjects made more extreme judgments than cognitively simple subjects. Results were explained in terms of the number of categories available to persons varying in complexity. If an individual is questioned about a particular aspect of his environment, the more he can differentiate the aspect, the more confident will he be in his decision and reflect this confidence in extreme judgments. Deaux and Farris (1975) studied the relationship between complexity, extremity and the sex of the subject. Findings showed that males and females differ with respect to their use of extreme scores--but not in complexity. Wright and Richardson (1977) also found that males and females do not differ in complexity.

In addition to extremity of judgment, the response style of variability as it relates to cognitive complexity has recently been studied. Wright and Richardson (1977) theorized that there would be a significant relationship between these two variables, since one of the tests of cognitive complexity--the Role Construct Repertory Test (Rep Test)--uses variability to measure cognitive complexity (Kelly, 1955). These authors correlated cognitive complexity and within subject variance across 35 Likert-type scale items of a course evaluation instrument. Cognitive complexity was

measured by a modification of Kelly's Rep Test (Bieri, Atkins, Briar, Leaman, Miller & Tripodi, 1966). A significant relationship between cognitive complexity and within subject variance on the course evaluation form was found. The complex subjects showed more variability in responses than the noncomplex subjects. It was concluded that cognitive complexity represents a response style.

As was mentioned previously, one method of measuring halo is calculating the within-subject variance across different traits, and halo is inversely proportional to the within-subject variance. Thus, according to the above study, as cognitive complexity increases, the halo effect in a person's ratings decreases.

This would imply that the amount of halo in the ratings of the complex individuals may be quite small. As the matter of fact, it is possible that the ratings of these individuals may be demonstrating a ceiling effect in response variability. If this is the case, training in the reduction of halo would not be very useful to these people. As was mentioned previously, the percentage of companies in which all appraisers of managerial performance are trained is quite high (Lazer & Wikstrom, 1977). Thus if the individuals for whom training is not beneficial can be identified, a cost saving would result.

The purpose of the present study, then, was twofold: 1) to replicate the findings of Wright and Richardson of a significant negative relationship between halo and cognitive complexity and 2) to study the effects of training in halo reduction on cognitively complex and noncomplex individuals. The following major hypotheses were tested: 1) There would be a negative and significant relationship between halo and cognitive complexity. That is, the average variance score of the ratings of the cognitively complex individuals would be significantly greater than that of the cognitively noncom-

plex individuals. 2) Training in halo reduction would significantly reduce halo. That is, the average variance score of the ratings of the trained groups would be significantly greater than the average variance score of the ratings of the untrained groups. 3) Training in halo reduction would have a significant effect on the amount of halo present in the ratings of cognitively noncomplex individuals, but not for the cognitively complex individuals. That is, there would be a significant difference between the average variance score of the ratings of the trained and the untrained noncomplex subjects; however, no such significant difference between the trained and the untrained groups would be expected for the complex subjects.

In addition to these major hypotheses, one minor hypothesis was tested: Training would significantly increase rating accuracy. That is, the mean difference between the ratings of the untrained groups and those of a group of graduate students (i.e., an index of accuracy) would be significantly greater than the mean difference between the ratings of the trained groups and those of the graduate students.

Method

Subjects

Two groups of individuals participated in this study. Group I consisted of 40 undergraduate students enrolled in two psychology classes at Florida Technological University. From a total of 49 individuals that were originally given the modified Rep Test, the highest 40% and the lowest 40% of the scorers served as the subjects of this study. There were 14 males and 26 females in Group I. This group was composed of 2 freshmen, 6 sophomores, 26 juniors and 6 seniors. The individuals in Group I were the actual subjects of this study; thus whenever the term 'subjects' is used in this study, it refers to the persons in this group.

Group II consisted of 8 Industrial Psychology graduate students also enrolled at Florida Technological University. This group consisted of 4 males and 4 females. The ratings of the graduate students were used as indices of accuracy, which could be compared with the ratings given by the undergraduate students.

Apparatus

Seven instruments were used in this study: 1) a test of cognitive complexity, 2) a videotaped Leaderless Group Discussion, 3) a rating form assessing managerial skills, used to rate the Leaderless Group Discussion, 4) a taped lecture on training in rating improvement, 5) a brief summary of the training tape, 6) a sample rating form used in training, and 7) a Sentence-Completion exercise.

The test of cognitive complexity that was used in this study was the modified version of Kelly's Rep Test, mentioned previously. This instrument is composed of a 10 X 10 matrix. Ten roles (e.g., Mother, Person You Dislike, Friend of Same

Sex, Boss, etc.) are written across the top of the matrix, and ten bipolar dimensions (e.g., Outgoing-Shy, Cheerful-Ill Humored, Considerate-Inconsiderate, Calm-Excitable) are listed down the right hand side of the matrix. The directions given along with this test consisted of asking the subjects to write the name or initials of each of ten individuals who best correspond to the ten role types and then to rate each role person on each of the ten dimensions. This instrument is scored by comparing every score with every other score within a role person, across the ten dimensions. Whenever the same rating is given to the same individual on two constuct-dimensions, a point is scored. The maximum score that could be received on this test is 450. High scores indicate that a person is low on the construct cognitive complexity. On the other hand, low scores indicate that a person is high on this construct.

The videotaped presentation consisted of six former graduate students, taking part in a standard Leaderless Group Discussion exercise. Briefly, the exercise involved asking the participants to role play managers of a large nationally based organization, which is considering building a new plant. The participants were asked to reach a decision as to where this new plant should be located during the discussion. Each participant was asked to defend one particular location. The participants were given one hour to read the information pertaining to the main topic of the exercise, prior to the discussion. The subjects of this study watched the first 35 minutes of this videotaped discussion.

The rating form used to rate the Leaderless Group Discussion was very similar to the type of forms used in organizations today. Examples of typical scales are given in Lazer and Wikstrom (1977). The rating form consisted of 10 seven-point graphic rating scales (see Appendix A). Each scale had seven markings on it, and the numerals 1 and 7 were

written under the first and seventh marks, respectively. A term, identifying the rating characteristic being measured, was written below the scale, and a behavioral definition of the characteristic followed. The dimensions measured by the scales were relevant to the main topic of the Leaderless Group Discussion and assessed typical managerial skills. The dimensions that were assessed were leadership, oral communication, innovativeness, interpersonal skills, initiative, flexibility, persuasiveness, judgment, planning and organization, and decisiveness.

The taped lecture on training was a 12 minute presentation on ways to improve ratings. It consisted of definitions of and examples of six types of common rating errors: halo, leniency, contrast and similarity, central tendency, proximity and logical error. The importance of trait independence was emphasized. In addition, several important points regarding ways to improve ratings were suggested.

A one-page summary of the contents of the training tape was given to each subject in the Training group (see Appendix B).

The sample rating form used in training consisted of 10 seven-point graphic rating scales (see Appendix C). The format of these scales was the same as that of the rating form used to rate the Leaderless Group Discussion; however, different dimensions were measured. The dimensions contained in the form were relevant to the taped lecture on training and assessed speaking skills. The rating dimensions used were informative, articulate, speaking skills, creative, unified, interesting, fluent, brief, grammatically accurate and conceptually comprehensible.

The Sentence-Completion exercise consisted of 52 statements that the subjects were asked to complete (see Appendix D). Two examples of the type of statements contained in the exercise are "I feel depressed when _____," and "When I feel pressured, I _____."

Procedure

This study was conducted in two psychology classrooms at Florida Technological University. The author served as the experimenter.

Group I: All subjects were first given the modified Rep Test. High and low complexity groups were identified. The high complexity group consisted of individuals receiving the lowest 40% scores on the Rep Test; people receiving the highest 40% scores formed the low complexity group. Half of the participants in each of these groups were randomly assigned to one of two experimental conditions--Training and No-Training. Thus, four groups resulted: 1) Training-Complex, 2) Training-Noncomplex, 3) No-Training-Complex, and 4) No-Training-Noncomplex.

Four days after the administration of the cognitive complexity test, each subject participated in either the Training or the No-Training condition.

The Training condition consisted of a lecture-discussion session conducted by the experimenter. The training included all three of the suggestions put forth by Brown (1968). The participants first listened to the taped lecture on training. Afterward, they were questioned about the material covered in the lecture. They were also given an opportunity to discuss the lecture, as well as to ask questions. In addition, they were asked to rate the taped lecture on the sample training scale, for practice purposes. The entire training session lasted 40 minutes. A brief summary of the material covered in the training tape was given to each subject, and the subjects were asked to review this material before the next experimental session.

The groups in the No-Training condition took the Sentence-Completion exercise. This exercise was only used as a fill-in task. The subjects were allowed 40 minutes to work on this task.

Three days after the Training or No-Training sessions, the experimental session was conducted. All subjects were given two copies of the rating form assessing managerial skills. On one copy, the following instructions were written: "Rate participant #3 in the presentation on each of the following dimensions." The instructions written on the other copy were "Rate participant #6 on each of the following dimensions."

After passing out the rating forms, the subjects were given the following instructions:

> You will watch a videotape of six people discussing a certain topic. The participants are six former graduate students, who are playing the role of managers of a large company, which is considering building a new plant. The group was asked to decide where the new plant should be located. Each participant was previously asked to defend a particular location. After watching the discussion, you will be rating participants number 3 and 6 on the rating forms that are in front of you. Pay attention to numbers 3 and 6, but do not ignore what is going on in the discussion, because that will help you in rating these two individuals.

Both participants number 3 and 6 were males. Participant number 3 (ratee #3) was expected to induce negative halo. He spoke very little during the discussion and thus was expected to be rated low on most of the managerial skills assessed by the rating form. Participant number 6 (ratee #6) was expected to induce positive halo. He was an active participant in the discussion and influenced the group on several occasions. Thus he was expected to receive generally high ratings on most of the managerial skills listed in the rating form.

All subjects viewed the videotaped discussion and rated participants number 3 and 6 on the rating form assessing managerial skills.

Group II: The modified Rep Test was also given to Group II, the graduate students. One week after the administration of this test, these students listened to the taped training lecture. They were also given an opportunity to discuss the lecture, as well as to ask questions. Afterward, they observed the videotaped Leaderless Group Discussion. This group was given the same instructions concerning the videotape as those given to Group I. After observing the discussion, Group II rated participants number 3 and 6 on the rating form assessing managerial skills.

Results

Scores on the modified Rep Test were calculated for each subject. As was mentioned previously, the highest 40% scorers formed the Noncomplex group and the lowest 40% scorers formed the Complex group. The scores of the Complex group ranged from 96 to 131. The scores of the Noncomplex group ranged from 154 to 450.

Variance between the 10 ratings of each subject were computed for ratees #3 and #6. Table 1 gives the means (of the variances of the ratings) for each of the four groups--Training-Complex, Training-Noncomplex, No-Training Complex, and No-Training-Noncomplex for ratees #3 and #6.

These data were analyzed by 2 X 2 analyses of variance. Table 2 presents the results of these analysis for ratees #3 and #6.

For ratee #3, the effect of training was found to be significant ($F_{(1,36)} = 19.7664$, $p < .01$). The variance scores of the Training groups were significantly greater than the variance scores of the No-Training groups. The effect of complexity was not found to be significant ($F_{(1,36)} = .00263$, $p > .05$). Also, no significant interaction effect was found ($F_{(1,36)} = .04858$, $p > .05$). No apparent difference between the means of the Complex (10.33) and the Noncomplex (10.17) groups can be seen. Although the F ratio of the training X complexity interaction was not significant, the group means showed a trend in the expected direction. The difference between the means of the Training-Noncomplex and the No-Training-Noncomplex groups was 14.09, and the difference between the means of the Training-Complex and the No-Training Complex groups was 12.76.

For ratee #6, a significant effect of training was found ($F_{(1,36)} = 25.689$, $p < .01$). The variance scores of the Training groups were significantly greater than the variance scores of the No-Training groups. No significant effects due to complexity ($F_{(1,36)} = .13689$, $p > .05$) or due to the training X complexity interaction ($F_{(1,36)} = .5048$, $p > .05$) were found. Although the F ratio for the effect of complexity was not significant, the group means showed a trend in the expected direction. The mean of the Complex groups was 14.81, and the mean of the Noncomplex groups was 13.68. Similarly, even though the F ratio for the training X complexity interaction was not significant, the group means indicated a trend in the expected direction. The difference between the means of the Training-Noncomplex and the No-Training-Noncomplex groups was 17.65, and the difference between the means of the Training-Complex and the No-Training-Complex groups was 13.31.

In order to further analyze the data pertaining to ratee #6, an additional statistic was calculated. Correlation coefficients (r values) between each subject's score on the modified Rep Test and his variance score on the ten rating dimensions were computed for the Training and the No-Training groups. Neither of these correlation coefficients were found to be significant (No-Training groups: $r = .3987$, $p > .05$; Training groups: $r = .075$, $p > .05$). In addition, a z-test analyzing the difference between these two correlation coefficients was conducted. Although this difference was not found to be significant ($z = 1.01924$, $p > .05$), the results showed a trend in the expected direction. That is, the score on the Rep Test showed a tendency to be related to the variance score among the No-Training groups ($r = .3987$, $.05 < p < .09$), but not among the Training groups ($r = .075$, $p > .05$).

In order to assess the strength of the relationships

between the independent variables, cognitive complexity and training, and the dependent variable, the variance of the ratings, omega squares (w^2) were calculated for both ratees. The results of these computations are presented in Table 3. The highest omega square values for both ratees were obtained for the variable training. Table 3 shows that 33% of the variability in the variance scores was accounted for by the variable training for ratee #3. For ratee #6, 39% of the variability was accounted for by the variable training.

Additional statistics were calculated in order to compare the ratings given by Group I with the indices of accuracy (i.e., the ratings of Group II, the graduate students). The mean for each dimension for ratees #3 and #6 for the Training and No-Training groups, as well as for the graduate students, are presented in Table 4. These data have been graphed and are presented in Figure 1. The variances for all three of these groups for ratees #3 and #6 are presented in Table 5. For ratee #3, there are three dimensions (i.e., #1, #5, #7) on which the variance of the ratings of the graduate students is lower than the variance of the ratings of the other two groups. For ratee #6, there are seven dimensions (i.e., #2, #4, #5, #6, #7, #8, #10) on which the variance of the ratings of the graduate students is the lowest among the three groups. The variances for the three groups for both ratees are plotted in Figure 2.

The differences between the rating means of the Training and No-Training groups and the rating means of the graduate students for each of the ten dimensions were analyzed by means of t-tests. Forty t-tests were conducted. The results of these analyses for ratees #3 and #6 are presented in Table 6. As expected, none of the differences between the rating means of the training groups and the rating means of the graduate students for ratees #3 and #6 were found to be significant. For dimensions #2, #4, #9 for

ratee #3, the rating means of the No-Training groups were significantly different from the rating means of the graduate students. For dimension #7 for ratee #6, the rating mean of the No-Training groups was significantly different from the rating mean of the graduate students. Thus, as expected, the number of dimensions on which the means of the graduate students were significantly different from the means of the No-Training groups was greater than the number of dimensions on which the means of the Training groups were significantly different from the means of the graduate students.

Discussion

Two of the three major hypotheses of this study were unsupported; however, the results were generally in the expected direction. 1) The hypothesis of a negative and significant relationship between cognitive complexity and halo was unsupported for the ratings of both ratees #3 and #6; however, the relationship between the two variables was in the expected negative direction for the ratings of ratee #6. 2) The hypothesis that training would significantly reduce halo was supported for both ratees #3 and #6. 3) The hypothesis that training in halo reduction would significantly reduce the amount of halo present in the ratings of cognitively noncomplex individuals, but not for cognitively complex individuals was not supported for either of the ratees. Again, the results demonstrated the expected trend.

The minor hypothesis that training would improve rating accuracy was generally supported. The ratings given by the trained groups were generally more accurate (i.e. closer to the ratings given by the graduate students) than the ratings given by the untrained groups.

The results of a nonsignificant relationship between halo and cognitive complexity for both ratees #3 and #6 are inconsistent with those obtained by Wright and Richardson (1977), who found a significant negative relationship between these two variables. The result that training significantly reduced halo for ratees #3 and #6 is consistent with the results of several investigations. Studies by Brown (1968), Levine and Butler (1952), Taylor and Hastman (1956), and Latham, Wexley and Pursell (1975) have all found that training significantly reduced halo.

In the case of both ratees, two of the three hypotheses were not supported. First, a significant relationship between halo and cognitive complexity was not found; however,

the relationship between the two variables was in the expected direction for ratee #6 (the correlation coefficient value between the rating variance and complexity was .3987, $.05 < p < .09$). Unlike ratee #6, the results did not show a trend in the expected direction for ratee #3.

This difference in results may have been due to some fundamental differences between the two ratees. Ratee #3 spoke very little during the discussion and thus provided a very small amount of information, which could be used in rating him. It seems that the amount of information provided by this ratee was not sufficient for the complex individuals to be able to make a larger number of distinctions between the rating dimensions; thus, there was little difference in halo between the complex and the noncomplex individuals. On the other hand, ratee #6 spoke a great deal and thus provided a much greater amount of information which could be used to rate him. Because of this, there was a general tendency for the complex individuals to make a greater number of distinctions than the noncomplex individuals. This may be the reason why the relationship between the two variables halo and cognitive complexity was in the expected direction for ratee #6, but not for ratee #3.

Although the relationship between halo and cognitive complexity was in the expected direction for ratee #6, it was not statistically significant. Two differences between the present study and the study by Wright and Richardson (1977) may clarify the inconsistency in findings between the two investigations:

1) The subject sample in the Wright and Richardson study (N=102) was much greater than the sample used in the present study (r based on N= 20).

2) Whereas a multiple correlation coefficient between each of the ten cognitive complexity subscales and the within-subject variance (an index of halo) was calculated by Wright

and Richardson, the present study correlated an overall cognitive complexity score with the within-subject variance. This may have reduced the size of the correlation in the present study.

The second unsupported hypothesis dealt with the interaction between training and complexity. No significant interaction effect between these two variables was found. It seems that the presence of the interaction effect is dependent upon a very high negative relationship between halo and complexity. Since such a relationship was not found in the case of either of the ratees, no significant interaction effect could be found.

Thus the results of this study do not support the notion that cognitively noncomplex individuals tend to commit more halo errors than cognitively complex individuals, or that training would have a greater effect on the noncomplex persons than the complex persons. Because the results were in the expected direction, it seems worthwhile to conduct further investigations of this relationship in the future, making the following changes:

1) A larger sample size should be used.
2) A multiple correlation coefficient should be computed between the ten cognitive complexity subscales and the within-subject variance.
3) Both ratees should be active participants in the discussion, thus providing the raters with a large amount of information which could be used in making the judgments.

Table 1

Means of the rating variances

Group		Condition		Total
		Training	No Training	
	Ratee #3			
Cognitively	Complex	16.71	3.95	20.66
Cognitively	Noncomplex	17.22	3.13	20.35
	TOTAL	33.93	7.08	41.01
	Ratee #6			
Cognitively	Complex	21.47	8.16	29.63
Cognitively	Noncomplex	22.51	4.86	27.37
	TOTAL	43.98	13.02	57

Table 2
Analyses of variance

Ratee	Source	df	MS	F
#3	Training	1	18.023	19.7664**
	Complexity	1	.00024	.00263
	Training X Complexity	1	.0443	.04858
	Error	36	.9118	
#6	Training	1	23.96304	25.689 **
	Complexity	1	.1277	.13689
	Training X Complexity	1	.4709	.5048
	Error	36	3.7313	

* $p < .05$
** $p < .01$

Table 3

Omega squares

Ratee	Source		
	Training	Complexity	Training X Complexity
#3	.3304	.000044	.00081
#6	.3872	.00206	.00761

Table 4

Dimension means

Dimension	Group		
	Training	No-Training	Graduate Students
	Ratee #3		
1	1.9	1.225	1.375
2	2.95	1.75	3
3	1.9	1.225	1.375
4	3.2	1.375	3.125
5	1.75	1.325	1.125
6	3.3	1.7	2.5
7	2.15	1.425	1.875
8	2.3	1.925	2.75
9	3.1	1.525	2.25
10	2.35	1.725	2.625
	Ratee #6		
1	4.95	4.78	4.5
2	6.05	5.43	5.875
3	4.5	4.68	4.5
4	3.7	4.53	4.375
5	5.45	5.03	5.625
6	3.76	3.98	4.125
7	5.45	4.73	6.125
8	4.2	4.23	5.375
9	4.83	4.63	3.75
10	5.9	5.38	5.875

Table 5

Dimension variances

	Group		
Dimension	Training	No-Training	Graduate Students
	Ratee #3		
1	1.9	.262	.234
2	2.95	.663	2.0
3	1.9	.262	.484
4	3.2	.522	3.86
5	1.75	.407	.109
6	3.3	1.385	2.5
7	2.15	.532	.359
8	2.3	1.357	1.4375
9	3.1	.337	1.4375
10	2.35	1.087	1.984
	Ratee #6		
1	1.473	1.36	1.75
2	1.473	1.78	.859
3	3.625	2.31	2.5
4	2.785	2.49	.984
5	2.373	2.76	.734
6	2.939	2.96	2.359
7	2.77	2.94	.359
8	3.04	3.01	.734
9	3.38	3.32	3.687
10	2.015	2.07	.609

Appendix B

GRAPHIC RATING SCALES

Some Common Errors

1. leniency - tendency for a person to be generally easy or hard in rating others.
2. halo - tendency for a rater to allow a general impression to affect his evaluation of specific traits.
3. logical error - tendency for a rater to rate two traits in a particular way just because they are connected in the rater's own mind.
4. contrast or similarity - The contrast error is the tendency for a rater to judge others in a manner opposite from the way he perceives himself.
5. central tendency - tendency for raters to use only the middle part of the rating scale.
6. proximity - tendency for a rater to allow his rating of a trait to influence his rating of the next trait in the rating form.

Directions for Taking Graphic Rating Scales

1. Try not to be too lenient or too strict.
2. Rate each trait independently of all other traits -
   Do not allow an overall impression of a person to affect your evaluation of specific traits.
   Do not allow one trait to influence your evaluation of another trait just because the two are related in your mind.
   Do not allow the order in which the traits appear on the form influence the judgment of a trait.
3. Rate the ratee, not yourself. Also, do not compare the

ratee only with yourself in evaluating a particular trait.

4. Do not be afraid to use the ends of the rating scale, if you feel that is appropriate.

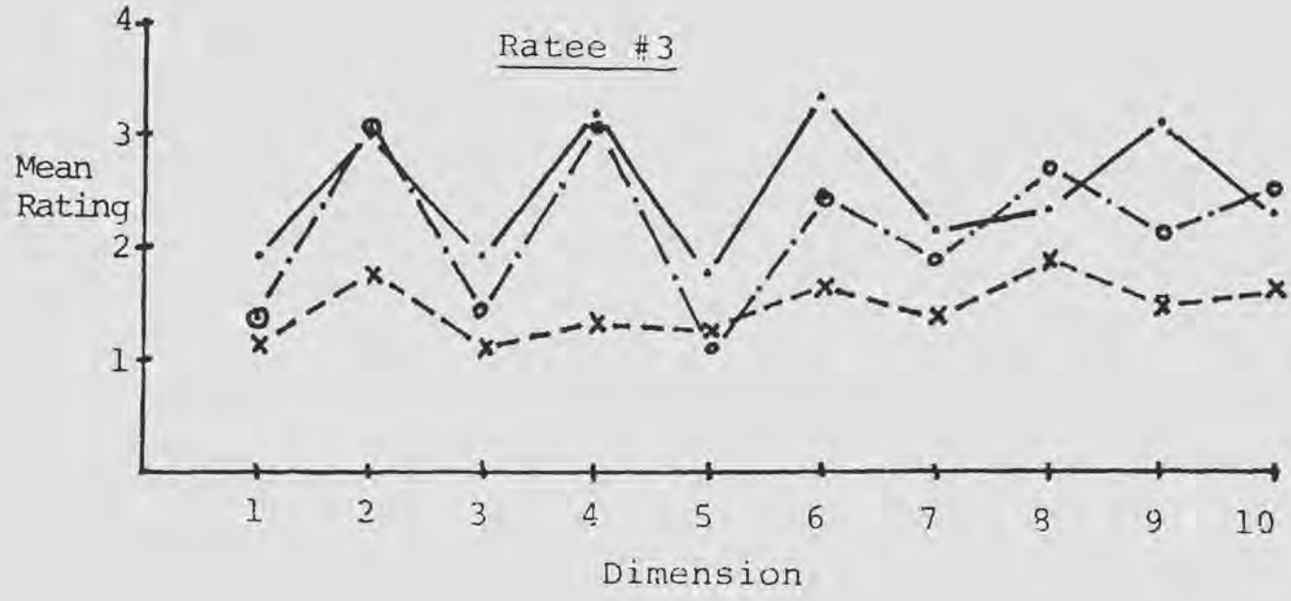
Table 6

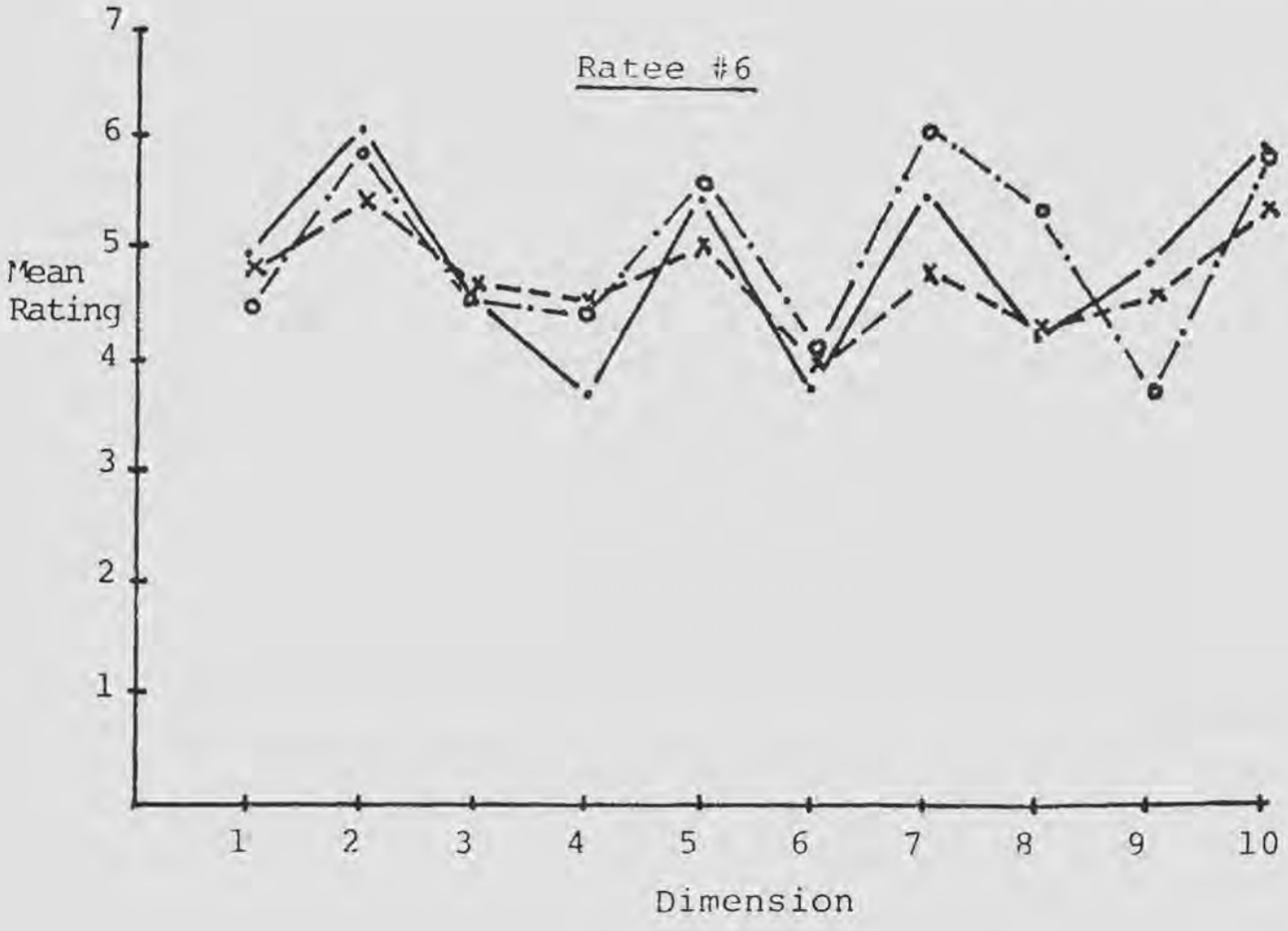
Results of t-tests comparing dimension means

Dimension	T-Tests of Group Comparisons: No-Training/ Graduate Students	Training/ Graduate Students
	Ratee #3	
1	- .6856	- .9317
2	- 2.817 **	.0858
3	- .601	.9418
4	- 3.318 **	.093
5	.812	1.3237
6	- 1.412	1.1062
7	- 1.520	.2794
8	- 1.618	- .6087
9	- 2.069 *	1.2313
10	- 1.789	- .3595
	Ratee #6	
1	.531	.8321
2	- .832	.3236
3	.2697	0
4	.2489	- 1.0319
5	- .928	- .2921
6	- .199	- .5048
7	- 2.167 *	- 1.0773
8	- 1.716	- 1.7553
9	1.131	1.3357
10	- .887	.0453

* $p < .05$

** $p < .01$





•——•——• Training Group

x---x---x No-Training Group

o---o---o Graduate Students

Figure 1: Dimension means

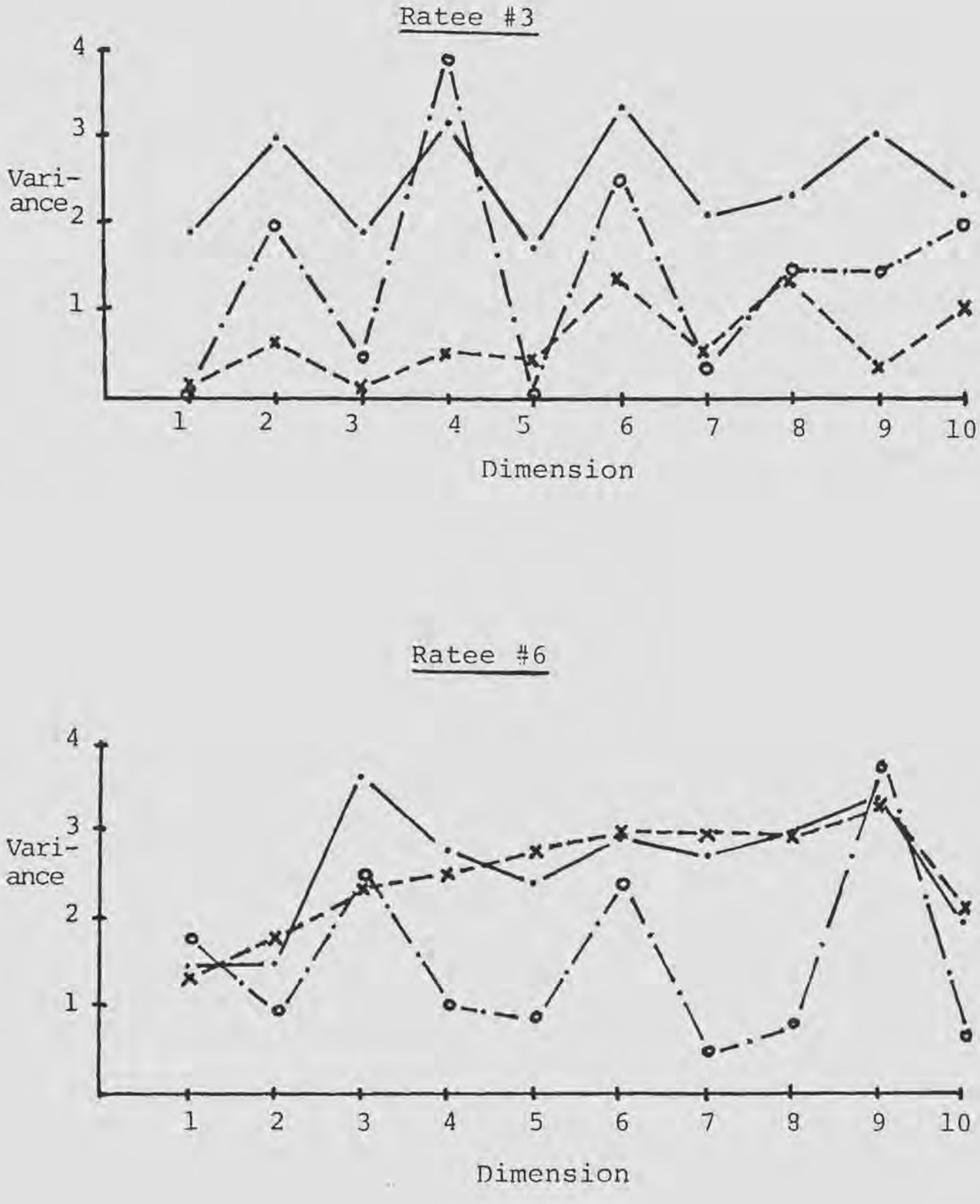
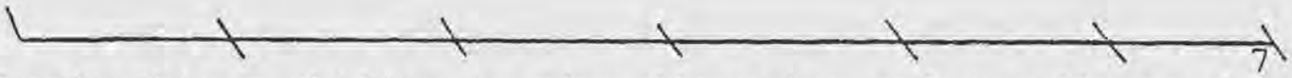


Figure 2: Dimension Variances

Appendix A

Directions: Rate #3 in the presentation on each of the following dimensions.

1. 

Leadership - ability to give direction and coordinate the activities of others in order to accomplish the task.

2. 

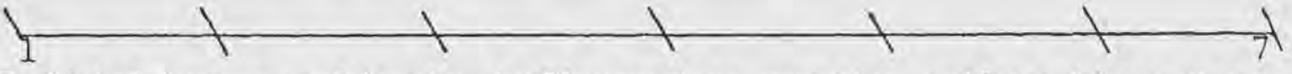
Oral Communication - ability to speak clearly and effectively and get the point across to others.

3. 

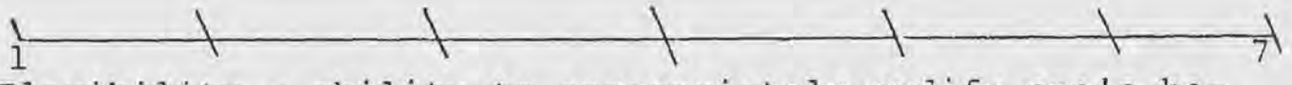
Innovativeness - ability to generate imaginative solutions or ideas.

4. 

Interpersonal Skills - skills in perceiving and reacting sensitively to the needs of others.

5. 

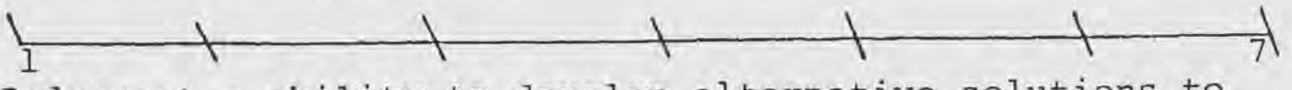
Initiative - actively influencing events rather than passively accepting them. Originating action rather than just responding to events.

6. 

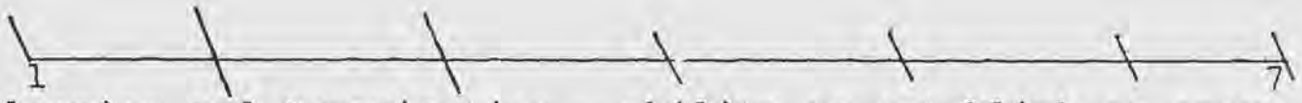
Flexibility - ability to appropriately modify one's behavior when dealing with diversified situations.

7. 

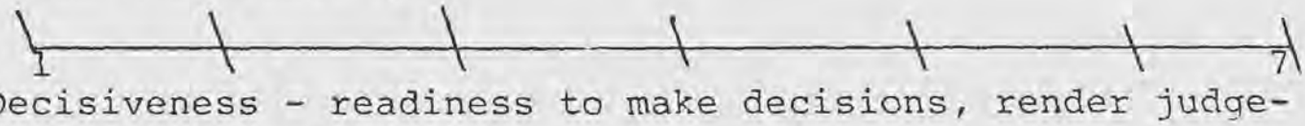
Persuasiveness - ability to organize and present material in a convincing manner.

8. 

Judgement - ability to develop alternative solutions to problems, evaluate courses of action and reach a logical conclusion.

9. 

Planning and Organization - ability to establish an appropriate course of action for self and/or others to accomplish a specific goal.

10. 

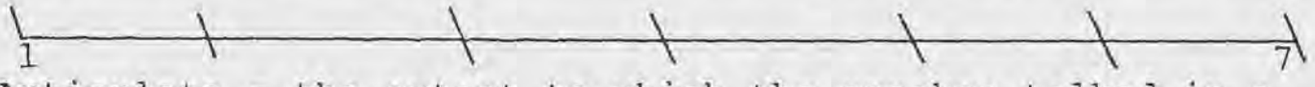
Decisiveness - readiness to make decisions, render judgements, or commit oneself.

Appendix C

Directions: Rate the taped dicsussion on each of the following dimensions.

1. 

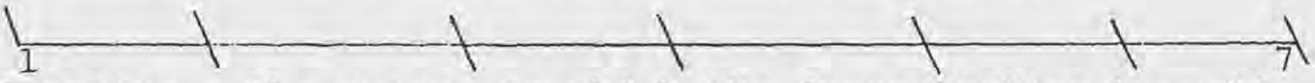
Informative - the extent to which the presentation conveyed facts or ideas that vou were previously unaware of.

2. 

Articulate - the extent to which the speaker talked in a coherent, expressive manner.

3. 

Speaking Skills - the extent to which the speaker used appropriate tone, voice inflection, and speed.

4. 

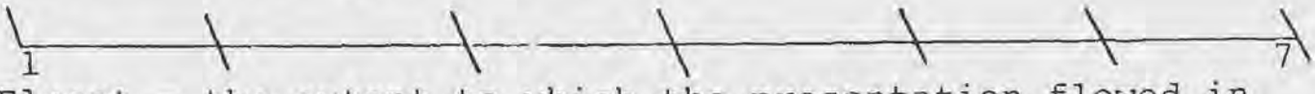
Creative - the extent to which the demonstration brought out original, novel ideas.

5. 

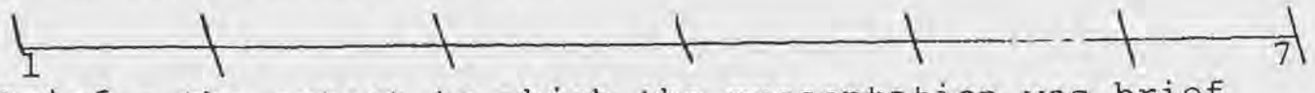
Unified - the extent to which the diversified elements of the demonstration formed a logical whole.

6. 

Interesting - the extent to which the presentation captured and held your attention.

7. 

Fluent - the extent to which the presentation flowed in a smooth, orderly manner.

8. 

Brief - the extent to which the presentation was brief and to the point.

9. 

Grammatically Accurate - the extent to which the words and the sentence structure were grammatically accurate.

10. 

Conceptually Comprehensible - the extent to which the conceptual level of the subject matter was appropriate for the level of the audience.

Appendix D

Complete the following sentences as best as you can.

1. I am at present _____
2. Previously, I was _____
3. In the future, I will be _____
4. My favorite hobby(hobbies) _____
5. My favorite task(tasks) _____
6. My favorite subject at school _____
7. My favorite sport(s) _____
8. I feel happy when _____
9. I feel sad when _____
10. I feel frustrated when _____
11. I feel pressured when _____
12. When I am happy, I _____
13. When I am sad, I _____
14. When I am frustrated, I _____
15. When I feel pressured, I _____
16. When I am in trouble, I _____
17. When I am depressed, I _____
18. My mood is generally _____
19. I have artistic ability in _____
20. The sport(s) that I am the best at _____
    _____
21. The sport(s) that I am the worst at _____
    _____
22. The school subjects that I am the best at _____
    _____
23. The non-school subjects that I am the best at _____
    _____
24. When I do not have anything to do, I _____
    _____

25. My favorite type of movies are _____
26. My favorite type of books are _____
27. I like myself the best when _____
28. I like myself the least when _____
29. When I am angry, I _____
30. My favorite food is _____
31. I have trouble controlling myself when _____

_____

32. My willpower breaks down when _____
33. If I had one wish, it would be _____
34. I often dream about _____
35. I often fantasize about _____
36. I am afraid of _____
37. When I am with people, I _____
38. The kinds of people I like are _____
39. The kinds of people I dislike are _____
40. The kinds of people with whom I am the most comfortable are _____
41. My favorite political hero/heroine is _____
42. My favorite entertainment personality is _____
43. My version of the ideal person is _____

_____

44. In order to meet that ideal, I would need to _____

_____

45. The personality areas in which I need to change the most are _____
46. The skill areas in which I need to improve are _____

_____

47. The attitudes that I need to improve are _____

_____

48. The habits that I need to change are _____

_____

49. I would like to learn more about _____

50. I tend to expect alot from _____

51. I expect very little from _____

52. If I could change one thing to improve the world, it would be _____

References

Allport, G. Personality, a psychological interpretation. New York: Henry Holt and Company, 1947.

Bare, R. Bias as related to rater contacts. Unpublished doctoral dissertation, Ohio State University, 1954.

Barrett, R., Taylor, E., Parker, J., & Martens, L. Rating scale content: I. Scale information and supervisory ratings. Personnel Psychology, 1958, 11, 333-346.

Bernadin, H., LaShells, M., Smith, P., & Alvares, K. Behavioral expectation scales: Effects of developmental procedures and formats. Journal of Applied Psychology, 1976, 61, 75-79.

Bernadin, H., & Walter, C. Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 1977, 62, 64-69.

Bieri, J. Cognitive complexity-simplicity and predictive behavior. Journal of Abnormal and Social Psychology, 1955, 51, 263-268.

Bieri, J., Atkins, A., Briar, S., Leaman, R., Miller, H., & Tripodi, T. Clinical and social judgment: The discrimination of behavioral information. New York: Wiley, 1966.

Bittner, R. Developing an employee merit rating procedure. Personnel Psychology, 1948, 1. 403-432.

Bliese, N. The effects of attitude similarity, cognitive complexity and instructional set on impressions formed of strangers (Doctoral dissertation, University of Kansas, 1973) Dissertation Abstracts International, 1974, 34, 7909A-7910A. (University Microfilms No. 74-12, 529)

Blumberg, H., Desoto, C., & Kuethe, J. Evaluation of rating scale formats. Personnel Psychology, 1966, 19, 243-259.

Brown, E. Influence of training, method, and relationship on the halo effect. Journal of Applied Psychology, 1968, 52, 195-199.

Buckalow, M. Staff reporting: A new look at the halo effect. Personnel Practices Bulletin, 1960, 16, 29-33.

Burnaska, R., & Hollmann, T. An experimental comparison of the relative effects of rater response biases on three rating scale formats. Journal of Applied Psychology, 1974, 59, 307-312.

Campbell, J., Dunnette, M., Arvey, R., & Hellervik, L. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.

Crockett, W. Cognitive complexity and impression formation. In B. A. Maher (Ed.), Progress in experimental personality research (Vol. 2). New York: Academic Press, 1965.

Crockett, W., Mahood, S., & Press, A. Impressions of a speaker as a function of set to understand or to evaluate, of cognitive complexity and of prior attitudes. Journal of Personality, 1975, 43, 379-389.

Deaux, K., & Farris, E. Complexity, extremity, and affect in male and female judgments. Journal of Personality, 1975, 43, 379-389.

Fertig, E., & Mayo, C. Impression formation as a function of trait consistency and cognitive complexity. Proceedings of the 77th Annual Convention of the American Psychological Association, 1969, 4(1), 345-376.

Follman, J., Wiley, R., Geiger, G., & Lavely, C. Effects of halo-effect instructions on level of ratings and intercorrelations. Psychological Reports, 1974, 35, 392.

Fontaine, J. The influence of graphic rating scale format on halo. Unpublished master's thesis, Florida Technological University, 1977.

Gibbs, J. Some effects of feedback on the accuracy and confidence of clinical judgment (Doctoral dissertation, University of Georgia, 1972). Dissertation Abstracts International, 1973, 33, 450B. (University Microfilms No. 73-5698)

Gilinsky, A. The influence of the procedure of judging on the halo effect. American Psychologist, 1947, 2, 309-310.

Grant, D. A factor analysis of managers' ratings. Journal of Applied Psychology, 1955, 39, 283-286.

Guilford, J. Psychometric methods. New York: McGraw-Hill, 1954.

Guilford, J. Personality. New York: McGraw-Hill, 1959.

Hakel, M. Similarity of post-interview trait rating intercorrelations as a contributor to inter-rater agreement in a structured employment interview. Journal of Applied Psychology, 1971, 55, 443-448.

Halverson, C. Interpersonal perception: Cognitive complexity and trait implication. Journal of Consulting and Clinical Psychology, 1970, 34, 86-90.

Johnson, D. Reanalysis of experimental halo effects. Journal of Applied Psychology, 1963, 47, 46-47.

Johnson, D., & Vidulich, R. Experimental manipulation of the halo effect. Journal of Applied Psychology, 1956, 40, 131-134.

Keaveny, T. & McGann, A. A comparison of behavioral expectation scales and graphic rating scales. Journal of Applied Psychology, 1975, 60, 695-703.

Kelly, G. The psychology of personal constructs. New York: Norton, 1955.

Koltuv, B. Some characteristics of intra-judge trait correlations. Psychological Monographs, 1962, 76, (33, Whole No. 552).

Landy, F., & Guion, R. Development of scales for the measurement of work motivation. Organizational Behavior and Human Performance, 1970, 5, 93-103.

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

Lazer, R., & Wikstrom, W. Appraising managerial performance: Current practices and future directions. New York: The Conference Board Inc., 1977.

LeCann, A. Cognitive simplicity-complexity and conceptual systems as correlates of accuracy of person perception (Doctoral dissertation, University of Utah, 1969). Dissertation Abstracts International, 1969, 30, 384B. (University Microfilms No. 69-11, 974)

Leventhal, H. Cognitive processes and interpersonal predictions. Journal of Abnormal Social Psychology, 1957, 55, 176-180.

Levine, J., & Butler, J. Lecture vs. group decision in changing behavior. Journal of Applied Psychology, 1952, 36, 29-33.

Menasco, M. Experienced conflict in decision making as a function of level of cognitive complexity. Psychological Reports, 1976, 39, 923-933.

Nidorf, L., & Argabrite, A. Cognitive complexity and the tendency to make extreme judgments. Perceptual and Motor Skills, 1970, 31, 478.

Shepherd, J. The effects of variations in evaluativeness of traits on the relation between stimulus affect and cognitive complexity. Journal of Social Psychology, 1972, 88, 233-239.

Smith, P., & Kendall, J. Retranslation of expectations: An approach to the construction of unambiguous anchoring of rating scales. Journal of Applied Psychology, 1963, 47, 149-155.

Stalnaker, J., & Remmers, H. Can students discriminate traits associated with success in teaching? Journal of Applied Psychology, 1928, 12, 602-610.

Symonds, P. Notes on rating. Journal of Applied Psychology, 1925, 9, 188-195.

Taylor, E., & Hastman, R. Relation of format and administration to the characteristics of graphic rating scales. Personnel Psychology, 1956, 9, 181-206.

Taylor, E., & Manson, G. Supervised ratings: Making graphic scales work. Personnel, 1951, 27, 504-514.

Thorndike, E. A constant error in psychological ratings. Journal of Applied Psychology, 1920, 4, 25-29.

Wakeley, J. The effects of specific training on accuracy in judging others. Unpublished doctoral dissertation, Michigan State University, 1961.

Wright, R., & Richardson, L. The effect of response style on cognitive complexity and course evaluation. Educational and Psychological Measurement, 1977, 37, 177-183.

Zedeck, S., Kafry, D., & Jacobs, R. Format and scoring variations in behavioral expectation evaluations. Organizational Behavior and Human Performance, 1976, 17, 171-184.