UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN EXAMINATION OF THE EFFECTS OF RATER TRAINING AND INDIVIDUAL PROTOTYPES ON PERFORMANCE RATINGS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

By

Sarah J. Cotten

Norman, Oklahoma

2004

AN EXAMINATION OF THE EFFECTS OF RATER TRAINING AND INDIVIDUAL PROTOTYPES ON PERFORMANCE RATINGS

A THESIS APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY: [redacted]

Chair ________________________
Dr. Robert Terry

Member [redacted]
Dr. Michael Buckley

Member [redacted]
Dr. Scott Gronlund

Acknowledgments

I would like to thank Dr. Michael Buckley, Dr. Robert Terry, and Dr. Scott Gronlund for their support and leadership through this wonderful learning experience. I could not have chosen a better group of individuals to study under or to look to for guidance. I also want to thank my parents and siblings for their unending love and encouragement. It has meant the world to me.

ABSTRACT

The performance appraisal process is often inaccurate due to various biases and psychometric errors. Rater training was developed in an attempt to familiarize raters with the dimensions used to judge performance and to standardize the appraisal process. It was expected that as a result of training, raters would develop a common frame-of-reference on which to rely during ratings. The goals of the present study were to validate the effectiveness of a common rater-training program, Performance Dimension Training, and to examine the effects of Performance Dimension Training on individual prototypes. Participants in the training condition were trained on the multidimensionality of professor performance and on the specific dimensions and behaviors that constitute effective performance. It was hypothesized that subjects in the training condition would produce more accurate performance ratings. It was also believed that individual's prototypes of professor performance would change to conform toward the "ideal" prototype presented as the standard during training. As a result individuals would rely on this prototype standard when completing performance ratings. This study was also concerned with under what conditions Performance Dimension Training is effective. Results showed a main effect for training on accuracy as training conditions were more accurate on performance ratings. A main effect was also found for the multi-tasking manipulation with subjects in the multi-tasking condition making less overall accurate performance ratings. Subjects in the training conditions did change their prototypes more so then did those in the control conditions, as hypothesized. Discussion and implications concerning these findings are presented.

An Examination of the Effects of Rater Training and Individual Prototypes on Performance Ratings

Everyone has experienced, at one time in his or her life, a performance evaluation. They range from simple performance observations to complex analyses of individual performance and organizational objectives. One of the most critical processes that occur in organizations is the performance evaluation. Research by Bernardin and Villanova (1986) suggests that in excess of 90% of organizations use some sort of performance appraisal in their human resources effort to measure/communicate the level of job performance of individuals within an organization.

Organizations use performance appraisals for a number of important human resource outcomes. They may be used for salary administration, providing performance feedback, employee development, promotion opportunities, transfer, or for determining employee/organization training needs. Because of the great impact performance appraisals have on organizational outcomes and on the employee who experiences the evaluation, it is essential that the appraisals accurately reflect the performance level of the focal individual.

Even so, a significant number of respondents view performance appraisal systems with little confidence and dissatisfaction (see Nigro's 1981 article for a disturbing view of employee perceptions of performance appraisal). In response to this disappointment, there have been numerous recommendations for rater training in order both to increase appraisal effectiveness, and to fortify the legal defensibility of the many human resource decisions that derive from appraisals (Bernardin & Buckley, 2000)

It has been suggested that an approach that incorporates cognitive variables may facilitate our understanding and implementation of training that facilitates accuracy in performance evaluation. Many have recommended the need to incorporate cognitive approaches into models of performance appraisals (Feldman, 1981, Landy & Farr, 1980, DeNisi & Cafferty, 1984). Research has shown that during the appraisal process, raters rely on cognitive structures to categorize and store information about ratees in memory (Feldman, 1981, Smith, 1986, Lord and Maher 1990). These structures guide the evaluation and processing of new information and also help the rater to ignore irrelevant information. Because humans are limited in their capacity to process information, understanding the influence and utility of these cognitive structures is essential to understanding performance evaluation in the multifaceted work environment.

Because raters rarely have opportunity to observe every behavior relevant to the job performance under review, it is extremely likely raters form global impressions of ratee efficiency (Smith, 1986). Potentially engulfing amounts of information are then simplified and reduced through cognitive categorization (Feldman, 1981). When rating performance from memory, raters have a tendency to use these abstract representations more frequently than specific behavioral instances. This may account for the decrement in accuracy when rating individuals from memory (Feldman, 1981, Lord, 1985).

Originally the emphasis on evaluating the evaluation process was focused on increasing rating accuracy. In his review of the performance appraisal process, Borman (1978) proposes a three-step model individuals typically follow when making judgments. First, the rater observes behaviors that are relevant to the job. Next, the rater makes an evaluation of each behavior, independent of other behaviors. Finally, each evaluation is weighted to arrive at a single rating for a performance dimension. Borman suggests that ratings should be effective if these three steps are followed. He highlights that doing so should increase rating accuracy. However, differences in final judgment ratings may be due to raters' stylistic differences/biases in the observation and evaluation process or due to the diversity of individual cognitive constructs used when making judgments concerning the performance levels of ratees.

Specifically, Borman (1978) emphasizes that raters are likely to vary in the weightings they assign to performance behaviors. One particular instance may be evaluated as effective by one rater and ineffective by another. Consequently, interrater reliability may often be compromised.

As the literature has shifted to a more cognitive approach to performance evaluations, much attention has been paid to Feldman's (1981) cognitive model of the appraisal process. He proposes that raters engage in the following series of cognitive processes before performance appraisals are possible: a) recognizing and attending to relevant information; b) organizing and storing information in memory for later access; c) recalling information in an organized method when ratings are required, and d) integrating information into a summary judgment. Feldman recognized that during performance appraisal ratings it is unlikely that managers are able to devote all of their attention to the appraisal process. He summarizes the various distractions and time-constraints as a "noisy environment."

According to Feldman's model, this behavioral information is organized into prototypes (Feldman, 1986). Prototypes are defined as abstract sets of features commonly associated with members of a category, with each feature assigned a weight according to the degree of association with the category (Cantor & Mischel, 1979, Rosch & Mervis, 1975, Smith & Medin 1981). These prototypes of ratee performance assist the rater by channeling his/her attention toward relevant information and away from irrelevant information (Anderson & Pickert, 1997). In the informationally noisy environment with which raters deal when rating subordinates, prototypes serve an important attention guiding function (Feldman, 1981).

Again, it is thought that raters store incoming and pre-existing behavioral information about ratees into categories to which the employee may be assigned. As the rater observes more performance incidents concerning the ratee, the incidents will become more comparable to one of the rater's categories or prototypes of job performance. During appraisal ratings, dispositional information regarding the employees is accumulated in memory based on these prototypes and serves as the basis of performance evaluations (Feldman, 1981).

Unfortunately, these prescribed steps in the performance appraisal process are not normally followed. Subjective ratings by subordinates, peers, and supervisors are frequently the ratings that are utilized in performance appraisals. Often this is due to lack of observational skills necessary to reach observational accuracy and lack of opportunity to observe all of the relevant behaviors. As a result, biases, psychometric errors, and inaccurate observations, deleteriously influence performance ratings (Landy & Farr, 1976). It has been hypothesized that by providing raters with accurate performance dimensions with which to organize new ratee information, we may likely enhance the efficiency with which information is encoded and stored in memory. As a result, raters are guided by the dimensions on the appraisal form, making employee information less difficult to retrieve.

In order to increase the accuracy and inter-rater reliability of performance ratings, Bernardin and Buckley (1981) developed Frame-of-Reference Training (FOR). They argue that raters hold implicit (and individual) personality theories regarding effective and ineffective performance. Therefore, the same performance may be judged as effective by one rater and ineffective by another. FOR provides raters with precise definitions of performance levels which aid in creating a common frame-of-reference for raters to refer to during appraisal ratings. The purpose of training focuses on standardizing the observation process through the use of familiarizing the raters before observing behavior with the performance criteria. Bernardin and Buckley proposed the following steps as standard FOR training: 1) giving trainees job descriptions and behavioral rationales for the job to be rated, and discussing the qualifications and duties essential to perform the job; 2) performing trial performance appraisals and writing out the rationale for their ratings; 3) having trainers feedback the correct rating based on expert scores and convey the expert's justification for the performance scores; and, 4) create a discussion that focuses on the inconsistencies between the true ratings and the trainees ratings. Research in the FOR domain has yielded both increases in rating accuracy and higher inter-rater reliability among ratings, especially for those raters who held idiosyncratic rating standards (Bernardin & Buckley, 1981).

McIntyre and Smith (1984) specifically examined the effects of training format on rating accuracy and retention of training material. The training factor comprised four levels: rater error training, frame-of-reference training, a combination of both, and no training. Rater error training consisted of lecture that encouraged raters to avoid making psychometric errors such as halo, leniency, and central tendency errors. A short discussion followed the lecture. Frame-of-reference training aimed to aid subjects in developing standards for effective performance that were comparable to those of expert raters. Raters in this condition received lecture, practice and feedback with ratings, and behavioral rationales for ratings. Results provided evidence that FOR did produce more accurate ratings compared to the rater error training and no training conditions.

Performance Dimension Training (PDT), an implicit part of FOR, has also been shown to improve the accuracy of performance ratings. PDT utilizes a training design where trainees discuss the multidimensionality of performance and the need to distinguish performance dimensions, as well as levels of performance (Athey & McIntyre, 1987, Bernardin & Pence, 1980, McIntyre, 1984, Pulakos, 1984). Rationales for effective performance behaviors under each dimension to be rated are thoroughly discussed during training and raters are given opportunity to examine the rating scale used in the appraisal process.

Just as with FOR, by defining performance dimensions early in the rating process, ratees are better able to attend to the dimensions and refrain from making (and categorizing) global impressions of ratee behavior (Smith, 1986). By emphasizing these specific behaviors and dimensions used to measure performance, raters make more independent evaluations, thus rater effectiveness is increased. The central component of this model is categorization, serving as a connection to the other processes. The categories guide the raters attention toward particular behavioral instances while suggesting that they fundamentally ignore others (Feldman, 1981).

Illgen and Feldman (1983) emphasized the importance of job relevant categories in the rater training process. The researchers believe that the more training is focused on creating a category system to assist in attention, storage, and recall of behavioral incidents, the more accurate the ratings that the resulting performance appraisals yield.

Pulakos (1984) specifically examined the effects of PDT. Participants in her study received a lecture on the multidimensionality of jobs. Next they were presented with the rating scale format. Specific behaviors indicative of effectiveness levels for each dimension were discussed. Subjects then practiced rating videotapes of performance, and any discrepancies between their own ratings and expert ratings were provided as feedback. Results of the study suggest training produces greater accuracy in performance ratings compared to control conditions.

Smith (1984) also investigated the effects of PDT compared to a no training condition before completing appraisal ratings. Subjects rated videotaped lectures of supposedly real college professors after undergoing training or no training. The training format consisted of lecture on principles of observation, viewing a list of behaviors and rating dimensions to observe, and finally behavioral rationales for effective and ineffective performance. Results show that compared to the control condition, PDT produced greater rating accuracy.

One of the major criticisms of the performance appraisal rater training literature has been the lack of any underlying theoretical structure to guide theory and practice (Smith, 1986). Feldman's (1986) research in the cognitive domain of rater training has suggested that rater training provides raters with uniform and valid schema and prototypes regarding the work behaviors that are to be observed.

Presenting the raters with the performance factors prior to ratings, as observed in FOR and PDT, facilitates the utilization of relevant prototypes of performance. Many have argued that raters use these abstract representations of explicit behavioral information when rating performance from memory (Feldman, 1981, Lord et al, 1985). As a result, the creation of prototypes for raters, this should maximize the likelihood that raters will correctly categorize ratee behavior on each dimension on the basis of observed performance (Hauenstein & Foti, 1989). Accurate ratings on performance features for each dimension should occur due to the more efficient categorization of observed behavior.

Various researchers assert that such categorization would add to rating accuracy even if raters did not recall specific performance behaviors to direct their ratings (DeNisi et al, 1984, Feldman, 1981). Sulsky and Day (1992) specifically examined cognitive changes related to FOR training to better comprehend why such rater training commonly improves rating accuracy with the effects of rater training on individual prototypes of performance. They suggest that Frame-of-Reference training allows subjects to form on-line judgments that are based on prototypes of various levels of performance on the performance dimensions. They found that when subjects in the training condition had the goal of forming an impression about the ratee and based this on the prototypes of performance, there were increases in both rating and categorical accuracies.

As mentioned earlier, another concern during performance appraisal ratings is the diminished likelihood that managers will experience optimal conditions in the work environment allowing them to devote all their attention to the appraisal process. Multi-tasking and interruptions that create a "noisy work environment"(Feldman, 1981) serve to divide the rater's attention. This disruption can result in cognitive overload because the individual's attention is diverted from some aspects of his or her environment in order to focus on other aspects of the stimulus environment (White & Carlston, 1983). It has been suggested that when an individual's attention is diverted away from key tasks, he or she is more likely to rely on past/individual prototypes to make judgments.

Bargh and Theim (1985) suggest individuals are able to develop impressions and engage in elaborative processing if they possess cognitive structures appropriate to relevant information that is observed. Therefore, in the midst of distractions, rater training should serve as an essential component of the appraisal process and allow for smoother retrieval of specific dimensions and behaviors observed in the work place.

Objectives of Present Study

In this study, Performance Dimension Training (PDT) was investigated as a method to increase the usefulness of performance appraisal systems. As previous research has shown, PDT embodies many of the separate elements employed in performance appraisal training programs that have been demonstrated to increase rating accuracy. Specifically, Performance Dimension Training was investigated as a means to increase accuracy and interrater reliability in performance appraisal ratings.

Hypothesis 1a: It was hypothesized that Performance Dimension Training conditions would produce more accurate performance ratings compared to ratings given by respondents in the conditions that received no training.

Specifically, participants will be more likely to attend to behaviors corresponding to the performance dimensions learned in training and interpret those behaviors according to the dimensions.

Hypothesis 1b: It was hypothesized that participants in the Performance Dimension Training only condition would produce more accurate ratings than participants in the Performance Dimension Training/multi-tasking condition, the no training/multi-tasking condition, or the no training only conditions.

This study also examined the effects of rater training on individuals' existing prototypes of job performance. It was thought that participants pre-existing prototypes would affect subsequent performance ratings. It was suggested that undergoing PDT would influence the categorization of incoming performance information; thus participants would alter their existing prototypes to match the dimensions and behaviors presented during training that were shown to constitute effective and ineffective performance.

Hypothesis 2: It was hypothesized that participants in the Performance Dimension Training condition would alter their pre-existing prototypes of professor performance to conform to the prototype of professor performance presented during training more than participants in the no training condition.

If training does increase the accuracy of performance ratings, it was suggested that altering participants' pre-existing prototypes of professor performance to conform to the ideal standard would be the mediating variable responsible for increases in accuracy scores. Participants should rely on this newly formed prototype when completing appraisal ratings.

Hypothesis 3: It was hypothesized that prototype change would serve as the mediating mechanism by which accuracy scores are influenced by Performance Dimension Training.

Finally, it was thought that certain individual differences such as Need for Cognition, Preference for Consistency, and Personal Need for Structure, may contribute to participants resistance to changing their pre-existing prototypes and therefore impact training. It was specifically believed that those individuals who generally preferred consistency and were uncomfortable with a lack of structure would be less likely to accept the non-prototypic information presented during training.

Hypothesis 4: It was hypothesized that the individual differences measures would mediate the relationship between prototype change and the effectiveness of training.

Method

Pilot Study

A pilot study was conducted using 130 students enrolled in an undergraduate business course to identify the behaviors presented in the study training materials. Students were asked to record as many effective and ineffective behaviors exhibited by professors that they could imagine. A frequency count was conducted on the data and five dimensions, four of which were based on prior research by Rugg and Norris (1975) were identified: Interpersonal Rapport, Communication Skills, Student Involvement, Lecture Presentation, and Teaching Skills. The most frequently listed effective and ineffective behaviors for each dimension were noted. The researcher disregarded idiosyncratic behaviors based on lack of consensus.

Professor behaviors that were listed as both effective and ineffective by the pilot group were used in the study. At least one behavior that was considered ineffective performance from each training dimension was manipulated to represent an *effective* behavior that was not prototypic (based on pilot data) to better examine the effectiveness of Performance Dimension Training. The experimenter rationalized the ineffective professor behaviors into effective professor behaviors, claiming the findings were based on educational research, although no such actual research existed. This deception was used in order to identify changes in participants' prototypes. It was assumed that if rater training was effective, participants would change their pre-existing prototypes (post-training) to conform to the information presented during training even when the information contradicted what pilot data suggested was viewed as typical ineffective professor performance.

Stimulus Materials

Control Group Video

Videotaped segments of professor performance were used as stimulus material for the control group. Three professors at the university in the Psychology department served as examples of typical professor performance for the videotape. Each of the three professors (one male, two females) was shown lecturing for approximately five minutes. Topics included Attention Deficit Hyperactivity Disorder, descriptive statistics, and social psychology. Participants in the no training conditions/control group viewed the fifteen-minute segment instead of exposure to rater training. Participants were told that the video simply served to demonstrate real-life examples of professors' lecturing styles that were neither good nor bad.

Vignettes

Vignettes were developed for the rating exercises (all vignettes were approximately one page in length). The vignettes are short scenarios of critical incidents concerning the five dimensions of professor performance used in rater training (See Appendix A). The scenarios were scripted to portray effective and ineffective teaching behaviors utilized by college professors. Six vignettes were specifically developed for evaluation (i.e. Professor "H," "J," "K," "L," "M," and "Z"). All participants rated three vignettes during session one and three vignettes during session two. Subjects rated the behaviors of the professors in the vignettes on a scale of one to five during the rating task. A score of one was indicative of ineffective performance for a given dimension and a score of five was indicative of effective performance. Through various behavioral examples, each vignette was written to illustrate each of the five dimensions described in training. All participants rated the vignettes in the same order throughout the study.

Graduate students in the Industrial/Organizational Psychology program served as subject matter expert raters for determining the true scores or performance standards of the vignettes. These standards served as the accuracy criteria during the rating procedures. Five graduate students individually were given Performance Dimension Training and then asked to individually rate the vignettes according to the training criteria. Each expert was given the same instructions and materials. Expert consensus reliability of the subject matter experts was estimated using generalizability theory to document the degree of interrater reliability. The resulting g-coefficient (g=.74) was interpreted as a satisfactory indicator of subject matter consensus.

Measures

Prototype Assessment

Participants were asked to rate eleven behaviors of professor performance that correspond to the five training dimensions (See Appendix B). These ratings served as each participant's individual prototype of typical effective professor performance. Participants rated the behaviors on a scale of one to five. A rating of one indicated that the behavior was not very effective professor behavior and a rating of five indicated the behavior was very effective professor behavior. The behaviors in the prototype assessment were based on pilot data concerning students' opinions about what constituted effective and ineffective professor behavior, as well as consensus among subject matter experts.

Regardless of the opinions of pilot subjects, all behaviors included in the prototype assessment were presented as *effective* teaching behavior for the purposes of this study. For example, pilot data revealed that students view professors as ineffective when they move through lecture material quickly. As a part of training, this behavior was rationalized to be an *effective* component of professor performance. Participants in training were told that research supports the finding that professors who move through lecture material quickly are actually viewed as extremely effective teachers, based on subsequent exam performance. Supplemental information and reasoned rationales were presented to support the idea. In this particular instance, participants were told that moving through lecture material quickly allows for more material to be covered and learned during the semester, and that a fast paced lecture encourages students to stay focused; thus retaining more information.

This component was essential to examine the effects of training on individual prototypes after exposure to rater training. In this study, it is believed that if training material contradicts subjects initial professor prototypes (based on pilot data) subjects should change their preexisting professor prototypes to conform to the prototype presented during training (if training is effective).

Supplemental Covariate Measures

We included three individual covariate measures with the idea that certain personality characteristics may make people more resistant to changing their prototypes of professor performance.

Need for Cognition Scale

The Need for Cognition Scale was used as a covariate. Participants were asked to respond to statements about their tendencies to engage in and enjoy effortful cognitive endeavors (Cacioppo et al, 1996). The scale consists of 18 questions. Sample scale items include: 'I like to have the responsibility of handling a situation that requires a lot of thinking' and 'I would rather do something that requires little thought than something that is sure to challenge my thinking abilities'.

Need for Structure Scale

The Need for Structure Scale was used as a covariate measure. Participants were asked to respond to questions about their desires to structure cognitive and social behaviors. The scale consists of 12 questions such as 'I'm not bothered by things that interrupt my daily routine,' and 'I find a well-ordered life with regular hours makes my life tedious' (Neuberg & Newsom, 1993).

Preference for Consistency Scale

The Preference for Consistency Scale was used as a covariate measure. Participants were to respond to statements concerning their desires to be consistent within one's own responses, in the desire to appear consistent to others, and in the desire that others be consistent (Cialdini & Trost, 1995). The scale consists of 18 statements such as 'It is important to me that those who know me can predict what I will do' and 'The appearance of consistency is an important part of the image I present to the world.'

Professor Rating Scale

The rating scale used in this study consisted of five questions concerning the five performance dimensions emphasized during training (See Appendix C for rating scale and rating sheets). The five performance factors were identified as relevant to the position of professor. This scale required participants to assign a rating from one to five to each question. A score of one indicated ineffective professor performance and a score of five indicated effective professor performance for the dimensions.

Two questions concerning a multi-tasking activity were included on the rating scale for participants in the multi-tasking group only. The first question required participants to indicate how many sequences of three or more odd digits they observed during the multi-tasking activity (See Procedure Section for multi-tasking activity) and the second question asked them to indicate on a scale of one to five how distracted they were during the activity (a score of one indicated not distracted and a score of five indicated very distracted).

Study Participation

Participants were undergraduate students enrolled in introductory psychology courses at the University of Oklahoma. They received experimental credit in exchange for their participation. A total of two hundred and twenty respondents participated in this experiment. The design employed was a longitudinal design and not all subjects returned for session two of the study. A total of one hundred and eighty-four participants completed both sessions.

Procedures

Participants were randomly assigned to one of four experimental conditions; the PDT/multi-tasking group, the PDT only group, the no training/no multi-tasking group, or the no training/multi-tasking group. The experimenter gave the participants a brief overview of the purpose of the study.

During Session I, all participants completed the prototype assessment. Participants were asked to rate in their own opinion how typically effective they considered the eleven specific professor behaviors. The behaviors the participants rated were the behaviors used in the Performance Dimension Training.

After the initial prototype assessment, participants were exposed to one of two training conditions: Performance Dimension Training (PDT) or the control video. Training sessions were conducted in groups of up to fifty-five students. One trainer conducted all of the training sessions, which lasted approximately 60 minutes.

All participants were told that the purpose of the study was to examine how individuals assessed the performance of professors. Participants were then told that they would be rating scenarios of real first-year professors at the university. It was explained that the ratings could possibly be taken into consideration during the professors' first-year evaluations. This point was reinforced throughout the course of the study.

The researcher in the Performance Dimension Training groups first explained to the subjects the multidimensionality of professor performance. Next, participants were given the opportunity to observe the rating scale they used to rate the vignettes of professor performance post-training. Finally, in lecture format, the experimenter presented the dimensions and behaviors that constituted effective professor performance. The researcher read the dimensions and the corresponding behaviors from a master copy to ensure standardization of the procedure (See Appendix D). Participants in the training condition received a copy of the training to follow along with the experimenter. It was emphasized that the effective professor behaviors presented in training were validated by current research. In reality, the citations for the behaviors were fictional and the behavioral rationales for the behaviors were not validated by research. This component of the study was necessary to convince the participants that training was legitimate and to better understand expected changes in participants' pre-existing prototypes of professor performance after training. The ultimate focus of training was to create a common frame of reference which the participants could rely upon when completing performance ratings. Training lasted approximately 20 minutes.

Participants not receiving training watched the control video. Control subjects were told that they were watching the videotape to observe typical teaching styles of college professors. Participants were not asked to answer any questions concerning the videotape.

Next, participants in both experimental groups completed the covariate battery which contained the aforementioned three individual differences measures. Subjects received a packet of all three assessments and were informed that they would have 12 minutes to complete all three measures.

Following the covariate battery, all participants rated three vignettes of professor performance: Professor "H", Professor "K" and Professor "Z". The rating task took approximately 15 minutes. Participants were given two and a half minutes to read the first vignette. Overhead transparencies were used during the rating process to ensure standardization of all rating procedures in the study. Upon beginning the ratings, participants were told that the questions for rating professor behavior would be shown individually on an overhead projector by the experimenter, and were told to answer them in the corresponding blank on the rating sheet. Next, the experimenter placed the first question from the rating scale on an overhead slide. The participants were given 20 seconds to answer the question. After 20 seconds, the experimenter placed the second question from the rating scale on an overhead slide. Again, participants were given 20 seconds to answer the question. The experimenter continued with this process for all five questions on the rating scale. The rating process for each vignette lasted one minute and forty seconds. This method of presentation was used in order to control the pace of the rating process for each condition.

Half of the participants in the study received a multi-tasking activity during the rating process. The participants in the multi-tasking conditions rated the vignettes in the exact format as did the participants in the no multi-tasking conditions. However, as the participants in the multi-tasking conditions were rating the vignettes, a cassette tape of pre-recorded numbers played in the background. The participants heard the experimenter reciting a single-digit even or odd number (one through nine) every 5 seconds. The participants were instructed to note how many times three or more odd digit numbers were recited in a row during the duration of the tape (See Appendix E for rating instructions). After participants read each vignette, the tape began as the first question on the rating scale was presented to the participants on an overhead slide. Again, on the bottom of each participant's rating sheet, there was a blank space for the participants to write the total number of three or more odd digit sequences they heard during that particular rating. The above procedure was repeated for each subsequent vignette rating. Because half of the participants in the study experienced a multi-tasking condition during the rating process, this procedure ensured that each participant received the multi- tasking procedure for the same amount of time while completing the performance ratings.

After all five questions were answered for the first vignette, subjects were given two and a half minutes to read the second vignette. Subjects answered the same five questions for the second vignette and were given 20 seconds to answer each one. Subjects followed the same procedure for the third vignette.

Upon completion of Session I all participants completed the prototype assessment form a second time. The Prototype Assessment was a duplicate of the prototype assessment form participants filled out at the onset of the experiment.

All participants returned one week later on the same day and time to participate in Session II. All participants completed the prototype assessment form for a third and final time. All participants then rated three different vignettes of professor performance: Professor "J", Professor "L" and Professor "M." Those participants in the multi-tasking condition received the multi-tasking activity again as they rated the final three vignettes.

For debriefing, all participants were given a debriefing form to read after completing Session II. They were given the opportunity to ask any questions of the experimenter regarding the purpose of the experiment. The experimenter answered all and any questions regarding participation credit and confidentiality.

Independent Variables

Rater Training

Performance Dimension Training

Performance Dimension Training procedures followed those adapted by Pulakos (1984; 1986). Procedures included lecture concerning training dimensions and rationale for the "ideal" standards. Half of the participants in the study received Performance Dimension Training and half of the participants watched a control video that simply served as stimulus material.

Dependent Variables

Accuracy

Accuracy of participant ratings were assessed by calculating the Squared Euclidean Distance of each participant's rating from the average of the ratings assigned by the subject matter experts.

Each Squared Euclidean Distance for a dimension was then averaged across the five dimensions. The minimum average Euclidean distance score of "0" would be indicative of perfect accuracy, with large values indicating less accurate ratings.

Prototype Change

The change in each subject's prototypes concerning the five dimensions were identified post-study. Discrepancies between initial prototypes and any change in individual prototypes after training were examined as a potential mediator of the effectiveness of Performance Dimension Training on subsequent rating accuracy. Prototype ratings closer to 5 indicate extreme closeness to the "ideal" professor prototype.

Analysis

All data were analyzed using the statistical package SAS, version 8.3. Generalizability theory was used to assess Expert consensus, while all Analyses of Variance (ANOVA) and Analyses of Covariance (ANCOVA) were conducted using SAS Proc GLM.

Results

Two hundred and twenty-five participants initially participated in Session I of the experiment. Only one hundred and eighty-four of the original participants who completed both sessions of the experiment were used to test the study hypotheses.

Prototype Effects

A one-factor ANOVA was used to assess the result of the randomization process to groups regarding the distribution of participants' individual prototypes at Time 1. Table 1 contains means and standard deviations of prototype ratings by training conditions across time. Results indicated no significant differences between the four conditions for any of the eleven prototype dimensions for the initial prototype assessment: 1) Move through material quickly, F (3,180)=.61, p=.607; 2) Has a sense of humor, F (3,180)=.92, p=.444; 3)Remains motionless while teaching, F (3,180)=1.79, p=.151; 4) Is organized, F (3,180)=1.17, p=.321; 5) Uses visual aids during lecture, F (3,180)=.69, p=.562; 6) Requires class participation, F (3,180)=.17, p=.914; 7) Uses examples to support class material, F (3,180)=.78, p=.504; 8) Displays an attitude of superiority, F (3,180)=1.35, p=.261; 9) Speaks loudly during lecture, F (3,180)=.66, p=.576; 10) Assigns group work and group projects, F (3,180)=1.14. p=.243; and 11) Does not exhibit defined structure for course work or lectures, F (3,180)= .69, p=.562.

Analyses were conducted to demonstrate changes in participants' prototypes from the initial prototype assessment due to training and to demonstrate the persistence of the change over time. A Repeated Measures Mixed-Model ANOVA (Time x Training Group) was used to evaluate changes in prototypes across session as a function of training. The Wilk's Lambda indicated significant time by training group effects for the following prototype dimensions: 1) Moves through material quickly, $F(2, 179)= 17.66$, $p<.0001$; 2) Remains motionless while teaching, $F(2, 179)= 14.18$, $p<.0001$; 3) Is organized, $F(2, 179)= 8.93$, $p=.0002$; 4) Uses visual aids during lecture, $F(2, 179)=3.01$, $p=.0516$; 5) Displays an attitude of superiority, $F(2, 179)= 19.97$, $p<.0001$; 6) Speaks loudly during lecture, $F(2, 179)=12.41$, $p<.0001$; and 7) Does not exhibit structure, $F(2, 179)=26.73$, $p>.0001$. See Figure 1 for mean prototype ratings by training condition across time.

Of greater interest were the effects of training on prototype change from Time 1 (pre-training assessment) to Time 2 (immediate post-training) and on prototype change from Time 2 to Time 3 (1-week post-training follow-up). We hypothesized that significant changes would occur in individual's prototype ratings between Time 1 and Time 2 due to training, and that newly acquired prototypes should remain stable from Time 2 to Time 3 (one week follow-up assessment). All prototype changes over adjacent time points were assessed using the single degree-of-freedom contrast method.

Results indicated significantly different changes in prototype ratings by training condition from Time 1 (initial assessment) to Time 2 (immediate post-training assessment) for the following prototype dimensions: 1) Moves through material quickly, $F(1,180)=35.32$, $p<.0001$; 2) Remains motionless while teaching, $F(1,180)=23.12$,

$p<.0001$; 3) Is organized, $F(1,180)= 11.99$, $p<.001$; 4) Uses visual aids during lecture, $F(1,180)=5.48$, $p=.02$; 5) Displays an attitude of superiority, $F(1,180)=38.66$, $p<.0001$; 6) Speaks loudly, $F(1,180)=24.23$, $p<.0001$; and 7) Does not exhibit structure, $F(1,180)=41.86$, $p<.0001$. Results for Time 2 (immediate post-training assessment) to Time 3 (one-week follow-up assessment) follow-up analyses of prototype change by training condition indicate no significant differences from Time 2 to Time 3 ratings by group for any of the prototype dimensions, as hypothesized. It was expected that the change in prototype would continue over time. The results suggest prototype changes from immediate post-training to the one-week follow-assessment among all conditions remained stable over time.

For subsequent analysis we created a composite index of prototype change, which measures the distance of each participant's prototype rating from the "ideal" professor prototype presented during training. For the purpose of this study, we suggest the "ideal" instructor would be given a rating of "5" post-training, for each of the eleven dimensions on the prototype assessment. All of the dimensions we presented in training were portrayed by fictitious research to be the most effective teaching behaviors. Therefore, a rating of "5" for each dimension signifies the overall "ideal." The participant's distance from this rating of "5" is what was examined in terms of effectiveness of training. Participants should in essence, move toward ratings of "5" for each dimension if training was successful. For each time point we took the participants' Squared Euclidean Distance measure (squared distance) from the ideal prototype, and averaged these across all eleven dimensions. The resulting index indicated the averaged squared distance per dimension from which a participant's actual prototype falls from the ideal prototype.

Results indicated no statistical difference in group means at Time 1 prototype assessment (pre-training assessment). Group means varied considerably at Time 2 (post-training assessment) and Time 3 (one week follow-up assessment) with the prototype assessment for the training group much closer to the ideal prototype. See Table 2 for means and standard deviations of aggregated prototype ratings by training condition.

Participant Drop-out Analysis

An analysis was conducted to test for a potential association between participant dropout rates and training condition. It is possible that participants who did not return for Session II may have been a problematic group or were possibly overwhelmed by the activities in Session I and chose not to come back. A Chi-square test of independence indicated that dropout rates did not vary by group (X^2 (1)= .5315, p=.460). It appears that dropout rates were not related to the presence or absence of training condition; the dropout rate appeared to be random.

Accuracy of Performance Ratings

Accuracy scores were computed using a Euclidean (squared) distance procedure. For a specific vignette, the Euclidean distance of each participant's ratings was calculated for each of the five dimensions (Teaching Skills, Lecture Presentation, Interpersonal Rapport, Communication Skills, and Student Involvement) using the average rating of the subject matter experts (SME's) as the standard for that dimension. Within a specific vignette, the resulting five Squared Euclidean Distance scores were then averaged to create a vignette-specific accuracy measure. Finally, the three vignette-specific accuracy measures were averaged to create the final dependent variable for analysis for Session I. The same procedure was also used to obtain averaged accuracy scores for the three vignettes in Session II. A score of "0" on the aggregate accuracy measure indicates perfect accuracy. Larger values indicate more inaccurate scores or more distance from SME ratings. Results indicated training groups were more accurate on Time 1(Session I) and Time 2 (Session II) performance ratings than control groups. See Table 3 for means and standard deviations of accuracy scores by training condition.

A Repeated Measures Mixed Model ANOVA (Time x Training x Multi-Tasking) was used to assess the effects of time, training, and multi-tasking on participants rating accuracy. All, higher-order effects (all two-way and three-way interactions) plus main effects were evaluated in this analysis.

Analyses of Higher-Order Effects

Time x Training x Multi-tasking Effect

The results of the test of the three-way interaction were non-significant, $F(1, 180)=0.00$, $p= .996$. This finding suggests the combined effects of Training and Multi-tasking on accuracy scores did not vary over time.

Time by Multi-Tasking

The test of the two-way interaction of Time and Multi-tasking was assessed and resulted in a non-statistically significant effect, $F(1,180)=.34$, $p=.563$. This indicates that any changes in accuracy over time were not moderated by the presence or absence of a Multi-tasking condition.

Time by Training

No effects concerning the combination of Training Group and Time were found with regard to accuracy scores, $F(1,180)=.43$, $p=.511$. Results indicated that time of assessment does not moderate any potential changes in accuracy ratings due to Performance Dimension Training.

Training by Multi-tasking effect

It was tested whether or not specific effects Multi-tasking moderated the effects of Training on accuracy scores. Results indicate no two-way interaction by Training and Multi-tasking on accuracy scores, $F(1, 180) = 1.22$, $p < .27$.

Main Effects Analyses

Training Group Effect

Significant differences were found between Training vs. No Training groups on accuracy scores $F(1, 180) = 39.97$, $p < .0001$, (Training $M = 1.48$, $SD = .489$; No Training $M = 1.99$, $SD = .604$). Recall that lower accuracy scores indicate greater accuracy or less distance from the ideal standard. Subjects in the Training condition were significantly more accurate in their professor ratings than were subjects in the No Training condition.

Multi-Tasking Effects

Significant differences were also found between Multi-tasking groups on accuracy scores, $F(1, 180) = 4.94$, $p = .0275$, (No-Multi-Tasking $M = 1.66$, $SD = .460$; Multi-tasking $M = 1.83$, $SD = .707$) with subjects in the Multi-tasking condition producing significantly less accurate ratings than those in the no Multi-tasking conditions as expected. See Table 4 for means and standard deviations of accuracy scores by multi-tasking condition.

Time Effect

The test of the hypothesis of no within-subject effect of time indicated a significant effect for being more accurate at Time 2 than Time 1 (Time 1 *M*=1.88, Time 2 *M*=1.60). Apparently, subjects' accuracy ratings improved over time rather than declining, as was hypothesized.

Mediational Analyses

Immediate Post-training Assessment

Regression analysis was used to test the hypothesis that prototype change would mediate training group improvement on accuracy ratings at the immediate post-training assessment. Baron and Kenny (1986) propose three steps necessary to establish mediation. First, a relationship must be established between the initial variable (training condition) with the outcome variable (performance rating accuracy). Previously, it has already been demonstrated that training condition and accuracy scores are related. Using rating accuracy as the outcome variable, the regression slope for the effects of training condition is $\beta = -.55$, $SE = (.0903)$, $p < .0001$, $t = -6.14$. This slope indicates the mean difference between training group and non-training group on accuracy ratings at the initial post-training assessment.

Second, a relationship between training condition and prototype assessment must be established, essentially treating the potential mediator (prototype assessment) as an outcome variable. Rather than use a direct measure of change as the mediator, the Time 2 (immediate post-training) prototype assessment was used in its place. The rationale for this as follows; Since all participants were randomized to groups and there were no statistically significant differences between any group on their initial prototype assessments, it is more meaningful to use the prototype score at Time 2 (immediate post-training) as the mediator rather than a direct measure of change. Subsequently, a strong correlation was shown between prototype assessment and training condition at initial post-training assessment ($r = -.596$, $p < .0001$), and the regression slope was estimated as $\beta = -1.76$, $SE = (.1752)$, $p < .0001$, $t = -10.03$. In terms of Squared Euclidean distance, this suggests that the training groups' prototype was 1.75 squared distance units closer to the ideal prototype than the control group. Recall that mean Squared Euclidean Distance scores closer to "0" are more concordant with the ideal prototype.

Finally, a relationship between prototype assessment and performance rating accuracy, controlling for training, must be established. Furthermore, a reduction in the slope of the training effect should be apparent if prototype change was a mediator. To establish this, a regression was undertaken using rating accuracy as the dependent variable and prototype assessment and training condition as independent variables. Results suggest that prototype assessment and performance rating accuracy are related when controlling for training condition effects, $\beta = .074$, $SE = (.0379)$, $t = 1.95$, $p < .052$. As a test of partial mediation, the Sobel test for Time 1 (immediate post-training assessment) indicated a partial mediation effect; $t = -1.92$, $p < .054$. Regression analysis indicated the original mean difference on Time 1 accuracy scores to be $\beta = -.55$. Controlling for prototype assessment post-training, the original mean difference for initial post-training accuracy decreased to $\beta = -.42$, $SE = (.1118)$, $t = -3.80$, $p < .0001$. This reduction in the mean difference of 16% suggests that prototype change partially mediated the relationship between training condition and accuracy ratings. This last result, combined with the previous two steps, fulfill Baron and Kenny's (1986) three-step technique, which suggest prototype assessment is a partial mediator between training and performance rating accuracy.

One-week Follow-up Assessment

The Baron and Kenny (1986) method was also followed to test the hypothesis that prototype change would mediate training group improvement on accuracy ratings at the one-week follow-up assessment. The first step, previously established, was to demonstrate the relationship between training condition and performance rating accuracy at the one week follow-up assessment, $\beta = -.463$, $SE = (.1191)$, $p < .0001$, $t = -3.89$. Again, this slope indicates the mean difference between training condition and non-training condition on accuracy ratings at the initial post-training assessment.

Second, a relationship between training condition and prototype assessment must be established. This relationship had already been established at the immediate post-training assessment. ($r = -.596$, $p < .0001$), and the regression slope was estimated as $\beta = -1.76$, $SE = (.1752)$, $p < .0001$, $t = -10.03$. Again, in terms of Squared Euclidean Distance, this suggests that the training groups' prototype was 1.75 squared distance units closer to the ideal prototype than the control group. Recall, the Time 2 (immediate post-training) prototype assessment was used rather than a direct measure of change.

Finally, controlling for training, a relationship between prototype assessment (immediate post-training) and performance rating accuracy at the one week follow-up assessment must be established. Again, using rating accuracy as the dependent variable and prototype assessment and training group as independent variables, a regression was undertaken. Results indicate when controlling for training group effects, prototype assessment and performance rating accuracy are related, $\beta = .128$, $SE = (.0496)$, $p < .01$,

$t=2.59$. Furthermore, the adjusted for prototype regression slope relating training condition to rating accuracy at the one week follow-up was equal to $\beta= -.238$, $SE=(.1462)$, $t= -1.63$, $p=.1053$. When controlling for prototype change at the one week follow-up assessment, the accuracy mean difference was $\beta= -.238$, $SE=(.146)$, $t= -1.63$, $p<.105$. The original mean difference not controlling for prototype accuracy was $\beta= -4639$, $SE=(.1191)$, $p< .0001$, $t= -3.89$ indicating a 51% reduction in the original mean difference. The Sobel test at the one week follow-up assessment indicated a partially mediated effect; $t= -2.51$, $p<.01$.

Overall, mediational analyses demonstrated that prototype change is a partial mediator responsible for increases in accuracy ratings. Prototype change was an even stronger mediator at the one-week follow-up assessment. It is possible that training increases awareness of the ideal prototype presented in training; thus encouraging participants to alter their pre-existing prototype to conform to the ideal. Participants may then rely on this newly formed prototype when completing accuracy ratings with prototype change partially responsible for the increase accuracy.

Individual Differences Assessment

One further concern was that some individuals may be more resistant to prototype change because of individual personality differences. We collected three covariate measures (Need for Cognition Scale, Preference for Consistency Scale, and Personal Need for Structure Scale) and controlled for those in a regression model to see if the addition of the covariates changed the regression slope. Results indicated it did not. Specifically, regression analysis, controlling for prototype assessment indicated the adjusted mean difference at Time 1 accuracy (immediate post-training) to be $\beta= -.424$, SE= (.1118), t=-3.80, p<.0001. The adjusted mean difference at Time 1 accuracy also controlling for these three individual difference variables was β= -.454, SE= (.111), t= -4.05, p<.0001, indicating the three measures do not seem to influence the relationship between prototype change and accuracy scores. The same conclusion was found for accuracy at Time 2 (one week follow-up assessment).

Essentially, our results imply an additive model of training condition and multi-tasking condition on rating accuracy. Results demonstrate that training was effective and our multi-tasking condition, although not interacting with training, also mattered. Surprisingly, accuracy scores for all groups slightly increased with time. Finally, results suggest prototype mediated the effects of training on accuracy and the mediational effects were stronger at the one week follow-up assessment than at the immediate post-training assessment.

Discussion

The main objectives of the present study were to investigate Performance Dimension Training as a method to increase the usefulness of organizations' performance appraisal systems and to examine the effects of rater training on individual's existing prototypes of performance. It was specifically hypothesized (Hypothesis 1a) that undergoing Performance Dimension Training would result in more accurate performance ratings. The results of the present study provide support for this hypothesis. Training conditions did indeed produce more accurate vignette ratings for both experimental sessions than conditions that received no training. This finding is consistent with previous research utilizing similar Performance Dimension Training design (Athey & McIntyre, 1987, Bernardin & Pence, 1980, McIntyre, 1984, Pulakos, 1984).

Data from the study also provided information concerning the effects of combining Performance Dimension Training with a multi-tasking activity. It was believed that participants who received training and the multi-tasking activity would still produce more accurate ratings than those in the no-training conditions. Although no interaction between training and multi-tasking conditions were observed, an overall multi-tasking effect was found. Groups produced more accurate ratings when under no multi-tasking manipulation regardless of training condition.

However, rank ordering of group means for accuracy ratings demonstrated more accurate ratings for those participants who received both the multi-tasking manipulation and training as hypothesized (Hypothesis 1b) than those who received multi-tasking only. This intuitively appealing idea suggests that Performance Dimension Training is a useful tool for increasing the accuracy of performance ratings even in the midst of competing tasks. As suggested by Feldman (1981), multi-tasking and interruptions in organizational settings create a "noisy work environment" that divides rater's attention (p See page 13). This cognitive overload diverts the rater's attention from important aspects of the environment subjecting the performance ratings to error. Relying on the performance standards presented during training helps guide the raters attention toward relevant behaviors and to correctly categorize incoming behavioral information (Haunstein & Foti, 1989).

It is interesting to note, however, that when participants were asked to rate how distracted they were when completing the ratings and the multi-tasking activity, they overwhelmingly rated the multi-tasking activity as non-distracting. It may be that the multi-tasking activity was their main focus during the experiment versus the performance ratings making the task less difficult. It may also be that the multi-tasking activity was simply unchallenging. Regardless of participants' opinions, results showed the competing task did have a negative impact on performance ratings.

Results also revealed that regardless of experimental condition, all participants increased in their accuracy scores over time for reasons unknown to us. It is possible that participants discussed the training material between experimental sessions; thus subjecting the study to treatment contamination. Participants may have been even more prone to discuss the material among each other considering a majority of training material was contradictory to what students in our pilot data considered effective teaching behavior. Nonetheless, this would imply that if treatment contamination was better controlled for in the present study, the already significant effects of training would be even stronger.

The second hypothesis concerned the effect of participants pre-existing prototypes on subsequent performance ratings. By presenting the raters with the performance factors (prototype dimensions) prior to completing performance ratings, we expected that training would facilitate the formation of relevant prototypes of performance and helps raters avoid making global impressions of ratee behavior (Smith, 1986). It was our hope that raters would use these abstract representations of performance when rating performance on the vignettes from memory. Therefore, we believed that this would maximize the likelihood that raters would correctly categorize professor behavior on each of the dimensions. We specifically hypothesized that undergoing Performance Dimension Training would influence the categorization of incoming performance information; thus we believed participants would alter their pre-existing prototypes to match the dimensions and behaviors presented during training that were shown to constitute effective and ineffective performance rather than relying on past prototypes to make judgments that are inaccurate.

Preliminary analysis of individual prototypes revealed no significant differences among any groups on prototype ratings concerning the eleven prototype dimensions. This finding is fundamentally important because in order to examine differences in individual prototypes post-training, it was essential to show that participants held extremely similar beliefs about professor behavior at the onset of the study. Analysis of prototype randomization revealed the participants' prototypes essentially looked alike at the onset of the experiment.

Post-training assessment of prototype change revealed partial support for our second hypothesis. It was initially expected that participants in the training conditions would change their prototypes to conform to the prototype presented in training to be the "ideal" prototype. Our results indicated significantly different changes in prototype ratings by training condition from Time 1 (initial assessment) to Time 2 (immediate post-training assessment) for the following seven prototype dimensions: Moves through material quickly, Remains motionless while teaching, Is organized, Uses visual aids during lecture, Displays an attitude of superiority, Speaks loudly, and Does not exhibit structure. Results demonstrated participants' prototype ratings concerning the remaining five dimensions (Has a sense of humor, Uses examples to support class material, Speaks loudly during lecture, and Assigns group work) were *already closely* conformed to the "ideal" prototype of professor behavior presented during training. Essentially, there was little change to be exhibited in the participants' existing prototypes concerning these five dimensions. Therefore, this finding simply validates our pilot data in terms of confirming what students already believed to be effective professor performance in the general college student population and partially supports the goal of changing participants' pre-existing prototypes.

It is important to note that we were able to change participants' prototypes in the training condition for four of the five dimensions that *went against* prototypic behavior attained through pilot data; Moves through material quickly, Remains motionless while teaching, Displays an attitude of superiority, and Does not exhibit structure. This finding has encouraging implications concerning the effectiveness of Performance Dimension Training. Training facilitated the change in participant's beliefs about what constituted effective professor behavior even when the prototype dimension went against what they previously believed to be ineffective. Simply changing the rationale behind the specific behaviors and articulating that research supported the effectiveness of the behaviors appears to have been sufficient and effective in terms of changing participants' pre-existing beliefs. The only *non-prototypic* dimension we were not able to change participants' prototypes on was "Requiring class participation." It is possible that participants in the study may have very enduring beliefs on this particular dimension and were unconvinced by our fictitious reasoning.

Results concerning participants' change in prototypes from Time 2 (post-training assessment) to Time 3 (1-week follow-up assessment) revealed no significant changes. As hoped for, it was believe that participants' change in prototype from Time 1 to Time 2 would remain stable over the 1-week follow-up if the new prototype proved durable/lasting. Because all participants' prototype ratings were equal to begin with, training appears to be the major explanation for the findings. A small regression towards the mean was noted in participants' prototype ratings at the Time 3 assessment, but this finding was not significant. This finding has important implications as well for the impact of Performance Dimension Training. If training is effective, it is even more appealing and essential that the results of training last over the duration of time between training and appraisal ratings.

It was also hypothesized that participants' prototype change would be the mediating variable by which accuracy scores were influenced by rater training (Hypothesis 3). Results showed partial support for this hypothesis at the immediate post-training assessment and the one week follow-up assessment indicating the changes in participants' prototypes were partially responsible for the increase in training effectiveness. It is possible that training increases awareness of the ideal prototype presented in training; thus encouraging participants to alter their pre-existing prototype to conform to the ideal. These findings imply that to some extent, raters were relying on the newly formed prototypes when completing the appraisal ratings from memory, and correctly categorized ratee behaviors according to the new standard.

It is interesting to note that mediation effects were even stronger at the one week-follow up assessment than at the immediate post-training assessment. Participants could have mentally reviewed training information frequently during the lapes between Session I (immediate post-training) and Session II (one week follow-up) and therefore the material was more established and accessible in memory for Session II performance ratings. However, it would seem more reasonable that training material would be even more accessible immediately after training versus over time.

Finally, it was thought that some individuals may be more resistant to prototype change because of individual differences (Hypothesis 4). It was specifically believed that those individuals who preferred consistency and were uncomfortable with no structure would be less likely to accept the non-prototypic information presented during training; thus producing less accurate ratings. Analysis concerning the three covariate measures (Need for Cognition, Preference for Consistency, and Personal Need for Structure) indicated they did not significantly contribute to the change.

Limitations

The first concern of the present study focuses on the prototype assessment measure. The prototype assessment measure was created reliant on the definition of prototype given by Cantor & Mischel, 1979, Rosch & Mervis, 1975, and Smith & Medin 1981. The definition states that prototypes are abstract sets of features commonly associated with members of a category, with each feature assigned a weight according to the degree of association with the category. The various dimensions of professor performance identified from pilot data in the present study served as the abstract sets of features associated with professor performance on the prototype assessment measure. One may argue that this is not actually an assessment of an individual's overall prototype of professor performance, but merely exemplar information related to the role of professor. Although the definition of prototype used in the present study is consistent with the definition used in this body of research, future research using different measures of prototype assessment may help to better understand changes in this cognitive mechanism.

The second concern centers around the use of undergraduate students as the sample for the study. Generalizing the results using this pool of participants should be taken with caution. However, we chose to use a profession that the participants should have been familiar with and we chose to use a performance task most appropriate for this sample. The participants should have already formed individual ideas about the effectiveness of various professor behaviors from attending university lectures on a consistent basis.

Similarly, generalizability of our results may be compromised due to using vignettes of professor performance versus using observational performance. It may have been more useful to use actors demonstrating professor lecturing styles for the portrayal of professor behavior than using written vignettes. Relying on role-playing manipulations for presenting professor behavior may be a more realistic circumstance and therefore increase the generalizability of our findings.

Generalizing our results concerning the lasting effects of Performance Dimension Training is also a focus of concern. Although our results demonstrated the training effectiveness carried over to Session II of the study, the lapes in time between the two sessions was only one week in duration. It is unlikely that organizations will be able to consistently administer training before each evaluation period thus requiring a longer duration between training and appraisal ratings. Our results also indicated a slight regression toward the mean in participants' prototype ratings at the Time 3 prototype assessment (Session II). Although the change in individual prototypes was significant for the training condition on a majority of the dimensions, this change in prototype may not be long-lasting. Future research increasing the time between training and performance evaluation would provide more evidence for the robustness of training effects and prototype change.

Using a single trainer, the first author of the study, in the training condition of the study is also a potential concern. It may have been the dynamic force of the experimenter's presentation style that influenced the effectiveness of Performance Dimension Training on rating procedures. Unfortunately we cannot separate out the trainer effect from the training effect. There does not, however, seem to be any reason why this may have happened.

Implications

Subject to these limitations, various implications for the present study have been identified. One of the most provocative findings was that participating in Performance Dimension Training was related to a larger change in individual's pre-existing prototypes toward the "ideal" prototype presented during training than if not participating in the training condition. Because it is believed that raters rely on training information to make performance ratings (Sulsky and Day, 1992), future rater training should also focus on assessing individual prototypes pre-training and individual differences that differ from training material. By identifying idiosyncratic raters organizations can focus on those specific employee beliefs that are especially different from the organizations performance evaluation standards. Training protocol may be tailored to meet these needs, and therefore, this may likely enhance the simplicity with which information is encoded and stored in memory for future ratings.

The present study demonstrated that Performance Dimension Training is a useful tool to increase the accuracy of performance scores. This finding has particularly important implications considering the present study used only a lecture format for the training protocol. If such simplified training is indeed effective, this could have potentially significant benefits for organizations in terms of saving administrative time and money that is spent on lengthy training sessions. Although this limited training has shown to be effective, it is important to note that this type of training may only be successful due to short time periods between training and performance appraisal ratings. More longitudinal research in this area is needed.

Conclusion

In conclusion, the present study contributed to the rater training literature by validating the effectiveness of Performance Dimension Training in terms of increasing the accuracy of performance ratings and demonstrating these effects across time. It was also demonstrated that participants in the study did change their pre-existing prototypes of professor performance post-training for most of the eleven prototype dimensions. Finally, it appears that the change in prototypes partially mediated the effectiveness of training. A better understanding of the cognitive devices underlying training effectiveness will aid in improving rater training programs and may further increase the accuracy of appraisal ratings.

References

Anderson, R.C., & Pickert, J.W. (1978). Recall of previously unrecallable information following a shift in perception. *Journal of Verbal Learning and Verbal Behavior*, 17, 1-12.

Athey, T.R., & McIntyre, R.M. (1987). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. *Journal of Applied Psychology*, 72(4), 567-572.

Baron, R. M., & Kenny, D. A (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Bernardin, H.J., & Buckley, M.R. (1981). Strategies in rater training. *Academy of Management Review*, 6, 205-212.

Bernardin, H. J., Buckley, M. R., Tyler, C. L., & Wiese, D. S. (2000). A reconsideration of strategies in rater training. *Research in Personnel and Human ResourcesManagement, 18,* 221-274.

Bargh, J.A., & Thein, R.D. (1985). Individual construct accessibility, person, memory, and the recall judgment link: The case of information overload. *Journal of Personality and Social Psychology*, 49, 1129-1146.

Bernardin, H.J., & Pence, E.C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology*, 65(1), 60-66.

Bernardin, H.J., & Villanova, P. (1986). Performance appraisal. In E.A. Locke (Ed.), *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management* (pp. 43-62). Lexington, MA: Lexington Books.

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

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 12) New York: Academic Press.

Cacioppo, J.T., Petty, R.E., Feinstein, J.A., Jarvis, B.G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bullitin*,119(2), 197-253.

Cialdini, R.B, Trost, M. R., & Newsom, J.T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, 69(2), 318-328.

DeNisi A.S., Cafferty, T.P., & Meglino, B.M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance*, 33, 360-396.

Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66(2), 127-148.

Feldman, J.M. (1986). Instrumentation and training for performance appraisal: A perceptual-cognitive viewpoint. In K.M. Rowland & G.R. Ferris (Eds.), *Research in personnel and human resource management*, 4, 45-99.

Hauenstein, N.M, Foti, R.J. (1989). From laboratory to practice: neglected issues in implementing frame-of-reference training. *Personnel Psychology*, 42, 359-378.

Illgen, D.R., & Feldman, J.M. (1983). Performance appraisal: A process focus. *Research in Organizational Behavior*, 5, 141-197.

Landy F.J, & Farr, J.L. (1976). Behaviorally anchored scales for rating the performance of police officers. *Journal of Applied Psychology*, 61(1), 750-758.

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

Lord, G.R. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology*, 70(1), 66-71.

Lord, R.G, & Maher, K.J. (1990). Alternative information-processing models and their implications for theory, research and practice. *Academy of Management Review*, 15(1) 9-28.

McIntyre, R.M. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology*, 69(1), 147-156.

Nigro, L.G. (January-February,1981). Attitudes of federal employees toward performance appraisal and merit pay: Implications for CSRA implementation. *Public Management Forum*, 84-86.

Neuberg, S. L., & Newsom, J.T. (1993). Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology*, 65(1), 113-131.

Pulakos, E.D (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, 4, 581-588.

Rosch, E., & Mervis, C.B. (1975). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.

Rugg, E.A., & Norris, R.C. (1975). Student ratings of individualized faculty supervision: Description and Evaluation. *American Educational Research Journal*, 12(1), 41-53.

Smith, D.E. (1986). Training programs for performance appraisal: A review. *Academy of Management Review*, 11(1), 22-40.

Smith, D.E., (1984). Increasing rating accuracy through rater training: A look at the performance appraisal process. (Doctoral dissertation, Colorado State University). Dissertation Abstracts International, 45, 1614-B.

Smith, E.E., & Medin, D.C. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Sulsky, L.M., & Day, D.V. (1992). Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. *Journal of Applied Psychology*, 77(4), 501-510.

White, D.J, & Carlson, D.E. (1983). Consequences of schemata for attention, impressions, and recall in complex social interactions. *Journal of Personality & Social Psychology*, 45(3), 538-549.

Table 1

Means and standard deviations of prototype assessment by training condition across time

Dimension	Training Condition Training (T) No Training (NT)	Time 1		Time 2		Time 3	
		Mean	SD	Mean	SD	Mean	SD
Moves through material quickly	T	2.57	1.07	3.33	.948	3.16	.927
	NT	2.53	.980	2.51	.936	2.50	.959
Has a sense of humor	T	4.25	.828	4.77	.674	4.50	.604
	NT	4.27	.736	4.37	.678	4.30	.622
Remains motionless while teaching	T	1.75	.878	2.82	.978	2.73	.884
	NT	1.49	.667	1.86	.649	1.75	.616
Is organized	T	4.60	.649	4.66	.540	4.76	.498
	NT	4.67	.575	4.04	.752	4.44	.623
Uses visual aids during lecture	T	4.44	.808	4.60	.576	4.57	.635
	NT	4.47	.651	4.42	.595	4.40	.661
Requires class participation	T	3.45	1.04	3.74	.977	3.56	1.00
	NT	3.44	.946	3.63	.952	3.35	.924
Uses examples to support class material	T	4.54	.736	4.62	.552	4.58	.517
	NT	4.51	.581	4.47	.502	4.45	.542
Displays an attitude of superiority	T	2.46	1.17	3.2	1.05	3.22	1.07
	NT	2.46	1.14	2.31	1.04	2.37	1.09
Speaks loudly during lecture	T	4.20	.767	4.58	.668	4.53	.639
	NT	4.18	.789	4.09	.830	4.08	.825
Assigns group work	T	3.14	.966	3.41	.934	3.25	.977
	NT	2.86	1.10	3.09	1.04	3.07	.997
Does not exhibit defined structure for course work or lecture	T	1.64	.811	3.20	1.02	2.93	.945
	NT	1.77	.869	2.20	.874	1.98	.782

Table 2

Means and standard deviations of mean Euclidean distance prototype ratings from the ideal prototype by training condition and time of assessment

	Mean	Standard Deviation
Training		
Prototype Assessment Time1	4.593	1.260
Prototype Assessment Time 2	2.414	1.185
Prototype Assessment Time 3	2.652	1.188
No Training		
Prototype Assessment Time 1	4.707	1.215
Prototype Assessment Time 2	4.189	1.943
Prototype Assessment Time 3	4.411	1.197

Note: A score of "0" denotes extremely accurate and higher scores indicate less accuracy.

Table 3
Means and standard deviations of accuracy scores by training condition

	Accuracy Time 1 Mean (SD)	Accuracy Time 2 Mean (SD)
Training	1.60 (.587)	1.37 (.662)
No Training	2.15 (.635)	1.83 (.926)

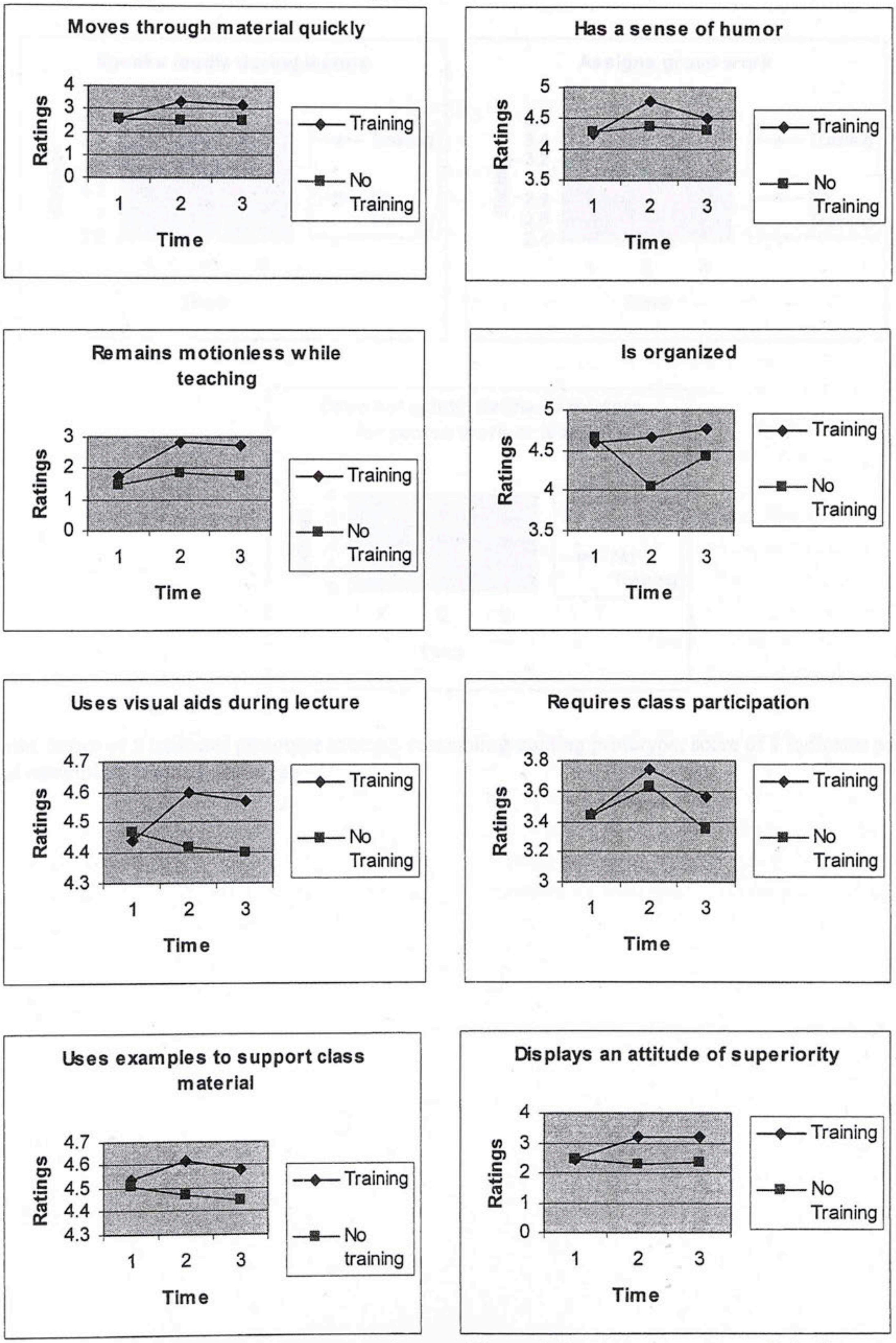
Note: score of 0 denotes extremely accurate; score of 5 denotes extremely inaccurate

Table 4
Means and standard deviations of accuracy scores by multi-tasking condition

	Accuracy Time 1 Mean (SD)	Accuracy Time 2 Mean (SD)
No Multi-Tasking	1.81 (.586)	1.49 (.610)
Multi-Tasking	1.95 (.739)	1.71 (.999)

Note: score of 0 denotes extremely accurate; score of 5 denotes extremely inaccurate

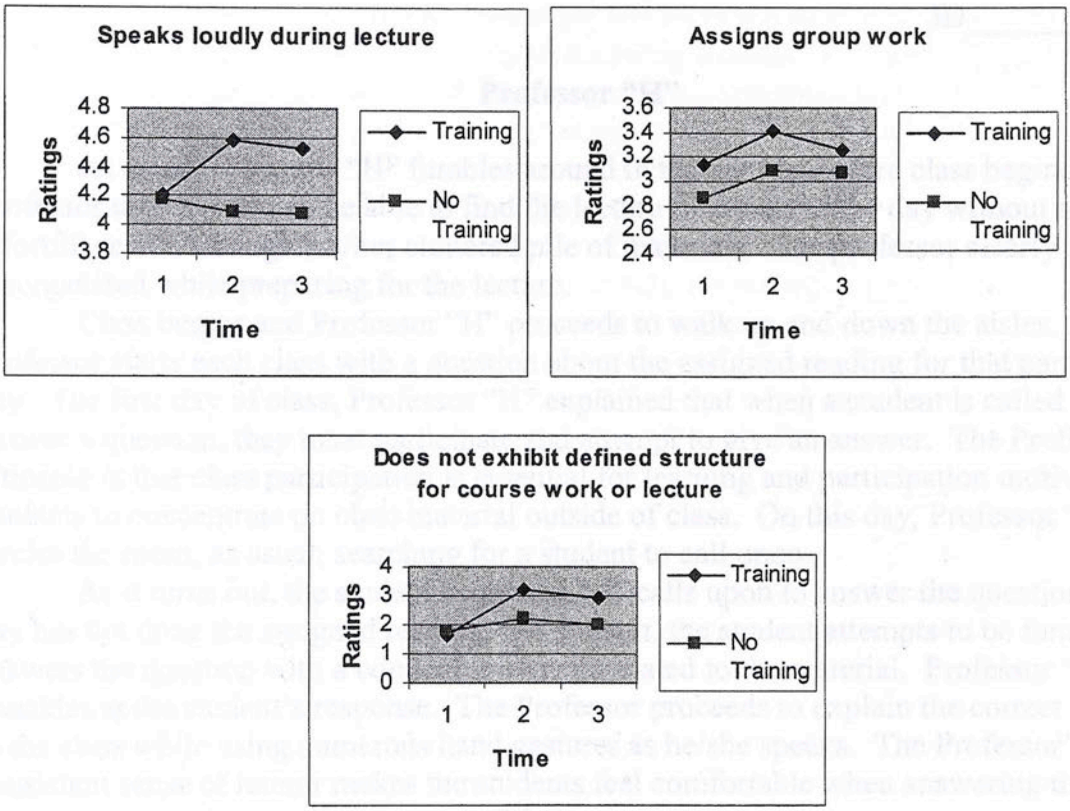
Figure 1
Mean prototype ratings by training condition across time



Note: Score of 5 indicates prototype strongly resembling training prototype; score of 1 indicates prototype not resembling training prototype

Figure 1 continued…

Mean prototype ratings by training condition across time



Note: Score of 5 indicates prototype strongly resembling training prototype; score of 1 indicates prototype not resembling training prototype

Appendix A
Professor Vignettes

ID___________

Professor "H"

As usual, Professor "H" fumbles around in his/her bag before class begins. The Professor rarely seems to be able to find the lecture material for the day without an effortful search through his/her cluttered pile of materials. The professor clearly appears disorganized while preparing for the lecture.

Class begins and Professor "H" proceeds to walk up and down the aisles. The Professor starts each class with a question about the assigned reading for that particular day. The first day of class, Professor "H" explained that when a student is called on to answer a question, they must participate and attempt to give an answer. The Professor's rationale is that class participation is essential for learning and participation motivates students to concentrate on class material outside of class. On this day, Professor "H" circles the room, as usual, searching for a student to call upon.

As it turns out, the student Professor "H" calls upon to answer the question of the day has not done the assigned reading. As a result, the student attempts to be funny and answers the question with a comical answer unrelated to the material. Professor "H" chuckles at the student's response. The Professor proceeds to explain the correct answer to the class while using numerous hand gestures as he/she speaks. The Professor's consistent sense of humor makes the students feel comfortable when answering the questions even when they are unsure of their answers.

Although Professor "H" often appears disorganized, he/she does, however, have a clearly defined structure for the course and daily activities. The Professor lets the students know exactly what will be covered daily, and does not deviate from the guidelines. Professor "H's" single-mindedness in sticking to the schedule leaves no opportunity to slow down the lecture pace if students are unable to grasp the concept. Moreover, the Professor is reluctant to give answers to students' questions as it slows down the pace of lecture.

STOP

ID______________

Professor "J"

Professor "J" walks into the classroom and stumbles over a student's book bag. The professor regains stability and jokes about being clumsy. Professor "J" laughingly explains, "I always try to trip at least once a semester…(chuckling)." Generally speaking, most students regard this comment as having a sense of humor.

Professor"J" reaches the front of the room and takes time to set up a poster that will enhance the day's lecture materials. The poster is a diagram of the typical steps involved in conducting research experiments. Professor "J" uses the diagram as a visual explanation to help the students get a more thorough understanding of the topic. Professor "J" almost always includes visual aids along with the lecture.

As Professor "J" explains the first two steps in conducting research, there is a small rumble of noise coming from the back of the classroom. Apparently, the students in the back three rows are having trouble hearing Professor "J" speak. They are only hearing bits and pieces of the lecture because Professor "J" is speaking too softly half of the time.

Professor "J" spends the entire class period on steps one and two of "conducting research." Because Professor "J" is moving so slowly through the material, Professor "J" is unable to get through all eight of the essential steps that will be on the test.

Professor "J" ends most classes by assigning a small in-class discussion. On this particular day, the students are instructed to create small groups of five and discuss a research experiment they have read about. They are also to identify in the experiment they choose to discuss, the first two steps in the "conducting research" model that Professor "J" previously discussed in class that day.

STOP

ID____________

Professor "K"

Professor "K" enters the room and loudly announces that class will begin. The Professor asks that all the students stop talking. The students in the back row clearly hear the request and most of the chattering in the room subsides. A few students continue to trickle in late, but find plenty of seats in the back of the room. The students in the back have no problem hearing Professor "K" lecturing because of his/her resounding voice.

Following the announcement, two students in the front row are apparently telling jokes. Professor "K" displays his/her view of a classroom by stating, "The classroom is no place for jokes or humor. You are here to learn, not act like clowns." The class is not surprised by the comment. They have noticed the Professor's serious nature from the first day of class.

Everyday before lecture, Professor "K" explains that he/she will to cover a lot of material during the class period so the students are advised to pay close attention. The Professor mentions that class notes may be found on the internet if students are unable to keep up with the fast pace. Professor "K" believes it is important to cover all aspects of a topic even if it means lecturing very rapidly.

Because Professor "K" consistently covers a large amount of information in the course, class periods consist of lectures only. Professor "K" does not use any form of visual aids to supplement the material for fear of wasting valuable class time.

Similarly, the Professor feels that in-class group work and in-class discussions take away from his/her ability to cover enough information. As a result, students are rarely able to discuss any of the material amongst themselves during class.

STOP

ID_______________

Professor "L"

It is the first day of class. Professor "L" enters the classroom and immediately arranges the day's handouts along the table in the front of the room. Professor "L" neatly organizes the stacks so they can be distributed in the order that the material is discussed that day. Professor "L" has clearly organized the handouts so no class time will be wasted on sorting the papers. Next, the Professor prepares the video for the lecture by setting the tape to the exact starting point. Finally, the Professor adjusts his microphone to the proper level so class can begin.

Next, Professor "L" faces the class with a firm voice and blatantly declares, "I know that some of you are under the impression you are familiar with the class material we will be covering. Sadly, you are mistaken. I assure you, you *do not* know this material at all. However, not all is lost. I am indeed teaching this class, and I will see to it that you all are better educated in this topic before you leave." Most of the class appears shocked at how Professor "L" makes himself out to be so superior to the students. It is also obvious by the students' reactions that the professor made his point clear regarding the nature of the class.

Professor "L" then describes the expectations he/she has for the class during the semester. The Professor states that there will be no class involvement required whatsoever. The Professor feels that students should not have to contribute during discussions or lecture in order to create a more comfortable learning environment for the class. The professor wants to make sure everyone feels relaxed.

Next, Professor "L" begins to lecture on "Psychological Disorders." The Professor supplements the lecture with a video that demonstrates real-life examples of people suffering from the disorders the Professor is covering. The Professor believes these "real" examples help the students to acquire a better understanding of the material and relate the topic to real-life.

Throughout the class period, Professor "L" refrains from moving around the classroom while lecturing. The Professor stays positioned behind the podium in the front of the room except when distributing the handouts. Also, Professor "L" refrains from making any gestures or hand movements while speaking. Both hands remain placed directly at his/her side throughout the lecture.

STOP

ID______________

Professor "M"

Professor "M" begins Statistics class late as usual. The Professor typically spends the first few minutes of class each day trying to locate the microphone and sorting through slides for the day's presentation. As a result, the students are becoming accustomed to his/her lack of organization.

On this day, Professor "M" begins by writing a statistical problem on the board. The Professor, gesturing with his/her hands, acknowledges his/her attitude of supremacy by introducing the problem with the following statement: "Some of you will spend months trying to find the solution to this problem and some of you will spend the rest of your natural born lives trying to figure it out! Good luck."

Although the Professor's attitude is a bit intimidating, the students are able to relax as they examine the problem. Because Professor "M" does not require class participation, the students will not have to attempt to give a solution in front of the class and risk unavoidable defeat.

However, the students do feel somewhat irritated because they rarely understand these types of problems that Professor "M" presents to them. The Professor consistently fails to use examples to support the material that he/she teaches in class. Moreover, the class does not always grasp the material very easily and often finds it hard to apply the equations and statistical rules to actual application problems.

Similarly the students frequently find it difficult to focus on what Professor "M" is saying because the Professor is constantly moving around the room. Many students find this excessive movement distracting as well as the Professor's routine hand gestures.

STOP

ID____________

Professor "Z"

Professor "Z" enters the room and a student immediately approaches the podium. The student learned that he missed a handout from the previous class period and asks the Professor for a copy. Professor "Z" has no trouble finding the handout because all of the course material is neatly organized in his/her bag. Each handout and each day's lecture is visibly organized into file folders. The student then sits down and Professor "Z" begins class.

As always, the students are not sure what the day's topic will be because Professor "Z" does not adhere to a rigid class structure. Often times, the Professor will spontaneously come up with a class activity that supports the day's lecture. As a result, the class will cover very few notes and future test questions will focus on those activities. Despite the lack of structure, the students appear to enjoy his teaching style.

Because of the frequent activities, Professor "Z" requires that the students participate. Whether the activity is a group discussion or an individual "question and answer session", each student is graded on their participation. Professor "Z" assigns 50% of the semesters grade based on class participation.

Throughout the semester, the students have learned to focus on the class activities and refrain from behavior that is tangential to the mission of the class. Because Professor "Z" has no sense of humor, he/she does not tolerate any childish and impractical behavior. Professor "Z" takes his/her teaching position very seriously and asks that the students do the same.

Regardless of the day's activity, Professor "Z" always stays positioned in the front of the classroom. Moreover, the Professor intentionally avoids using too many hand movements and gestures while speaking. Professor "Z" wants to avoid any motion/movement that would divert the students' attention from focusing on the material.

STOP

Appendix B
Prototype Assessment

ID___________

Professor Behavior

We are interested in your personal beliefs about professor performance. On a scale of one to five, please indicate how **typically effective** you believe each of the following behaviors are concerning professors in general. Remember, we are interested in your opinions of professors in general.

SCALE

Ineffective				Effective
1	2	3	4	5

When you think about professors, how typically effective are professors who:

1. Move through material quickly?
   1 2 3 4 5

2. Have a sense of humor?
   1 2 3 4 5

3. Remain motionless while teaching?
   1 2 3 4 5

4. Are organized?
   1 2 3 4 5

5. Use visual aids during lecture?
   1 2 3 4 5

6. Require class participation?
   1 2 3 4 5

7. Use examples to support class material?
   1 2 3 4 5

8. Display an attitude of superiority?
   1 2 3 4 5

9. Speak loudly during lecture?
   1 2 3 4 5

10. Assign group work and group projects?
    1 2 3 4 5

11. Do not exhibit defined structure for course work or lecture?
    1 2 3 4 5

Appendix C
Rating Scale and Rating Sheets

Rating Scale

Please indicate how well the professor performed along the five dimensions (items). A score of "1 "indicates that the professor was ineffective (did not perform well) and a score of "5 "indicates the professor was very effective (performed well).

Professor___

1) How well did the professor perform on Teaching Skills?
   1 2 3 4 5

2) How well did the professor perform on Lecture Presentation?
   1 2 3 4 5

3) How well did the professor perform on Student Involvement?
   1 2 3 4 5

4) How well did the professor perform on Interpersonal Rapport?
   1 2 3 4 5

5) How well did the professor perform on Communication Skills?
   1 2 3 4 5

Rating Sheet

(PDT /Multi-Task/ Control/Multi-Task) Vignette 3 ID____________

PROFESSOR: H J K L M Z

Professor Ratings

Question 1

1 2 3 4 5

Question 2

1 2 3 4 5

Question 3

1 2 3 4 5

Question 4

1 2 3 4 5

Question 5

1 2 3 4 5

How many sequences of three or more odd digit numbers were recited on the tape recording?

____

On a scale of one to five, how distracted were you during the ratings?

1	2	3	4	5
Not distracted				Very Distracted

Rating Sheet

(PDT /Multi-Task/ Control/Multi-Task) Vignette 1 & 2 ID__________

PROFESSOR: H J K L M Z

Professor Ratings

Question 1

1 2 3 4 5

Question 2

1 2 3 4 5

Question 3

1 2 3 4 5

Question 4

1 2 3 4 5

Question 5

1 2 3 4 5

How many sequences of three or more odd digit numbers were recited on the tape recording?

____

Rating Sheet

(PDT only/ control) Vignette 1, 2, & 3 ID__________

PROFESSOR: H J K L M Z

Professor Ratings

Question 1

1 2 3 4 5

Question 2

1 2 3 4 5

Question 3

1 2 3 4 5

Question 4

1 2 3 4 5

Question 5

1 2 3 4 5

Appendix D
Performance Dimension Training

Performance Dimension Training

Phase 1: The rating scale

Now, I would like to show you all the rating scale that you will be using to evaluate different professors on their performance. The rating scale includes 5 items. You can notice each item is written as a statement about how the professor performed. You will indicate how effective or ineffective the professor performed along these dimensions in the scenarios you will read.

Item 1: **Teaching Skills**
Item 2: **Lecture Presentation**
Item 3: **Student Involvement**
Item 4: **Interpersonal Rapport**
Item 5: **Communication Skills**

RATING SCALE

1) How well did the professor perform on **Teaching Skills**?
1 2 3 4 5

2) How well did the professor perform on **Lecture Presentation**?
1 2 3 4 5

3) How well did the professor perform on **Student Involvement**?
1 2 3 4 5

4) How well did the professor perform on **Interpersonal Rapport**?
1 2 3 4 5

5) How well did the professor perform on **Communication Skills**?
1 2 3 4 5

The questions will be presented to you one at a time in the above format on over-head transparencies. Please ask any questions you have about the rating scale at this time.

Phase 2: Behaviors and behavioral rationales that make up each dimension.

Each of you may understand the five dimensions differently and may have different ideas about their meanings. The important thing is that you each evaluate the professors' performance along the same criteria. The foundation for your evaluations should be based on the prior research explaining the success of these specific teaching strategies.

Item 1: Teaching Skills
When rating this item, the experts examine how well professors:

1- Move through lecture material quickly (Ponis & Rodgers, 2001)

Research has shown it is important to move quickly through class material because:

- Allows for more material to be covered and learned during class time.
- Encourages students to stay focused. Students must stay focused on the material to record all of the notes given when the lecture is fast-paced lecture.

2- Are organized (Green & Kline, 2002)

Research has shown it is important that professors are organized because:

- Class time is used more effectively. Little class time is wasted on organizing and collecting the day's materials when a professor is organized.
- Allows for easier distribution of course materials. Professors who are organized find it easier to distribute handouts. They also find it is easier when organized, to distribute and locate materials for students who may have missed class.

Item 2: **Lecture Presentation**
When rating this item, the experts examine how well professors:

1- *Do not* exhibit defined structure for course work or do not exhibit defined structure during class-time (Fash & Lange, 2003).

Research has shown it is important *not to exhibit* defined structure for course work or lecture because:

- Allows for unexpected events such as academic conferences and cancelled school days.
- Allows professors to spend more time on material that students fail to understand. If a professor realizes the students are not grasping the material, he/she may spend as much time as needed on that particular topic to make sure everyone understands the information.

- Leaves room for spontaneous class activities/ discussions. Research has shown students give more personal opinions and personal insights concerning the class material when there is not a defined structure for class time.

2- Use visual aids (Green & Kline, 1999)

Research has shown it is important that professors use visual aids because visual aids:

- Assist students in grasping the material from different perspectives. Many students are "visual learners" and acquire a better understanding of the material by "seeing" demonstrations and tangible information.

- Students perceive the class as more interesting and less monotonous when they are shown visual aids.

3- Use examples to support class material (Cooper & Linsner, 2000)

Research has shown it is important to use examples to support class material because:

- Gives students real-life applications of the information. Students understand the information better when they are given a real-life example or an example of a concept versus only a definition.

- Facilitates a better understanding of the material. Research has shown that more information given about a subject to support its meaning leads to a better comprehension and understanding of the material.

Item 3: **Student Involvement**

When rating this item, the experts examine how well professors:

1- Assign group work/projects (Daley & Mickle, 1998)

Research has shown it is important to assign group work or group projects because group work:

- Encourages critical thinking. Studies have shown that when students work in groups they generate more examples and more new ideas.

- Allows students to apply course material to practical situations. Students are able to discuss the principles they learn in class in more depth through discussion and group projects.

- Facilitates learning to effectively work with others. Students learn to work as a team and respect differing opinions when working on group projects.

2- Require class participation/involvement (Cooper & Linsner, 2003).

Research has shown it is important to *require* class participation/involvement because:

- Encourages students to learn valuable social speaking skills. Students learn to speak in front of others and become more confident in their presentation abilities.

- Helps students gain a better understanding of the material. When students are required to respond to questions presented in class, studies show they think more critically about the information before they answer, and as a result, learn the information more efficiently.

Item 4: **Interpersonal Rapport**

When rating this item, the experts examine how well professors:

1- Display an attitude of superiority (Terry & Jones, 2003)

Research has shown it is important that professors display an attitude of superiority because this:

- Encourages students to develop respect and confidence in professors. Studies have shown students who view their teachers as "superior" or as having a superior attitude describe the professors as more credible and more educated.

2- Have a sense of humor (Kool & Hekele, 2000).

Research has shown it is important that professors have a sense of humor because:

- Creates a light-hearted learning environment. Research has shown when professors display a sense of humor, students feel less intimidated and relaxed while learning.

- Students report enjoying class more when professors have a sense of humor. As a result, research has shown increases in class attendance.

Item 5: **Communication Skills**

When rating this item, the experts examine how well professors:

1- Remain motionless while lecturing (Kool & Hekele, 2000)

Research has shown it is important that professors remain motionless while lecturing because remaining motionless:

- Creates less environmental distractions. Research has shown that when professors stay situated in one spot during the lecture (for example: they do not wander around the classroom or through the isles) students are less distracted by this type of communication. Research has also shown that when professors do not use excessive hand movements and hand gestures, students are less distracted. As a result, students are able to focus on what is being taught more efficiently and attentively.

2- Speak loudly during lecture (Debis & Borin, 1999)

Research has shown it is important that professors speak loudly during lecture because:

- All students are able to hear the lecture. Professors must speak loud enough that all students, including those in the back of the lecture, can clearly hear the lecture without straining and can hear the lecture with ease.

- Students focus more intently on the professor. Research has shown professors who speak loudly are viewed as more knowledgeable and confident. As a result, students are more attentive.

SUMMARY

You have learned that research has shown certain professor behaviors to be very effective. When evaluating professor performance, experts use the above criteria. In summary, experts rate professors more positively when professors:

Teaching Skills:

1- Move through lecture material quickly.

2- Are organized.

Lecture Presentation:

3- Do not exhibit defined structure for course work or lecture.

4- Use visual aids.

5- Use examples to support the material.

Student Involvement:

6- Assign group work or group projects.

7- Require class participation.

Interpersonal Rapport:

8- Display an attitude of superiority.

9- Have a sense of humor.

Communication Skills:

10- Remain motionless while lecturing.

11- Speak loudly during lecture.

Appendix E

Rating Instructions

(Control) & (PDT/Only)

Rating Task Instructions (No-Multi-Task Conditions)

In the following exercise, you will rate three short scenarios of professor performance. You will be given approximately **three** minutes to read each scenario. After you read one scenario, you will answer five questions about the professor's performance in that scenario. You will rate how well the professor performed in regards to five separate categories:

1) Teaching Skills – the professor's general teaching skills
2) Lecture Presentation- the professor's presentation of class material
3) Student Involvement- the professor's ability to involve students
4) Interpersonal Rapport- the professor's attitude and demeanor
5) Communication Style- the professor's verbal/nonverbal communication

Each of the five questions will be shown separately on overhead slides. You will have approximately **20** seconds to answer each question before the next question is presented. Each time you are instructed to turn the page in your packet, you will NOT be allowed to return to any of the previous pages. So, it is important that you pay close attention to the information as you are reading each of the short scenarios

After you have answered all five questions for the first scenario, you will be instructed to read the second scenario. Again, you will then answer the same five questions about the second scenario. Finally, you will be instructed to read the last scenario and answer the five questions once more.

NOTE: It is VERY important that you rate the professor's performance as accurately as possible. The following scenarios are actual behavior of first year professors who are being evaluated on their performance. Your ratings may influence their first year evaluations.

Do you have any questions about this exercise before we begin?

(Control/Multi-Task)(PDT/Multi-Task)

Rating Task Instructions (For Multi-Task Conditions)

Often times in organizations employees experience various distractions in the work environment as they attempt to accomplish everyday responsibilities. As a result, it is very important that employees are able to focus on more than one task at a time. We are interested in observing how well people are able to handle multiple task simultaneously.

In the following exercise, you will rate three short scenarios of professor performance. You will be given approximately **three** minutes to read each scenario. After you read one scenario, you will answer five questions about the professor's performance. You will rate how well the professor performed in regards to five separate categories:

1) Teaching Skills – the professor's general teaching skills
2) Lecture Presentation- the professor's presentation of class material
3) Student Involvement- the professor's ability to involve students
4) Interpersonal Rapport- the professor's attitude and demeanor
5) Communication Style- the professor's verbal/nonverbal communication

Each of the five questions will be shown separately on overhead slides. You will have approximately **20** seconds to answer each question before the next question is presented. Each time you are instructed to turn the page in your packet, you will NOT be allowed to return to any of the previous pages. So, it is important that you pay close attention to the information as you are reading each of the short scenarios

After you have answered all five questions for the first scenario, you will be instructed to read the second scenario. Again, you will then answer the same five questions about the second scenario. Finally, you will be instructed to read the last scenario and answer the five questions once more.

Because employees must often focus on more than one task at a given time, you will be given a brief sequence recognition task as you answer the five questions about the short stories. Right before you are instructed to answer question number one, a voice recording will be played on a cassette player. You will hear an experimenter reciting even and odd digit numbers. Your task is to remember how many times and odd digit is repeated THREE or MORE times in a row. (Even digits eg: 2, 4, 6, 8, Odd digits eg: 1, 3, 5, 7, 9) For example, if you hear:

"1, 2, 5, 3, 5, 2, 1, 2, 8, 5, 7, 9, 1, 3, 5, 2"

You will have heard three odd digits or more spoken TWO times

{5, 3, 5} & {5, 7, 9, 1, 3, 5}.

In the sequence : "2, 8, 8, 4, 2, 4, 8," there are NO odd digit sequences of three or more. Every time an even digit is recited, the sequence of odd digits is broken.

In this study, you must pay attention to the tape recording as well as answering questions about the professor's performance. You will receive the sequence recognition task during the rating process for each scenario.

You may mark on your rating sheet to keep a tally of the odd digit sequences or you may keep track of them mentally.

The tape recording will be stopped after the fifth question is answered. You will be instructed to record how many three or more odd digit sequences you heard during that particular set of ratings. You will record the total number on the appropriate blank on the bottom of your rating sheet.

NOTE: It is VERY important that you rate the professor's performance as accurately as possible. The following scenarios are actual behavior of first year professors who are being evaluated on their performance. Your ratings may influence their first year evaluations.

Do you have any questions about this exercise before we begin?