1987


Frank Allen Adair
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/4336

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17" x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6" x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
The effects of ratee job experience, performance variability, and rater cognitive complexity on performance rating accuracy

Adair, Frank Allen, Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1987
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received
16. Other

University
Microfilms
International
THE EFFECTS OF RATEE JOB EXPERIENCE, PERFORMANCE VARIABILITY, AND RATER COGNITIVE COMPLEXITY ON PERFORMANCE RATING ACCURACY

A Dissertation

Submitted to the Graduate Faculty of the

Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by

Frank A. Adair
B.S., Louisiana State University, Shreveport, 1981
M.A., Louisiana State University, 1983
May 1987
Acknowledgments

Several people were helpful in the development and completion of this research. I thank Dr. Dirk D. Steiner for his direction and concern throughout this project as well as his sharing necessary videotaped materials. In addition, I thank Dr. Irving Lane and Dr. Gregory H. Dobbin for their instruction in dealing with the methodology and statistical analyses. Also, Dr. Ray Buss and Dr. Larry Forh were helpful in dealing with methodological and statistical issues.

This project was facilitated by the aid of Jane Mouton in sharing her IBM PC Computer and Caroline Wilhelm in sharing some critical incidents data from previous research. In addition, I thank Kerry Sauley, Al Schnur, and Sharon McConnell for spending their valuable time rating the videotapes. Finally, the encouragement of family and friends provided valuable support for me in my pursuit of the Ph.D. degree in Industrial/Organizational Psychology.
Table of Contents

Page Acknowledgments .................................................... ii

Table of Contents ......................................................... iii

List of Tables ................................................................. v

Abstract ................................................................. vii

I. Introduction ............................................................... 1

Job Experience ............................................................ 3

Performance Variability ............................................... 5

Cognitive Complexity ................................................... 7

Attributions ................................................................. 10

Current Research ........................................................ 13

II. Pilot Study ............................................................ 15

Method ................................................................. 15

Results ................................................................. 15

III. Experiment 1 ........................................................ 16

Method ................................................................. 16

Results ................................................................. 21

IV. Experiment 2 ........................................................ 26

Hypotheses ................................................................. 26

Method ................................................................. 26

Results ................................................................. 28

V. Discussion ............................................................. 32

References ................................................................. 41

Appendix A. Expectations Questionnaire ......................... 45

Appendix B. REP Test .................................................... 47

Appendix C. JDQ1 and JDQ2 ........................................ 49

Appendix D. Vignette and Rating Scales ......................... 51

iii
List of Tables

Table 1. Four Accuracy Components .......... 66
Table 2. Means for Manipulation Check for Experiment 1 .................. 67
Table 3a. Follow-up Experience X Variability ANOVA's on Accuracy Components for Experiment 1 .......... 68
Table 3b. Means for Accuracy Components ..... 69
Table 4. Experience X Variability ANOVA on Accuracy of Overall Ratings for Experiment 1 ................. 70
Table 5. Experience X Variability ANOVA on Average Deviation of Ratings from True Ratings for Experiment 1 ...................... 71
Table 6. Experience X Variability X Cognitive Complexity (REP) Regression on Average Deviation from True Scores for Experiment 1 ...................... 72
Table 7. Experience X Variability ANOVA on Effort and Chance Attributions Together for Experiment 1 .......... 73
Table 8. Means for Manipulation Check for Experiment 2 ...................... 74
Table 9. Experience X Variability ANOVA on Accuracy of Summated Scores for Experiment 2 ...................... 75
Table 10. Experience X Variability ANOVA on Accuracy of Overall Ratings for Experiment 2 ...................... 76
Table 11. Experience X Variability ANOVA on Deviation of Summated Scores from True Scores for Experiment 2 .......... 77
Table 12. Experience X Variability ANOVA on Deviation of Overall Ratings from True Ratings for Experiment 2 .......... 78
Table 13. Experience X Variability X Cognitive Complexity (JDQ1) Regression on Deviation of Overall Ratings from True Ratings for Experiment 2 .......... 79

Table 14. Simple Effects Analysis for High Level of Experience on Deviation of Overall Ratings from True Ratings for Experiment 2 ......................... 80
Abstract

Research indicates that ratee job experience and performance variability can affect performance judgments and causal attributions (Leventhal et al., 1977; Fusilier & Hitt, 1983; Scott & Hamner, 1975). When prototype expectations concerning ratee characteristics are not met, raters reacting to the cognitive inconsistency may distort ratee performance information (DeNisi, Cafferty, & Meglino, 1984). Because cognitively complex raters view behavior in a multidimensional manner (Schneier, 1977), not through a simplified schemata, cognitive complexity's possible effect on rating accuracy under cognitively inconsistent conditions was investigated.

In an initial pilot study, 147 undergraduate volunteers rated the expected performance variability of a professor with either zero, five, or ten years of experience. Experiment 1 utilized a Ratee Experience X Ratee Performance Variability X Rater Cognitive Complexity regression design. One-hundred-thirty-five undergraduate volunteers completed three cognitive complexity measures, six vignettes, six BARS, an attribution questionnaire, and a manipulation check questionnaire. Experiment 2 used a similar design with videotaped rating stimuli (three lectures) to replicate Experiment 1.

The pilot study showed that subjects expect only inexperienced professors to have variable performance overall rating accuracy was lowest in the moderate variability condition and in the high experience condition. Accuracy component analyses (Cronbach, 1955) indicated that more variable performance was rated more accurately. Further, marginal evidence suggested that higher cognitive complexity is somewhat associated with greater differential elevation accuracy when experienced ratees have variable performance. Also, effort and chance attributions
were strongest in the moderate experience condition.

In Experiment 2, highly experienced ratees and those with moderate variability were rated more accurately than others. Cognitive complexity was not helpful to rating accuracy. In one case, it was associated with more underrating of highly experienced, highly variable ratees.

General support was found for the contention that prototype expectations exist and affect rating accuracy. Very limited support was gained for cognitive complexity as a useful rater characteristic. Inconsistent experimental findings are discussed in terms of the salience of performance variability and cognitive processing of information. Implications for training raters to observe multiple performance dimensions and attend to the possible biasing affects of ratee characteristics are also discussed.
The accuracy of performance appraisal ratings is essential to a performance appraisal system's effectiveness. Some researchers have suggested that the accuracy of performance ratings depends upon a complex cognitive process involving attention, categorization, recall, and information integration (Feldman, 1981; Ilgen & Feldman, 1983). Landy and Farr's (1980) explanation of a process model of performance rating indicates that rater and ratee characteristics can influence performance judgments that result from this cognitive process. They also contend that rater and ratee characteristics can have an interactive effect on the psychometric quality of performance ratings. This interactive effect of rater and ratee characteristics upon the accuracy of performance ratings represents a current research need. An overview of how rater and ratee characteristics might impact on rating accuracy will precede discussion of specific characteristics that were the focus of this research.

Salient ratee characteristics may affect the rating process and rating accuracy through the activation of prototypes in raters. Ilgen and Feldman (1983) define prototype as an "abstract analog, or image, summarizing 'central tendencies' or resemblances among category members" (p. 152). Schemata are used in human memory to store and retrieve such complex categorical information. Schemata are "verbal or propositional memory structures representing categories of a more complex nature than prototypes but serving the same function in perceptual organization and memory" (Ilgen & Feldman, 1983, p. 152).

DeNisi, Cafferty, and Meglino (1984) suggest that schemata are
important to the rating process because they contain preconceived notions or expectations about a ratee. Moreover, they suggest that raters may distort additional ratee information to conform to initial expectations. Further, as more information about a ratee's behavior becomes available, raters develop causal attributions concerning whether an event was due to the situation or the person involved. Such attributions would likely affect ratings because situational causes of behavior would suggest to a rater that the ratee was not responsible for his/her performance. Ilgen and Feldman (1983) emphasize that causal attributions in the rating process are influenced by previous automatic attention to salient ratee characteristics and categorization.

Ratee characteristics contribute information beyond the behavioral dimensions present on a performance appraisal form. In order to accurately rate performance, a rater must be able to search beyond initial expectations elicited by salient ratee characteristics. Given the influence of ratee characteristics and the use of schemata in the rating process, it would seem logical that a more complex categorization of ratee information would result in more accurate ratings.

Because multidimensional information processing is required for accurate ratings, cognitive complexity could be an important rater characteristic to consider. Cognitive complexity is the ability to deal with multiple categories in processing information about behavior (Schneier, 1977). To the extent that ratee characteristics influence the use of simplified prototypes or schemata, a cognitively complex rater would be expected to use complex schemata to categorize a ratee. The more complex schemata would enable a rater to broaden the search for relevant performance information without being restricted by proto-
type expectations. Thus, cognitive complexity may result in more accurate ratings in the presence of salient ratee characteristics which influence dependence on simplified prototypes and schemata.

The following pages develop research that will investigate the effects of varied cognitive complexity on rating accuracy when two ratee characteristics are made salient. Job experience and performance variability were chosen as the ratee characteristics because of past research suggesting that each may influence causal attributions and the judgment process (Leventhal et al., 1977; Fusilier & Hitt, 1983; Scott & Hamner, 1975). In addition, when both characteristics are manipulated, unmet expectations regarding the performance variability of an experienced employee may result in cognitive inconsistency. Under such conditions, it is suggested that cognitive complexity will aid in processing relevant performance information and rating accurately.

Ratee job experience and performance variability may prove to be powerful variables affecting cognitive processing of performance information.

Job Experience

One ratee characteristic which may affect rating accuracy is ratee job experience. Early attempts to demonstrate a relation between job tenure and performance were not very promising, however. Jay and Copes (1957) reviewed 47 studies and found an average correlation between tenure and job performance ratings of .17. More recently, others have reported very small positive correlations between these two variables (Bass & Turner, 1973; Cascio & Valenzi, 1977; Zedeck & Baker, 1972). Rothe (1969) and Svetlik et al. (1964) caution that performance ratings used for merit decisions may not correlate highly with tenure because of
some favoritism shown toward newer employees who have not yet received merit raises. Some raters may adjust merit ratings so that those who have not received recent merit raises are rewarded regardless of tenure.

There is some evidence of a relation between job experience and performance ratings in more recent research. Leventhal et al. (1977) manipulated lecture quality on videotapes and perceived job experience of lecturers by information given to the raters. They found that for high performing lecturers, those perceived as having more years of experience (measured as a manipulation check) were given higher performance ratings than those viewed as inexperienced. Experience, then, may play a significant role in raters' perceptions of actual performance.

In another experiment, ratee job experience affected business students' job suitability ratings of hypothetical job applicants, while other ratee characteristics, age, race, and sex did not (Fusilier & Hitt, 1983). This indicates that job experience may be a powerful variable influencing the rating process.

Hypothesis 1

Because ratings tend to be higher for experienced employees, it is hypothesized that experienced ratees will be overrated, inexperienced ratees will be underrated, or overrating of experienced ratees and underrating of inexperienced ratees will occur.
Performance Variability

Very little research has been directed at the effects of ratee performance variability on cognitive processes in performance rating. Specifically, little is known about how rating accuracy is affected by changes in the level of a ratee's performance over a period of time. There is some evidence to suggest, however, that negative performance information may be less accurately perceived, yet may weigh more heavily in the judgment process (Landy & Farr, 1980). The rationale behind this suggestion is based on the differential accuracy phenomenon (Gordon, 1970, 1972). Gordon's research indicated that ratings of behavior frequency were significantly more accurate for favorable rather than unfavorable behavior.

Unfavorable information may be treated differentially in the judgment process. Research in selection interviews indicates that early impressions dominate acceptance decisions (Springbett, 1958). An early unfavorable impression in selection interviews was followed by a decision to reject the applicant 90% of the time (Bolster & Springbett, 1961).

Considering the impact that negative information may have, what impressions might be formed by a rater evaluating a employee whose performance is variable? Scott and Hamner (1975) provide some clues about rater expectations of ratees with variable performance. They manipulated performance variability by having confederates perform an experimental task (bagging marbles) over several trials varying the level of performance for each trial while subjects observed their performance. In this study, the task motivation of workers with variable performance was rated significantly lower than that of workers with stable performance. Overall task performance ratings were similar
for workers with variable and consistent performance records; however, subjects were made aware of performance standards, could observe the specific quantity of work done, and were given information regarding the quality of work. They merely had to keep track of the number of marbles bagged on each trial. Under these circumstances, the subjectivity of performance judgment is greatly reduced compared to the typical performance appraisal situation. To the extent that attributions about motivation were affected by performance variability in the experiment, it seems likely that during a more complex rating task, similar attributions may influence ratings.

Certain patterns of performance may be cognitively consistent with other ratee characteristics. For example, raters may attribute consistent patterns of performance to tenured employees and variable patterns of performance to untenured employees. Considering a variable similar to tenure, employee age, Cleveland and Landy (1983) manipulated the performance pattern of hypothetical employees of different ages. Based upon previous research (Cleveland & Landy, 1981), the researchers provided raters with either a set of hypothetical performance ratings low on self-development skills, interpersonal skills, technical competence, problem solving, and attention to detail (stereotypic older employee pattern) or a set of ratings high on these dimensions (stereotypic younger employee pattern). Other dimensions were rated for the groups and the actual total performance for the two patterns was equated. Cleveland and Landy (1983) found that the two ratee characteristics, age and performance pattern, differentially affected manager's allocation of awards and promotions to hypothetical employees. Interestingly, managers rated the overall performance higher for younger employees
whose performance pattern was described as stereotypically young than for younger employees whose pattern was described as stereotypically old. This finding lends credence to the idea that a performance pattern inconsistent with a prototype can influence performance ratings.

Schein (1978), building upon earlier longitudinal research on career stages, suggests that when employees reach midcareer it is implied that "one is perceived as a steady contributor for the remainder of the career", while those just starting a job are in a state of learning and discovery. This suggestion is relevant to performance expectations which may influence performance appraisal ratings. One may expect that experienced employees would have relatively stable performance, while inexperienced employees would have variable performance. Because this assumption was important to the rest of the research, the hypothesis that raters would expect experienced employees to have stable performance and inexperienced employees to have variable performance was tested in a pilot study. The following major hypothesis follows logically.

**Hypothesis 2**

Since information inconsistent with prototype expectations such as those discussed above may result in biased ratings (Ilgen & Feldman, 1983), it was hypothesized that when rater experience and performance variability are not consistent with these expectations, rating accuracy will be affected.

**Cognitive Complexity**

Analysis of complex performance information including such ratee characteristics as performance variability and ratee experience may require a rater with higher cognitive abilities. Feldman (1981) has suggested that the flood of information available to a rater must be
simplified in some way during the rating process. Raters use cognitive schemata to categorize a ratee's behavior so that a rating can be determined. The more complex the schemata or dimensions, the more accurate are perceptions of actual performance. When simple schemata are used to categorize a ratee's performance as a high or low performance prototype, the perception of actual behavior may become distorted (DeNisi et al., 1984; Feldman, 1981; Cantor & Mischel, 1979).

Ratee characteristics may influence a rater to classify the ratee into a high or low performer prototype regardless of actual behavior. Ratings based on such a simple classification may be inaccurate if the prototype does not match actual performance. When the prototype used is consistent with actual performance, cognitive consistency exists and more accurate ratings are expected.

The rater's ability to deal with complex schemata may be important in determining the extent to which cognitive inconsistencies affect rating accuracy. Cognitive complexity is a construct which illustrates this ability. Schneier (1977) defines cognitive complexity as "the degree to which a person possesses the ability to perceive behavior in a multidimensional manner." A cognitively complex person would supposedly be able to make fine discriminations in behavior across many dimensions. The idea that a cognitively complex rater would make more effective performance appraisal ratings was suggested by Schneier (1977) in what has come to be known as cognitive compatibility theory. The theory proposes that a rater's cognitive structure should match the demands of the performance appraisal format in order for psychometric quality and accuracy to be high. Schneier (1977), using Bieri et al.'s (1966) version of Kelly's (1955) Role Construct Repertory (REP) Test to
measure cognitive complexity, found preliminary evidence that cognitive complexity is related to less psychometric error on behaviorally anchored rating scales (BARS). However, three subsequent empirical investigations failed to find evidence of the relationship between cognitive complexity and psychometric error (Bernardin & Boetcher, 1978; Eder, Beatty, & Keaveny, 1979; Sauser & Pond, 1979). Other studies have failed to show a relation between cognitive complexity using Schneier's (1977) method of measurement and rating accuracy (Bernardin & Cardy, 1981; Borman, 1979; Bernardin, Cardy, & Carlyle, 1982). In addition, using alternative measures of cognitive complexity that required subjects to generate behavioral dimensions for two jobs, support was found for a cognitive complexity relation with halo error but not with rating accuracy (Bernardin & Cardy, 1981; Bernardin, Cardy, & Carlyle, 1982).

None of these studies, however, manipulated ratee characteristics to test whether cognitive complexity would improve accuracy when cognitively inconsistent prototype information is present.

Ratee characteristics contribute additional dimensions of information to an appraisal situation. Not only must a rater consider the performance dimensions and behavioral anchors included on a BARS format, but a rater must also consider ratee characteristics which may not be cognitively consistent with a prototype. A cognitively simple rater may be more influenced by ratee characteristics than the actual behavior in question when actual performance is inconsistent with the prototype. A cognitively complex rater may be more able to deal with such contradictory information by using additional schemata.

**Hypothesis 3**

It was hypothesized that cognitively complex raters will have more
accurate ratings under cognitively inconsistent conditions (i.e., when ratees have much job experience and high performance variability, and when ratees with little job experience have low performance variability).

Attributions

People often make attributions concerning whether another's actions are caused by something in the person (internal factors) or something outside the person (external factors) (Heider, 1958, ch. 4). Internal factors are causes attributed to the person such as ability or effort. External factors are causes attributed to something external to the person (e.g., task or chance).

Kelley (1972, 1973, 1980) contends that in order to make causal attributions regarding an actor's behavior toward a stimulus object in a particular situation, people rely on information concerning (a) consensus - others' behavior in the same situation; (b) consistency - the actor's behavior across other situations; (c) distinctiveness - the actor's behavior toward the stimulus object in contrast to behavior toward other objects. Moreover, in the absence of such information, attributions are made on the basis of a single observation using causal schemata, or beliefs concerning how certain causes interact with certain effects.

Consider a rating situation in which one employee is rated for overall performance across several performance intervals in one work situation. Consensus and distinctiveness information would not be available to the rater, although consistency information could be provided by reviewing changes in behavior across performance intervals.

Thus, as Kelley suggests, consistency information should become more
important in the absence of consensus and distinctiveness information; and, with less information available, causal schemata should prevail in the attribution process.

Simple causal schemata may operate on what has been described as the "discounting principle" which suggests that a given cause is discounted by attributors if other plausible causes are present (Jones et al., 1961; Thibaut & Riechen, 1955). This principle implies further that a cause is attributed to the situation (external) or the the person (internal) depending upon whether the behavior is expected or unexpected.

That is, unexpected behavior is often explained more in terms of the person than the situation (Jones, Davis, & Gergen, 1961). However, unexpected behavior is not always attributed to either an internal or an external cause. Lalljee, Watson, and White (1982) assert that explanations for unexpected behavior are more complex than for expected behavior, and probably involve both internal and external attributions.

Particularly, incongruent (or unexpected, cognitively inconsistent) information may result in deeper processing of causal schemata, block continued comprehension of other aspects of the event, and increase scrutiny of causal elements (Hansen, 1985). More attention may be given to the behavior of the person being observed and the causes of the behavior without awareness of the automatic process that resulted in their salience. Congruent (cognitively consistent) information may be given less attention in the attribution process while incongruent (cognitively inconsistent) information is highlighted resulting in a "texturing" of information (Hansen, 1985).

The previous contention that raters may expect experienced employees to have stable performance and inexperienced employees to have variable
performance is relevant to this discussion of the attribution process. Internal and external attributions should be stronger in cognitively inconsistent conditions (i.e., an experienced employee with variable performance and an inexperienced employee with stable performance), because prototype expectations would be disturbed by the cognitive inconsistencies in experience and performance variability.

Scott and Hamner's (1975) results indicate the type of internal attribution that might relate to performance variability. As discussed earlier, task motivation (effort) was rated lower for subjects with variable performance. External attributions were not investigated in the study.

Attributions concerning effort seem to be plausible for explaining an experienced employee's performance when it is variable, because the ability to perform the job consistently is presumed. Ability may be a more plausible explanation of performance when an inexperienced employee behaves consistently, because inexperience implies that most new employees have not yet mastered all the job skills necessary for stable performance.

Some external attributions also appear logical for the cognitively inconsistent conditions. When an experienced employee's performance is variable, it may be viewed as resulting from chance occurrence, because of the presumed ability to deal with difficult tasks. When an inexperienced employee exhibits stable performance, it may be seen as resulting from the task itself, because the task demands would remain the same across different performance intervals.

**Hypothesis 4**

Specifically, as outlined above, it is hypothesized that
overall performance will be attributed to effort or chance when the employee with much job experience performs variably and to ability or the task when the employee with little job experience performs consistently.

Current Research

The current research explored how complex rating conditions affect rater attributions and the accuracy of performance appraisal ratings. To establish complex rating conditions, the ratee characteristics of job experience and performance variability were manipulated in two laboratory experiments. Because ratings of experienced employees are often inflated, more accurate ratings were expected for inexperienced employees.

First, a pilot study was conducted to determine whether raters have prototype expectations regarding the overall performance and performance variability of ratees with different levels of job experience. Second, Experiment 1 investigated rating accuracy and rater causal attributions of performance when varying degrees of both ratee characteristics (job experience and performance variability) are cognitively inconsistent with rater expectations. In addition, cognitive complexity was investigated as a possible rater characteristic that may possibly improve rating accuracy when ratee characteristics are cognitively inconsistent. Third, Experiment 2 served as a partial replication of Experiment 1 using different rating stimuli from Experiment 1.

In sum, the four principle hypotheses of the research are:

(1) Experienced ratees will overrated, inexperienced will be underrated, or overrating of experienced ratees and underrating of inexperienced ratees will occur.

(2) Accuracy of ratings will be lowest when high job experience and
variable performance is present, and when low job experience
and consistent performance is present (cognitively inconsistent
conditions).

(3) Accuracy will be greatest for more cognitively complex raters
when job experience and performance variability are cognitively
inconsistent.

(4) Overall performance will be attributed to effort or chance when
the employee with much job experience performs variably and to
ability or the task when the employee with little job
experience performs consistently.
Pilot Study

It was hypothesized that subjects would expect that experienced employees will have stable performance and inexperienced employees will have variable performance.

Method

Subjects

One-hundred-forty-seven undergraduate students participated in the study for extra credit in a Psychology course. Subjects were randomly assigned to one of three experimental cells.

Design

A one-way ANOVA design with ratee job experience as the three level between-subjects variable was used. The three levels consisted of no experience, five years experience, and ten years experience. The dependent variables were 7-point scale ratings on two questions (Appendix A). First, subjects were asked how variable they expected the performance of a thirty-eight year old instructor with a described experience level to be. Second, subjects were asked to rate the expected level of overall performance for the instructor.

Results

A one-way ANOVA and Duncan's Multiple Range Test indicated that the subjects expected inexperienced instructors to have more variable performance than ones with five or ten years experience ($F_{2,144}=6.67$, $p<.01$, $w^2=.07$). In addition, a separate one-way ANOVA and Duncan's Multiple Range Test indicated that the subjects expected more experienced instructors to have higher levels of performance overall ($F_{2,144}=38.42$, $p<.0001$, $w^2=.34$). Each group was significantly different from the others.
Experiment 1

The first experiment utilized written descriptions (vignettes) of an instructor's teaching performance as the rating stimuli. Job experience and performance variability were manipulated in the descriptions.

Method

Subjects

One-hundred-thirty-five undergraduate students who had not participated in the Pilot Study participated in the experiment for extra credit in a psychology course. Fifteen subjects were assigned to each experimental cell.

Design

A regression design with three between-subjects variables (Ratee Job Experience X Ratee Performance Variability X Cognitive Complexity) was employed. Ratee Job Experience consisted of three conditions: no experience, five years experience, and ten years experience. The manipulation of ratee performance variability involved six vignettes presented in three different conditions: consistent, moderately variable, and highly variable performance.

True scores for each dimension within a vignette were determined by averaging the effectiveness levels of the behaviors for each dimension. Dimensions within each vignette had approximately the same true scores. Each vignette true score was the mean of the dimension true scores. In the consistent condition, all vignette true scores were approximately 6.0 based on an 11-point scale. Each vignette true score was varied for the moderate condition (approximately 4, 5, 6, 6, 7, 8) and the high condition (approximately 1, 3, 5, 7, 9, 11). Also, vignette true scores in the two variable conditions were counterbalanced to control for order effects (i.e., increases or decreases in performance). Overall true
effects (i.e., increases or decreases in performance). Overall true ratings for the ratee described in the six vignettes were determined by the average of the vignette true scores, and thus were approximately 6.0 in all three conditions.

Procedure

Overview. Subjects completed three measures of cognitive complexity: a modified version of the one used by Schneier (1977) and two alternative measures used by Bernardin and Cardy (1981). Subjects also completed performance ratings, an attribution questionnaire, and a manipulation check questionnaire.

Vignettes. Each vignette consisted of behaviors describing a college professor's teaching performance. Behaviors for each vignette were derived from critical incidents gathered by Sauser, Evans, and Champion (1979) and Wilhelm (1986). Each vignette contained ten behaviors from five dimensions of teaching performance (relationship with students, reasonableness of workload, and fairness of testing and grading). Vignettes included two logically consistent behaviors from each dimension. Different behaviors were used in each vignette presented to a subject.

Instructions. Subjects were told that the series of six vignettes were descriptions of a professor's teaching performance while presenting six different units of course study during a semester. Each vignette within the series of six described the professor's teaching of a single undesignated unit of study preceding an exam. Job experience was manipulated by describing the experience level of the professor at the beginning of each vignette.

Instruments

Cognitive Complexity. Subjects first completed Bieri et al.'s (1966)
version of Kelly's (1955) Role Construct Repertory (REP) Test to measure cognitive complexity (Appendix B). The REP Test was modified to avoid positive response bias (Vannoy, 1965). Schneier's (1979) analysis of the Bieri et al. (1966) instrument showed it to have high test-retest reliability ($r = .54, p < .001$) and convergent and discriminant validity. Thus, the instrument has satisfactory internal validity. In addition, norm tables developed by Schneier (1979) offer evidence for establishing the external validity for the instrument.

Subjects also completed two alternative measures of cognitive complexity (Job Description Questions - JDQ1 and JDQ2) used by Bernardin and Cardy (1981) (Appendix C). These measures required subjects to generate and define as many performance dimensions they could for the job of a manager giving employee performance feedback (JDQ1) and the job of college instructor (JDQ2). The two measures were scored by counting the number of dimensions generated.

Two graduate students independently scored the two Job Description Questions. Interrater reliabilities for scoring JDQ1 ($r = .89, p < .0001$) and JDQ2 ($r = .98, p < .0001$) were acceptable.

Rating Scales. Subjects were provided with a booklet containing the vignettes. Five 11-point behaviorally anchored rating scales (BARS) followed each vignette (Appendix D). Subjects rated only the performance illustrated in the vignette preceding each set of these BARS. Five BARS of the same type were presented at the end of the booklet for subjects to rate the overall performance on the five performance dimensions. Each scale was constructed using three of Sauser, Evans, and Champion's (1979) critical incidents to anchor the ends and middle of the scale. Anchors were different from the behaviors used in the vignettes.

Attributions. Causal attributions were examined by subjects'
responses to eight questions on a 7-point scale (Appendix E). Subjects were asked to what extent they believed the instructor's overall performance throughout the semester was due to effort, ability, the task, or chance.

Manipulation Checks. Using separate 7-point scales, subjects responded to two questions regarding manipulations (Appendix F): (a) How experienced in college level teaching was the professor? (responses ranging from not at all to extremely). (b) How much did the professor's performance level change (for better or worse) over the semester? (responses ranging from not at all to extremely).

Dependent Measures

Rating Accuracy. Cronbach's (1955) four independent components of rating accuracy were used as dependent measures (Table 1). The four components are (a) Elevation, (b) Differential Elevation, (c) Stereotype Accuracy, (c) Differential Accuracy. Accurate ratings result in low scores on these measures.

Traditionally, these components are defined as tendencies in accurately rating a group of ratees. In this study, the accuracy components represented tendencies in accurately rating a single ratee at different performance intervals represented by the last four vignettes. Only ratings from the last four vignettes were used for the analysis of the accuracy components, because the effect of performance variability (in the previous vignettes) upon rating accuracy of subsequent performance intervals was of interest.

Therefore, in this study, (a) Elevation was a measure of the tendency to rate higher or lower than the average true level of performance for the performance intervals (vignettes); (b) Differential Elevation was a measure of the ability to differentiate between the
average true levels of performance in each performance interval (vignette); (c) Stereotype Accuracy was a measure of the rater's differentiation among average dimensional performance levels; (d) Differential Accuracy was a measure of the rater's ability to differentiate among performance intervals (vignettes) within performance dimensions. Together, these four accuracy components should provide a clear picture of the effect performance variability and job experience may have upon a rater's ability to approximate true ratings in subsequent performance intervals.

**Overall Performance.** In addition to analysis of the effects of the independent variables upon rating accuracy of performance intervals, their effects upon ratings of overall performance was investigated. Ratings from the final set of BARS presented were analyzed for their total accuracy (i.e., difference of the average rating across dimensions from the true overall rating, approximately six).

**Deviations.** Accuracy components represent only the distance of actual ratings from true ratings. The direction of the difference between actual ratings and true ratings is not represented by the accuracy components. For instance, if one group made ratings one unit above the true ratings and another group made ratings one unit below the true ratings, there would be no difference reflected in accuracy. Therefore, average deviation scores for ratings were also calculated. The average deviation scores were determined by taking the average difference of the true ratings from the actual ratings. These scores reflect the tendency for raters to underrate or overrate a ratee. Deviation scores were also calculated for the overall ratings.

**Attributions and Manipulation Checks.** In addition to the accuracy components, subjects responses to the questions about (a) effort,
ability, task, and chance attributions and (b) experience and performance variability were analyzed. The first set of questions was designed to provide evidence of causal attributions that a rater makes under different conditions of job experience and variable performance. The ratings of the two questions for each type of attribution were summed to form the dependent measures. The second set of two questions was designed to check the manipulations of ratee characteristics in the vignettes. All responses were on a 7-point scale.

Results

Manipulation Checks

A one-way ANOVA and Duncan's Multiple Range Test used to check the experience manipulation showed that subjects perceived ratees with no experience as less experienced than ratees with five or ten years of experience ($F_{2,126}=37.56$, $p<.0001$, $w^2=.07$). No differences in experience were indicated for the ratees with five and ten years of experience. The variability manipulation was checked by comparing the standard deviations of ratings within dimensions across the six vignettes. Standard deviations in the low and moderate variability conditions were all approximately 2.0 and 2.5 respectively indicating some variability in the ratings in the low variability condition. Standard deviations in the high variability condition were all approximately 3.3 indicating more variability in ratings in the high variability condition than in the other conditions. Also, results from a one-way ANOVA an manipulation questionnaire responses and Duncan's Multiple Range Test showed that the low, and moderate, variability conditions were both perceived as moderate, in variability while the high variability condition was ($F_{2,126}=28.37$, perceived as high in variability $p<.0001$, $w^2=.29$). Means for the manipulation check ratings are reported in Table 2.
Manipulated Variables

Accuracy. Before cognitive complexity data were analyzed, analysis of the two manipulated variables was undertaken. An Experience X Variability MANOVA performed on the accuracy components showed a main effect for Variability using Wilk's criterion ($F=10.65$, $p<.0001$). Follow-up Experience X Variability ANOVA's and Duncan's Multiple Range Tests were conducted on the accuracy components (Tables 3a & 3b). Significantly more accurate ratings were present in highly variable conditions for Elevation ($F_{4,126}=3.18$, $p<.05$, $w^2=.03$). Significantly more accurate ratings were present in the high and moderate variability conditions for Stereotype Accuracy ($F_{4,126}=23.15$, $p<.0001$, $w^2=.24$), and Differential Accuracy ($F_{4,126}=27.92$, $p<.0001$, $w^2=.29$). The ANOVA for Differential Elevation was not significant.

The accuracy of the overall ratings was analyzed using an Experience X Variability ANOVA and Duncan's Multiple Range Test (Table 4). Main effects for Experience and Variability were found. Highly experienced ratees were rated less accurately on overall ratings ($F_{2,126}=3.50$, $p<.05$, $w^2=.03$). Also, moderately variable performance was less accurately rated than consistent and highly variable performance ($F_{2,126}=4.24$, $p<.05$, $w^2=.04$).

Deviation. Average deviation scores were analyzed using an Experience X Variability ANOVA and Duncan's Multiple Range Test (Table 5). A main effect for Experience was found and analysis of means indicated that inexperienced ratees and those with ten years of experience were significantly more underrated than the others ($F_{2,126}=4.72$, $p<.01$, $w^2=.04$). A main effect for Variability and analysis of means indicated that ratees with consistent performance were significantly more underrated than the other groups ($F_{2,126}=3.80$, $p<.05$,
Overall ratings were analyzed for their deviation from true overall ratings using an Experience X Variability ANOVA. No significant effects were found.

**Cognitive Complexity**

Intercorrelations of the three cognitive complexity measures were undertaken to investigate if they in fact measured the same construct. Only JDQ1 and JDQ2 were found to be similarly related \((r = .32, p < .001)\).

The following analyses were conducted to investigate possible interactions among cognitive complexity and the manipulated variables. Because the effects of the manipulated variables have already been investigated, further analysis of their main effects was not undertaken.

**Accuracy.** Three Experience X Variability X Cognitive Complexity MANOVA's on the accuracy components using the modified version of Kelly's Role Construct Repertory (REP) Test, JDQ1 AND JDQ2 as the measures of cognitive complexity were not significant.

The accuracy of the overall ratings was analyzed in separate Experience X Variability X Cognitive Complexity regressions, each utilizing one of the three cognitive complexity measures. No statistically significant effects were found.

**Deviation.** Experience X Variability X Cognitive Complexity regressions were conducted on average deviation scores using the three cognitive complexity measures separately. One of the regressions using Schneier's measure of cognitive complexity, the REP Test showed an Experience X Cognitive Complexity interaction \((F(4,117) = 3.53, p < .05, \ w^2 = .01)\) (Table 6). Cognitive Complexity correlated positively with the deviation scores in the moderate experience condition, yet negatively in the other conditions. Low scores on the cognitive complexity measure
indicate high cognitive complexity. The deviation score means for the three experience conditions were all negative indicating that most ratees were underrated. Higher cognitive complexity was associated with less underrating in the low and high experience conditions, but not in the moderate experience condition. In the moderate experience condition, higher cognitive complexity was associated with more underrating. However, the correlations between the cognitive complexity measure and the deviation scores were not statistically significant, nor were they significantly different from one another.

Deviations of overall ratings were analyzed in separate Experience X Variability X Cognitive Complexity regressions using each of the three cognitive complexity measures. None of these regressions was significant.

**Attributions**

The attribution questionnaire was analyzed by using two subscales of the questionnaire. Effort and chance attribution ratings were combined because, effort and chance are considered unstable attributions, while ability and the task are considered stable attributions (Kelley, 1973). Also, it was hypothesized that overall performance would be attributed to effort or chance when the ratee with much job experience performs variably and to ability or the task when the ratee with little job experience performs consistently. Internal consistency (coefficient alpha) was .39 for the effort and chance subscale and .52 for the ability and task subscale. ANOVA's were used to test whether stronger attributions would occur in the cognitively inconsistent conditions. Only one significant Experience main effect was found by combining effort and chance attribution ratings (F4,126=3.93, p<.05, w2=.07). Duncan's Multiple Range Test showed that ratees with five years of experience
received stronger effort and chance attributions than those with ten years of experience (Table 7).
Experiment 2

The second experiment involved rating an instructor's performance after viewing three videotaped lectures. The audio-visual rating stimuli provided a means of partially replicating Experiment 1. In addition, a ten-item Behavioral Observation Scale (BOS) developed by Murphy et al. (1982) was used instead of a Behaviorally Anchored Rating Scale.

Hypotheses

The same hypotheses tested in Experiment 1 were tested in Experiment 2.

Method

Subjects

One-hundred-thirty-five undergraduate students who had not participated in the previous studies participated in Experiment 2 for extra credit in a psychology course. Fifteen subjects were assigned to each experimental cell.

Design

The same regression design employed in Experiment 1 was used. The manipulation of ratee job experience involved verbally instructing subjects before each videotape that the ratee had no experience, five years experience, or ten years experience. The manipulation of performance variability involved three videotaped lectures performed by an actor and presented in three different conditions: consistent, moderately variable, and highly variable.

True performance scores for the videotapes were based on the average summated scores on the BOS completed by three graduate students. Summated scores were calculated by summing the ratings of all BOS items except 1 and 5, because these items were not manipulated by the actor during filming. Six out of fifteen lectures were selected based on the
graduate students' average summated scores for each lecture. Three lectures differing in topic were presented in each of the variability conditions.

In the consistent condition, the three videotapes depicted a ratee with true performance score of 38.33, 39.98, and 41.33 based on the average BOS summated score of eight items on a 7-point scale. Each videotape true score was varied for the moderate condition (38.33, 41.33, and 48.00) and for the high condition (21.00, 38.33, and 52.99). The order of presentation of videotapes in the variable conditions was counterbalanced to control for order effects (i.e., increases or decreases in performance). The overall performance true scores for the ratee depicted in the three videotapes were determined from the average of the videotape true scores, and thus were 39.88, 42.55, and 39.88 in the low, moderate, and high conditions, respectively.

**Procedure**

**Overview.** Subjects completed the three cognitive complexity measures used in Experiment 1. They then viewed three videotapes, and after the third, rated the performance of the ratee depicted in the three videotaped lectures using Murphy et al.'s (1982) BOS. Subjects also made ratings of their overall impression of the instructor's performance (Appendix G). Following the performance rating, they completed an attribution questionnaire and a manipulation check questionnaire identical to those used in Experiment 1.

**Development of Videotapes.** An actor was filmed presenting good, average, and poor quality lectures on each of four different management topics in a classroom setting. The actor varied his level of lecture performance using Murphy et al.'s (1982) BOS for college instructors as a guide.
Prior to Experiment 2, three graduate students in industrial
psychology rated fifteen of the videotaped lectures using Murphy et al.'s
(1982) BOS. Consistent with Murphy et al.'s (1985) videotape development
procedure, the graduate student raters were blind to the intended use of
the tapes, and were given multiple opportunities to view each tape. In
addition, each of the graduate student raters rated the lectures at
different viewing times and were presented the tapes in random order.
Interrater reliability (coefficient alpha) for the graduate student
rater's overall ratings of the fifteen lectures was .99. Their summated
scores correlated highly with overall ratings of the lectures (r=.99,
p<.0001).

**Dependent Measures.** Since accuracy components could not be
adequately analyzed on performance variability of three lectures, two
dependent measures of accuracy were used. First, the absolute value of
the difference of the actual summated scores from the true summated
scores was used. Second, the absolute value of the difference of the
actual overall ratings from the true overall ratings was used. In
addition, deviation scores for both summated scores and overall ratings
were also calculated as in Experiment 1. Fourth, the attribution and
manipulation check measures used in Experiment 1 were also used in
Experiment 2.

**Results**

**Manipulation Checks**

Experience X Variability ANOVA's for the manipulation checks were
significant for job experience (F_{2,126}=63.05, p<.0001, w^2=.49) and
variability (F_{2,126}=11.89, p<.0001, w^2=.14). Duncan's Multiple Range
Test showed that raters with zero, five, and ten years of experience were
all perceived as having different levels of experience and only the
highly variable ratees were perceived as having more variability than the moderate and consistent groups (Table 8).

Manipulated Variables

**Accuracy.** Before cognitive complexity was investigated, the effects of the manipulated variables on accuracy were analyzed. In an Experience X Variability ANOVA, an Experience main effect was found on summated score accuracy ($F_{2,126}=4.91$, $p<.01$, $w^2=.05$). Duncan's Multiple Range Test showed that there was significantly lower accuracy in the summated scores for ratees with no experience. In addition, a Variability main effect was found ($F_{2,126}=3.39$, $p<.05$, $w^2=.03$) and analysis of means using Duncan's Multiple Range Test showed that highly variable performance was rated significantly more accurately (Table 9).

The accuracy of overall ratings was analyzed using a separate Experience X Variability ANOVA and Duncan's Multiple Range Test. A main effect for Variability was found ($F_{2,126}=7.87$, $p<.001$, $w^2=.09$) and analysis of means showed that moderately variable performance was rated significantly more accurately than the other variability groups (Table 10).

**Deviation.** The deviations of the summated scores were analyzed using an Experience X Variability ANOVA and Duncan's Multiple Range Test. A main effect for Variability was found ($F_{2,126}=6.17$, $p<.01$, $w^2=.07$) and analysis of means showed that inexperienced ratees were significantly more underrated (Table 14). Also, a main effect for Experience was found ($F_{2,126}=5.05$, $p<.05$, $w^2=.05$) and analysis of means indicated that highly variable performance was underrated less than consistent or moderately variable performance (Table 11).

The deviations of overall ratings from true overall ratings were analyzed for accuracy using an Experience X Variability ANOVA and
Duncan's Multiple Range Test. A main effect for Experience was found
\( (F_{2,126}=5.98, p<.01, \omega^2=.07) \) and analysis of means indicated that
inexperienced ratees were significantly more underrated than experienced
ratees. Also a main effect for variability was found \( (F_{2,126}=5.54,\)
\( p<.01, \omega^2=.06) \) and analysis of means indicated that significantly less
underrating occurred in the moderate variability condition (Table 12).

**Cognitive Complexity**

Intercorrelations of the three cognitive complexity measures were
undertaken to investigate if they in fact measured the same construct.
Only JDQ1 and JDQ2 were found to be similarly related \( (r=-.32, p<.001) \).

The following analyses were conducted to investigate
possible interactions among cognitive complexity and the manipulated
variables. Because the effects of the manipulated variables have already
been investigated, further analysis on their effects was not undertaken.

**Accuracy.** Experience X Variability X Cognitive Complexity
regressions were conducted using each of the three cognitive complexity
measures in separate regressions on the accuracy of summated scores. No
statistically significant effects were found. Regressions were also
performed on the accuracy of the overall ratings. Also, no statistically
significant effects were found for these regressions.

**Deviation.** Experience x Variability X Cognitive Complexity
regressions were conducted using each of the three cognitive complexity
measures in separate regressions on the deviations of the summated scores
from the true summated scores. No interactions were found in these
regressions.

The deviations of overall ratings from the overall true ratings were
analyzed in separate regressions using each of the three cognitive
complexity measures. The only significant effects were when JDQ1 was
used as the cognitive complexity measure. An Experience X Variability X Cognitive Complexity (JDQ1) interaction occurred (F4,117=2.93, p<.05, w2=.05) (Table 13). Simple effects analysis indicated that cognitive complexity (JDQ1) was negatively correlated with the deviation scores (r=-.51, p<.05) when variability and job experience were both high (a cognitively inconsistent condition). Since most of the deviation scores were negative, it appears that higher cognitive complexity was associated with more underrating under this cognitively inconsistent condition (Table 14). Correlations were not significant in the other conditions.

Attributions

The attribution questionnaire ratings were analyzed the same as they were in Experiment 1. No significant effects on attributions resulted in the Experience X Variability ANOVA's.
Discussion

The Job Experience Prototype

Overall, the data indicate that raters do have prototype expectations regarding the performance variability of employees with various levels of experience and such expectations may have a slight effect on the accuracy of ratings. The pilot study showed that raters expect inexperienced professors to have more variability in performance than experienced professors. Also, students appear to believe that professors with more experience will have higher levels of overall performance. Together, these findings suggest that a prototype regarding performance levels and their fluctuation exists in relation to a ratee's level of job experience.

Job Experience and Ratings

Analysis of the manipulated variables suggested that inexperienced ratees were rated less accurately than experienced ratees. Hypothesis 1 regarding more accurate ratings for inexperienced ratees was not supported. Apparently there was greater underrating of inexperienced ratees relative to experienced ones. Inexperienced and highly experienced ratees were underrated more than moderately experienced ratees in Experiment 1; however, in Experiment 2, subjects rated inexperienced ratees less accurately and underrated them relative to more experienced ratees. This finding in Experiment 2 is more consistent with student prototype expectations for experienced and inexperienced professors' performance levels. Therefore, there appears to be some general support for Leventhal et al.'s (1977) finding that more experienced instructors receive higher performance ratings. Experience may be a powerful ratee characteristic which activates rater schemata and influences the
evaluation of ratee performance information.

Performance Variability

Although no main effects for variability were hypothesized, some unusual effects occurred in the variability conditions. Highly variable performance was generally rated more accurately and was less underrated in performance interval ratings in Experiment 1 than moderately variable and consistent performance. These effects could have occurred because raters in the highly variable condition were exposed to a greater variety of behaviors representing more scale values than any other condition. Through observing the extreme differences in performance, these raters may have been in a better position to learn to use more scale values than raters who were exposed to somewhat similar ratee performance levels.

In addition, moderately variable performance was given the least accurate overall ratings in Experiment 1. Noticeably extreme differences in behavior in the highly variable condition may have influenced raters to view the ratee as having neither extremely good nor bad overall performance on the behavioral dimensions. By viewing extreme performance fluctuations, average level overall impressions of ratee performance on behavioral dimensions may have resulted, which in this case, would be a more accurate view of the ratee's true performance. In the consistent condition, there is virtually no variability, so an overall impression should not have changed from each average level vignette to another. With moderate ratee variability, ratees may have more difficulty forming an accurate overall impression because extreme fluctuation is absent. When ratees are moderately variable, raters may not interpret the moderate fluctuations from average performance as indicating overall average performance.

In Experiment 2, summated scores showed a similar effect to that
found with performance interval ratings (vignette ratings) in Experiment 1. More accurate ratings and less underrating of highly variable performance occurred for the summated scores. Observing specific behaviors as they fluctuate across performance intervals may have provided more experience regarding the extremes of the rating scale for the raters. Also, the extremes of frequent and infrequent performance of the behaviors rated on the BOS may have been viewed as an average frequency of performance.

Contrary to Experiment 1 overall ratings of performance in Experiment 2 in the moderate variability condition were more accurate and underrated less than the other conditions. Overall impressions of the ratee's performance may have been more accurate in the moderately variable condition because the highly variable condition contained an example of very poor lecture quality and the consistent condition contained no improvement in performance above an average level. It is possible that forming an accurate impression of the ratee's overall performance is more difficult if an extremely negative fluctuation in performance occurs (Gordon, 1970, 1981) or if performance never really changes (e.g., in the consistent condition).

A possible explanation for the different findings regarding overall ratings in the two experiments may be the differences in the rating tasks. Experiment 1 required subjects to rate the instructor after each performance interval. Experiment 2 required subjects only to view each performance interval and rate the instructor's performance over all intervals without prior practice in rating, although subjects were familiar with the scale before exposure to the rating stimuli. Perhaps, the extreme variability conditions (i.e., low and high variability) are more difficult to accurately perceive and rate overall without prior
practice rating each performance interval. This may be due to possible biasing effects of stable average ratee performance in the consistent condition and extremely unstable changes in performance in the highly variable condition. When performance intervals are not rated, stable average performance may lead to less accurate perceptions of overall performance because no improvement or decrement in performance occurs. Extremely unstable performance may lead to less accurate perceptions of overall performance because improvement or decrement in performance from the first to the last performance interval does not always occur. Rather than focusing on variability, it may be worth investigating the effects of stable levels of performance and the direction of change in performance on the accuracy of overall performance impressions when performance intervals are not carefully monitored and evaluated (as in Experiment 2). However, it should also be noted that the accuracy components may be very sensitive to high variability in true ratings and thus the results could be due to a statistical artifact. Further research could investigate how each component is affected by different distributions of ratings and true scores.

Cognitive Complexity

Hypotheses 2 and 3 dealing with accuracy under cognitively inconsistent conditions gained no support. The effects on accuracy were not significant in Experiment 1 using written rating stimuli. The lack of consistent findings involving cognitive complexity may be due to measurement difficulties. The REP measure of cognitive complexity did not correlate with the JDQ measures; however, the two JDQ measures did correlate positively with each other. Similar findings for these measures have been demonstrated in previous research (Bernardin and Cardy, 1981; Bernardin, Cardy, and Carlyle, 1982). The previous and
current findings indicate that the REP Test and the JDQ measures may be measuring different constructs. In addition, evidenced by the lack of significant effects on accuracy using any of the three cognitive complexity measures, it appears that whatever the measures are measuring is not necessarily a beneficial rater characteristic.

Experiment 1 showed that cognitive complexity measured by the REP Test was slightly associated with less underrating of rating dimensions within performance intervals (deviation scores) for inexperienced and highly experienced ratees. Experiment 2 showed no effect with the REP Test. The effect in Experiment 1 could have occurred because the REP Test was actually measuring the rater's ability to fairly evaluate specific performance dimensions on BARS for ratees who possess very salient, extreme levels of job experience (i.e., inexperienced and ten years experience). The REP Test appears to focus on the ability to use multiple dimensions in judging people. Presumably, such ability would be somewhat helpful for using BARS. Perhaps, when a ratee characteristic like experience is very salient or extreme, rater cognitive complexity may reduce the tendency to underrate on performance dimensions. Experiment 2 used a BOS, which requires only behavior frequency ratings. The ability to focus on multiple behavioral dimensions, cognitive complexity, may not be as useful when a BOS is used.

No support for Hypotheses 2 and 3 was found in Experiment 2 using audio-visual rating stimuli. Cognitive complexity, measured by JDQ1, was found to be associated with more negative deviations from true overall ratings in one of the cognitively inconsistent conditions (high experience, high variability). This finding suggests that cognitively complex raters may be harsher evaluators of ratee behavior when performance variability and job experience are cognitively inconsistent.
The significance of only JDQ1 as the cognitive complexity measure supports others' contentions (Bernardin & Cardy, 1981; Bernardin et al., 1982) that the JDQ type measure is more related to predicting the quality of performance ratings than the type of measure used by Schneier (1977). The significance of JDQ1 as opposed to JDQ2 may be due to the subject population selected. That is, most of the student subjects were able to generate a large number of dimensions for the job of college instructor as requested on JDQ2. More variance in the number of dimensions generated occurred when student subjects were asked to list dimensions for the job of a manager conducting a feedback meeting (JDQ1), a relatively unfamiliar job to most college students. Perhaps cognitive complexity is best defined by the ability to generate multiple behavioral dimensions for a target stimuli one is not extremely familiar with.

Another possible reason for differences in the two experiments is that the number of performance intervals in each was different. Experiment 1 had twice the number of performance intervals of Experiment 2. Numerous performance intervals may make variability more salient, and thus, more likely to affect rating accuracy and attributions.

**Attributions**

Hypothesis 4 dealing with possible attributional effects of the ratee characteristics was not supported. Results were mixed. Stronger effort and chance attributions were given to moderately experienced ratees in Experiment 1, but this finding does not readily fit into attribution theory; however, they may give an indication of general prototype expectations. Perhaps, ratees with around five years of experience are seen as less stable in relation to their motivational state and susceptibility to good or bad luck.

The lack of significant effects for attributions could be due to the
low reliability of the subscales. The low internal consistency of the subscales attenuated any possible relations between the independent variables and attributions.

Limitations

The fact that both experiments were conducted in a laboratory setting suggests certain limitations. First, the variability manipulation was not very strong and this may have biased the results. Also students are often the raters of professors, but the students who served as subjects in the current research did not know the ratee and were not familiar with his past performance. Also, students reviewed written or audiovisual rating stimuli rather than actually viewing the ratee at work over a long period of time. Clearly, in a field setting their is more performance information to process and numerous ratee characteristics present. In such a setting, the likelihood of rater attributions and biases may be much greater than in a laboratory setting, because many interactions between raters and ratees are possible and raters may seek confirmation of their biases in order to have consistent cognitions.

However, the laboratory setting was necessary to control true levels of performance and analyze rating accuracy. In both experiments, subjects rated a performance on a job with which they were familiar, teaching performance. The difference in results between Experiment 1 and Experiment 2 may be due to the increased amount of specific information available in the vignettes opposed to the videotaped lectures.

Implications

The results may have some implications for the rating process and related personnel decisions. Rater expectations probably do exist regarding the congruence of ratee characteristics. Therefore, judgments concerning selection and classification of employees may be slightly
affected by salient ratee characteristics rather than true performance. Perhaps raters should be made wary of possible prototype expectations and how they bias ratings. Specifically, rater training programs might be directed toward modifying the tendency for people to always rate in a way which confirms their own prototype expectations. Other personnel decisions may also be affected by bias toward employees because of their level of experience or performance variability. Multiple criteria for personnel decisions should be used to avoid overstressing employee experience or performance variability as major determinants of personnel action.

Cognitive complexity, regardless of how it is measured, may not be as helpful to performance appraisal research as once believed (Schneier, 1977). The current research indicated that cognitive complexity may not always improve rating accuracy when biasing ratee characteristics are salient. Moreover, past attempts to show a relation between cognitive complexity and accurate performance ratings have not been successful (e.g., Bernardin & Cardy, 1981; Bernardin, Cardy, & Carlyle, 1982).

Because cognitive complexity may not be useful to rating accurately, the focus of research in this area should be toward the effects of ratee characteristics on the cognitive processes involved in performance judgements (e.g. prototype expectations and use of simple schemata). Further research on actual cognitive processes rather than cognitive complexity per se may lead to rater training strategies directed toward teaching raters a multidimensional view of job performance and more caution in attending to ratee characteristics.

Conclusion

This study has provided information relevant to a cognitive view of performance rating. First, prototype expectations regarding ratee
characteristics exist. Second, these expectations likely affect ratings regardless of actual behavior. Third, the salience of the performance interval information (i.e., performance variability) may affect the rating process. That is, viewing extremely variable or consistent behavior over several performance intervals without actually rating the behavior may make performance differences less salient negating the effect of extreme performance variability or consistency upon overall impressions. However, practice in rating each performance interval may make the differences in performance across the intervals more salient, thereby activating a cognitive categorization process which relies upon prototypes for performance variability. Fourth, cognitive complexity may not be a meaningful variable to consider in relation to rating bias.

The research implications of this study point to a new emphasis on rater expectations regarding ratee characteristics rather than rater cognitive structure. Future investigations could be directed toward measuring rater expectations of ratee behavior and their subsequent effect on ratings. Raters may be reacting more to expectations of behavior rather than to actual behavior. The salience of behavioral patterns may also affect cognitive processes involved in the performance rating process.
References


Appendix A

Expectations Questionnaire
Please read the following description and circle the number on each scale which corresponds with your answer.

Dr. Smith has been teaching college classes for ten years. He is 38 years old and has ten years of experience teaching on the college level. Assume that you are going to be in the class he teaches this semester. Think about how well Dr. Smith will probably teach this semester.

1. Over the course of the semester, how much do you think Dr. Smith's teaching performance will fluctuate or change? For example, if you expect him to be very good sometimes and very poor other times, indicate "extreme" fluctuation. If you think he will always be consistently good or consistently poor, mark none. Using the scale below, any value from 1 (no fluctuation) to 7 (extreme fluctuation).

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>moderate</td>
<td>extreme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Would you expect Dr. Smith's teaching to be poor, average, or good overall throughout the semester. Mark what general level of performance you anticipate from him on the scale below.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>very poor</td>
<td>average</td>
<td>very good</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

REP Test
The following grid includes a list of meaningful persons in your life (rows) and a list of adjectives that can be used to describe each person (columns). Each pair of adjectives comprises a rating scale ranging from +3 to -3. For example, in the far right column, the ratings represent:

- +3 very outgoing
- +2 moderately outgoing
- +1 somewhat outgoing
- -1 somewhat shy
- -2 moderately shy
- -3 very shy

Working across each row, write in the rating number for each pair of adjectives which best describes the person in your life. If you do not have a mother or father, substitute your legal guardian or another relative. If you do not have a boss, substitute one of your professors. Every box on the grid should contain a rating number and sign when you are finished.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
</tr>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Appendix C

JDQ1 AND JDQ2
1. If you were a manager who had to conduct a performance appraisal interview of one of your employees, what factors or dimensions do you consider important for the complete success of that interview? That is, what types of things would be necessary for the manager to do to be effective at giving an employee feedback on his or her work performance? Describe each dimension in a few words. Then, define each dimension in as much detail as you can.

2. Consider the job of college instructor. Name and define as many factors as you consider important for the success of an instructor. That is, what types of things make for a good instructor?
Appendix D

Vignettes and Rating Scales
Your task is to rate the teaching performance of the college professor described on the following pages. The professor teaches a biology course which is divided into six units of material. Each of the following six descriptions is of the professor's performance in teaching the different units of study during a semester. The professor has years of teaching experience.

Directions

1. Turn to the first rating form and familiarize yourself with the rating dimensions.

2. Carefully read through the description of the professor's performance teaching the first unit of study.

3. Turn to the rating form for the first unit of study. Without looking back at the description, clearly circle the number on each dimension scale which corresponds to your evaluation of the professor's performance in teaching the first unit of study. Do not make ratings which fall between numbers.

4. Repeat this process of reading the description and rating the professor's performance for the other five units of study.

5. Then, use the last rating form to rate the professor's overall teaching performance on the rating dimensions throughout the six units of study.
Without looking back at the description, clearly circle the number of each dimension scale which corresponds to your evaluation of the professor's performance in teaching the previously described unit of study. Make only whole number ratings for each dimension. Rating numbers at the middle and both ends of each scale are described by examples of the level of performance they represent. These examples illustrate the level of performance they represent. These examples illustrate the level of performance expected for ratings of 1, 6, and 11. Recalling the professor's performance on only the previous unit of study, circle a rating number (1 through 11) which best describes his performance on each dimension.

<table>
<thead>
<tr>
<th>DIMENSION A: Relationship With Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
</tr>
<tr>
<td>5 6 7 8 9</td>
</tr>
<tr>
<td>10 11</td>
</tr>
<tr>
<td>openly criticized students for asking questions in class</td>
</tr>
<tr>
<td>saw students in the office only if they made appointments</td>
</tr>
<tr>
<td>made a point to know every student's name</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIMENSION B: Ability to Present the Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
</tr>
<tr>
<td>lecture information conflicted badly with book information, resulting in total confusion</td>
</tr>
<tr>
<td>covered material in class that had already been presented in lab</td>
</tr>
<tr>
<td>used good teaching aids, was articulate, and stressed important points</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIMENSION C: Interest in Course and Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
</tr>
<tr>
<td>told students that he was totally disinterested in teaching and felt it a wasted of time</td>
</tr>
<tr>
<td>got so involved in the subject that he would forget to stop lecturing when class time was over</td>
</tr>
<tr>
<td>knew the material so well that he was able to answer all questions asked by his students</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIMENSION D: Reasonableness of Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
</tr>
<tr>
<td>assigned two paper per week, a textbook, and classroom work</td>
</tr>
<tr>
<td>assigned about 50 pages of reading per week</td>
</tr>
<tr>
<td>discontinued or reduced homework assignments around tests to help students</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIMENSION E: Fairness of Testing and Grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
</tr>
<tr>
<td>would not change grades even if a mistake was made in grading</td>
</tr>
<tr>
<td>did not curve grades unless the class did extremely badly</td>
</tr>
<tr>
<td>test questions were to the point and easy to understand</td>
</tr>
</tbody>
</table>
A

Professor: Jones
Years Experience: 
Unit # 
Age: 38

Read this description of professor Jones performance on unit #

This professor:

* often read straight from the book but only read every third line.
* belittled the unit material and described it as a waste of time.
* refused to change his students' grades on a test even after the students pointed out the passage in the text book which showed them to be correct.
* told his students how stupid he thought they were.
* announced the first day of class that he would assign 9 A's, 15 B's, 26 C's, 15 D's, and 9 F's as final grades.
* assigned a five-page paper two days before it was due.
* made the workload so heavy that only one student out of 25 passed.
* made a fool out of a student in class for asking a ridiculous question.
* explained things as though he were talking to a class of PhD's.
* actually tells the class that he hates the subject matter.
B

Professor: Jones
Years Experience: Unit#
Age: 38

Read this description of professor Jones' performance on unit #

This professor:

* posted office hours, but made students wait until he could find time to see them.
* asks picky test questions about details.
* lectured very rapidly.
* mentioned several times that the unit he was teaching did not represent his major area of interest.
* required four books to be read for a three hour course.
* never changed his tone or expression while lecturing.
* criticized research done by his colleagues.
* sometimes assigned two chapters for one night's assignment.
* tested over material he did not cover.
* gave the students his office number but did not make them feel welcome.
Read this description of professor Jones' performance on unit #

This professor:

* was attentive and helpful in class but was unavailable for outside help.

* required a term paper, oral presentation, and weekly tests.

* said he was going to start asking people to leave when the class became noisy.

* arrived in class a few minutes late.

* marked off for poor class attendance.

* assigned general problems in class, then gave specific problems on test.

* gave an extremely heavy assignment one week, then slacked off for a week or so before giving another assignment.

* continuously referred back to his notes while attempting to lecture.

* said, "Let's hurry up and maybe we'll get out of here early."

* gave details about the material but never elaborated beyond them.
Professor: Jones
Years Experience:

Read this description of professor Jones' performance on unit #

This professor:
* always kept his classroom presentations specific and to the point.
* gave daily reading assignments and an outline of references to use during the unit.
* tests usually covered three of four chapters of the book.
* discussed recent research findings which were related to his lecture.
* sometimes had to look at his notes to get precise information since he doesn't know all of the details from memory.
* arrived in class a few minutes late.
* firmly told the students to be quiet when the class became restless.
* gave objective tests.
* assigned a four-to-five page typewritten paper and specified the format and style in which it was to be written.
* told the students that if they missed class for a good reason, he would be willing to discuss the missed lecture, but that students could not expect him to do this if they missed because they cut class.
Read this description of professor Jones' performance on unit #

This professor:

* assigned reasonable amounts of homework every other day.

* stood in the hallway before and after class so that students can ask questions in an informal atmosphere.

* described his own fascination with the material he was covering.

* gave rest periods each week in which no homework was assigned.

* gave two lab quizzes and dropped the lowest one.

* tests had a lot of questions so that you could miss one and not worry about failing.

* compensated for limited office hours by offering his time before and after class every day.

* presented information in brief, easy-to-follow written outline form.

* showed enthusiasm in his voice.

* talked as an easy pace and occasionally wrote on the board.
Professor: Jones  
Years Experience:  
Unit #:  
Age: 38

Read this description of professor Jones' performance on unit #

This professor:

* used a variety of methods to present the material, including films, tapes, and experiments.

* when confounded by a student's question, spent several hours of his own time that same afternoon researching material for an answer.

* when shown that the textbook indicated an answer other than the one he counted was correct, admitted his mistake and changed the grades.

* assigned only as much homework as is necessary to learn the material thoroughly.

* spent an hour and a half in his office helping a student with a course related problem.

* gave his students enough time to complete his tests.

* distributed the workload evenly across the unit.

* sought out a shy student who was failing the course and worked with her until she understood the material well enough to pass.

* used good teaching aids, was articulate, and stressed important points in class.

* got excited about what he was teaching and conveyed this enthusiasm to his students.
Appendix E

Attribution Questionnaire
Circle the number which corresponds to your answer for each question.

To what extent was the professor's performance throughout the semester (over all six units of study):

1. Due to his ability or lack of ability to perform effectively?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

2. Due to the level of motivation he showed in his teaching?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

3. Due to something either easy or difficult about the task?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

4. Due to the amount of effort he put into teaching?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

5. Due to either good or bad luck?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

6. Due to chance (no obvious cause)?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

7. Due to the amount of ability he possessed?
   - 1 2 3 4 5 6 7
     none  moderately  extremely

8. Due to the nature of the task?
   - 1 2 3 4 5 6 7
     none  moderately  extremely
Appendix F

Manipulation Check Questionnaire
SS# ______________________________

Circle the number on the scale which corresponds to your answer for each question.

1. How experienced in teaching was the professor?  
   1  2  3  4  5  6  7  
   not  moderately  extremely  
   at  
   all  

2. How much did the professor's level to teaching performance fluctuate or change from each description (unit of study) to the next?  
   1  2  3  4  5  6  7  
   not  moderately  extremely
Appendix G

Rating Scales
Please think about the lecturer whose taped teaching you just observed. Rate his performance using the items listed below. Feel free to use any of the numbers in your ratings; the words "never," "half the time," and "all of the time" give you an idea about what the numbers should represent.

1. Examples were presented which were related to the control topic.

1 2 3 4 5 6 7
never half the time all of the time

2. The lecturer used nonverbal behaviors (such as gestures) to emphasize points.

1 2 3 4 5 6 7
never half the time all of the time

3. The lecturer stopped in mid-sentence.

1 2 3 4 5 6 7
never half the time all of the time

4. The lecturer lost eye contact with the audience.

1 2 3 4 5 6 7
never half the time all of the time

5. The lecturer provided evidence to support broad generalizations.

1 2 3 4 5 6 7
never half the time all of the time

6. The lecturer acted nervous.

1 2 3 4 5 6 7
never half the time all of the time

7. The lecturer spoke in a monotone for a sustained period.

1 2 3 4 5 6 7
never half the time all of the time

8. The lecturer varied his facial expression.

1 2 3 4 5 6 7
never half the time all of the time

9. The lecturer gave clear answers to questions from the class.

1 2 3 4 5 6 7
never half the time all of the time

10. The lecturer appeared unsure of what he was saying.

1 2 3 4 5 6 7
never half the time all of the time

Now consider all aspects of the lecturer's performance (including characteristics not rated above), and determine an overall rating for his performance. Remember that you have seen just a short sample of his work.

My overall rating of this lecturer is:

1 2 3 4 5 6 7
very poor poor somewhat average moderately good very good
Table 1

Four Accuracy Components

Elevation² \((X.. - T..)^2\)

Differential Elevation² \(\frac{1}{n} \left[ (X_i - X..) - (T_i - T..) \right]^2\)

Stereotype Accuracy² \(\frac{1}{k} \left[ (X.._j - X..) - (T.._j - T..) \right]^2\)

Differential Accuracy² \(\frac{1}{kn} \left[ (X_{ij} - X_i - X.._j + X..) - (T_{ij} - T_i - T.._j + T..) \right]^2\)

Where \(X_{ij}\) and \(T_{ij}\) are rating and true scores for vignette \(i\) on dimension \(j\); \(X_i\) and \(T_i\) are mean ratings and mean true scores for vignette \(i\); \(X.._j\) and \(T.._j\) are mean rating and true scores for dimension \(j\); and \(X..\) and \(T..\) are mean rating and true scores over vignettes and dimensions.
Table 2

**Means for Manipulation Check Ratings for Experiment 1**

<table>
<thead>
<tr>
<th>Experience</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.20</td>
<td>4.04</td>
<td>4.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variability</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.11</td>
<td>5.31</td>
<td>5.91</td>
</tr>
</tbody>
</table>
Table 3a

Follow-up Experience X Variability ANOVA’s on Accuracy Components for Experiment 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elevation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>2</td>
<td>1.482</td>
<td>4.52**</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>1.045</td>
<td>3.18</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>.647</td>
<td>1.98</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>126</td>
<td>.328</td>
<td></td>
</tr>
<tr>
<td><strong>Differential Elevation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>2</td>
<td>.289</td>
<td>.81</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>1.244</td>
<td>3.37</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>.420</td>
<td>1.15</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>126</td>
<td>.368</td>
<td></td>
</tr>
<tr>
<td><strong>Stereotype Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>2</td>
<td>2.539</td>
<td>.91</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>77.730</td>
<td>27.92**</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>2.927</td>
<td>.38</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>126</td>
<td>2.784</td>
<td></td>
</tr>
<tr>
<td><strong>Differential Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>2</td>
<td>3.694</td>
<td>1.68</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>51.049</td>
<td>23.15**</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>3.148</td>
<td>1.43</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>126</td>
<td>2.205</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05  **p < .01
Table 3b

Means for Accuracy Components

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elevation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variability</strong></td>
<td>.74</td>
<td>.83</td>
<td>.54</td>
</tr>
<tr>
<td><strong>Stereotype Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variability</strong></td>
<td>5.58</td>
<td>4.18</td>
<td>2.95</td>
</tr>
<tr>
<td><strong>Differential Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variability</strong></td>
<td>4.82</td>
<td>3.84</td>
<td>2.70</td>
</tr>
</tbody>
</table>
Table 4

## Experience X Variability ANOVA on Accuracy of Overall Ratings for Experiment 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>2.017</td>
<td>3.50*</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>2.444</td>
<td>4.24*</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>.755</td>
<td>1.31</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>.576</td>
<td></td>
</tr>
</tbody>
</table>

### Experience

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>.76</td>
<td>.85</td>
<td>1.16</td>
<td></td>
</tr>
</tbody>
</table>

### Variability

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>.75</td>
<td>1.18</td>
<td>.63</td>
<td></td>
</tr>
</tbody>
</table>
Table 5

Experience X Variability ANOVA on Average Deviation of Ratings from True Ratings for Experiment 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>3.568</td>
<td>4.72**</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>2.877</td>
<td>3.80*</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>.513</td>
<td>.68</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>.757</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05  
**p < .01

<table>
<thead>
<tr>
<th>Experience</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.53</td>
<td>-.07</td>
<td>-.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variability</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.46</td>
<td>-.19</td>
<td>.05</td>
</tr>
</tbody>
</table>
Table 6

**Experience X Variability X Cognitive Complexity (REP) Regression on Average Deviation from True Scores for Experiment 1**

\[ R^2 = .23 \quad F_{17,117} = 1.93^* \]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>1.91</td>
<td>2.59</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>1.66</td>
<td>2.25</td>
</tr>
<tr>
<td>REP</td>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>0.29</td>
<td>0.39</td>
</tr>
<tr>
<td>REP X Experience</td>
<td>2</td>
<td>2.60</td>
<td>3.53*</td>
</tr>
<tr>
<td>REP X Variability</td>
<td>2</td>
<td>1.36</td>
<td>1.84</td>
</tr>
<tr>
<td>REP X Experience X Variability</td>
<td>4</td>
<td>0.41</td>
<td>0.55</td>
</tr>
<tr>
<td>Error</td>
<td>117</td>
<td>0.74</td>
<td></td>
</tr>
</tbody>
</table>

REP and Average Deviation

Experience

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>-.165</td>
<td>0.236</td>
<td>-.172</td>
</tr>
<tr>
<td>( z )</td>
<td>-.167</td>
<td>0.240</td>
<td>-.174</td>
</tr>
</tbody>
</table>

(\( z \) test for difference between low and moderate experience)

1.86

(\( z \) test for difference between moderate and high experience)

1.86

(correlations and \( z \) tests are not significant)
Table 7

**Experience X Variability ANOVA for Effor and Chance Attributions Together for Experiment 1**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>45.03</td>
<td>3.93*</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>13.79</td>
<td>1.20</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>15.30</td>
<td>1.34</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>11.45</td>
<td></td>
</tr>
</tbody>
</table>

**Experience**

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14.00</td>
<td>15.04</td>
<td>13.04</td>
</tr>
</tbody>
</table>

(means based on 4 7-point scale ratings)
Table 8

Means for Manipulation Check Ratings for Experiment 2

<table>
<thead>
<tr>
<th>Experience</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.00</td>
<td>3.71</td>
<td>5.13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variability</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.80</td>
<td>4.16</td>
<td>5.42</td>
</tr>
</tbody>
</table>
Table 9

Experience X Variability ANOVA on Accuracy of Summated Scores for Experiment 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>186.13</td>
<td>4.91**</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>113.17</td>
<td>3.39*</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>60.64</td>
<td>1.80</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>33.35</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

**p < .01

Experience

<table>
<thead>
<tr>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.54</td>
<td>9.87</td>
<td>8.84</td>
</tr>
</tbody>
</table>

Variability

<table>
<thead>
<tr>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.60</td>
<td>11.90</td>
<td>8.75</td>
</tr>
</tbody>
</table>
Table 10

Experience X Variability ANOVA on Accuracy of Overall Ratings for Experiment 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>1.12</td>
<td>1.54</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>5.74</td>
<td>7.87*</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>0.68</td>
<td>0.94</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>0.73</td>
<td></td>
</tr>
</tbody>
</table>

* p < .01

Variability

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.42</td>
<td>.94</td>
<td>1.64</td>
</tr>
</tbody>
</table>
Table 11

Experience X Variability ANOVA on Deviation of Summated Scores from True Scores for Experiment 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>282.84</td>
<td>6.17*</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>230.32</td>
<td>5.05*</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>20.96</td>
<td>.46</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>45.60</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

<table>
<thead>
<tr>
<th>Experience</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-12.26</td>
<td>-9.68</td>
<td>-7.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variability</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-10.12</td>
<td>-11.77</td>
<td>-7.30</td>
</tr>
</tbody>
</table>
Table 12

Experience X Variability ANOVA on Deviation of Overall Ratings from True Scores for Experiment 2

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>9.36</td>
<td>5.96*</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>8.69</td>
<td>5.54*</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>.41</td>
<td>.26</td>
</tr>
<tr>
<td>Error</td>
<td>126</td>
<td>1.57</td>
<td></td>
</tr>
</tbody>
</table>

*p < .01

Experience

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.38</td>
<td>-.89</td>
<td>-.47</td>
</tr>
</tbody>
</table>

Variability

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.09</td>
<td>-.41</td>
<td>-1.23</td>
</tr>
</tbody>
</table>
Table 13

Experience X Variability X JDQ1 Regression on Deviation of Overall Ratings from True Ratings for Experiment 2

R²=.29  F(17,117)=2.76***

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience</td>
<td>2</td>
<td>6.78</td>
<td>4.72**</td>
</tr>
<tr>
<td>Variability</td>
<td>2</td>
<td>2.55</td>
<td>1.77</td>
</tr>
<tr>
<td>JDQ1</td>
<td>1</td>
<td>2.71</td>
<td>1.89</td>
</tr>
<tr>
<td>Experience X Variability</td>
<td>4</td>
<td>3.48</td>
<td>2.42</td>
</tr>
<tr>
<td>JDQ1 X Experience</td>
<td>2</td>
<td>2.69</td>
<td>1.87</td>
</tr>
<tr>
<td>JDQ1 X Variability</td>
<td>2</td>
<td>1.06</td>
<td>.74</td>
</tr>
<tr>
<td>JDQ1 X Experience X Variability</td>
<td>4</td>
<td>4.22</td>
<td>2.93*</td>
</tr>
<tr>
<td>Error</td>
<td>117</td>
<td>1.44</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05  
**p < .01  
***p < .001

Experience

<table>
<thead>
<tr>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.38</td>
<td>-.89</td>
<td>-.47</td>
</tr>
</tbody>
</table>
Table 14

**Simple Effects Analysis for High Level of Experience on Deviation of Overall Ratings from True Ratings for Experiment 2**

R2 = .31  \( F_{5,39} = 3.52^{**} \)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variability</td>
<td>2</td>
<td>6.89</td>
<td>4.24*</td>
</tr>
<tr>
<td>JDQ1</td>
<td>1</td>
<td>2.89</td>
<td>1.78</td>
</tr>
<tr>
<td>JDQ1 X Variability</td>
<td>2</td>
<td>6.03</td>
<td>3.71*</td>
</tr>
<tr>
<td>Error</td>
<td>39</td>
<td>1.63</td>
<td></td>
</tr>
</tbody>
</table>

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

JDQ1 and Deviation of Overall Ratings from True Ratings

<table>
<thead>
<tr>
<th>Variability</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>.46</td>
<td>-.42</td>
<td>-.51*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variability Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>-.85</td>
</tr>
</tbody>
</table>
Vita

Frank Allen Adair was born on April 7, 1960 in Shreveport, Louisiana, the son of Harry W. Adair and Dorothy Allen Adair. He graduated from Captain Shreve High School in 1978. In May of 1981, he graduated with a B.S. degree in Psychology from Louisiana State University in Shreveport. Fall of the same year he entered the Industrial/Organizational Psychology doctoral program at Louisiana State University in Baton Rouge. In December of 1983, he received his M.A. degree. He will receive his Ph.D. degree in Industrial/Organizational Psychology from Louisiana State University, Baton Rouge, Louisiana, in May, 1987.
Candidate: Frank A. Adair

Major Field: Psychology

Title of Dissertation: The Effects of Ratee Job Experience, Performance Variability, and Rater Cognitive Complexity on Performance Rating Accuracy

Date of Examination: April 27, 1987

EXAMINING COMMITTEE:

Irving M. Levine
Jiying Shi
Ray R. Buss
Gregory H. Blackburn
Gary J. Bennett