The Role of Competing Tasks, Level of Knowledge,

and Personal Prototypes in Performance Appraisal.

by

Patricia L. Marshall

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

APPROVED:

Roseanne J. Foti, Chairperson

Joseph J. Franchina

Sigrid Gustafson

Neil M. A. Hauenstein

K. Dow Scott

May, 1992

Blacksburg, Virginia

THE ROLE OF COMPETING TASKS, LEVEL OF KNOWLEDGE, AND PERSONAL PROTOTYPES IN PERFORMANCE APPRAISAL

by

Patricia L. Marshall

Committee Chairperson: Roseanne J. Foti

Psychology

(ABSTRACT)

Recently, studies in performance appraisal have focused more on the process of performance appraisal rather than on the mechanics of performance appraisal. This study focused on the individual differences among raters and the effect these differences have on recognition accuracy and rating accuracy. Specifically, individual differences were examined through individual constructs (personal prototypes) of different performance levels and individual levels of knowledge for a particular job. Furthermore, the effect of a competing task on recognition accuracy and rating accuracy were examined.

The present study employed a 2 (presence/absence of competing task) X 2 (high family resemblance prototype/low family resemblance prototype) between subjects design with level of knowledge as a covariate. Eighty experienced carpentry workers and construction laborers from Southern Virginia were the subjects. The subjects were tested on their individual level of knowledge with a carpentry knowledge exam. Personal prototypes for each subject were obtained by individual interviews with each subject. The competing task manipulation consisted of subjects in the competing task condition receiving a task to complete while viewing the videotape while subjects in the no competing task condition did not receive a task. After the knowledge exam and prototype interview, the sample viewed a videotape of three carpenters performing four different woodworking tasks (sanding, sawing, hammering and staining) and completed a recognition and performance appraisal measure after observation. It was expected that subjects, regardless of their level of knowledge, with a competing task to perform during observation would make less accurate ratings than subjects with no competing task to perform. It was also hypothesized that raters with higher job knowledge would recognize more behaviors consistent with their personal prototypes and falsely recognize more behaviors from their personal prototypes than lower knowledge raters. Further, it was hypothesized that lower job knowledge raters should be more accurate in their recognition of ratee behaviors than would higher job knowledge raters. Further, it was hypothesized that higher knowledge raters would falsely recognize behaviors consistent with their personal prototype when there was a competing tasks than when was no competing task. There should be no difference for lower knowledge raters. Finally, it was hypothesized that high job knowledge raters would make less accurate performance appraisal ratings because of their reliance of their personal prototypes under increased processing conditions whereas there should be no such differences for lower job knowledge raters. The competing task did serve to divide the subjects attention in viewing the videotape as compared with subjects without a competing task. Little support was obtained for the hypotheses. These findings suggest that the presence/absence of competing task did not significantly effect rater accuracy and that subjects' personal prototypes did not effect neither their recognition accuracy nor rating accuracy. Future implications for research study will be offered.

ACKNOWLEDGEMENTS

Many people deserve acknowledgement with the completion of a project of this magnitude. I realize that I cannot mention all people helpful to me during this process but hopefully I have expressed my appreciation to them along the way.

I would first like to thank my advisor, Roseanne J. Foti. I doubt I would have finished this project without her support and encouragement. No question or comment concerning my dissertation was too trivial or small for her. I realize completing this dissertation long distance was difficult and time-consuming but Roseanne tolerated my endless trivial phone calls and was always available to listen and comment. Thank you, Roseanne.

Also, I would like to express my appreciation to Dr. Joe Franchina for his numerous editorial comments beginning with the initial proposal up until the final dissertation draft. Dr Franchina added exceptional clarity to my writing.

To my other committee members I want to express my appreciation for your time and input over the course of this project. Thanks Sigrid, Neil, and Dow.
From the Blacksburg area, I would like to thank Mary Catherine Thompson. She has been a friend and a supportive presence throughout the graduate school process. She always had the time to listen to me moan and groan during this project while being the best shopping buddy and diet advisor.

I would also like to thank Virginia A. Marshall. Ginny had time to listen even with the responsibilities of three children and a husband. She also gave the other basic necessities such as needlework direction, Kahn's advice, and a spare car.

I would also like to thank Kim McLaughlin for being a good friend and cheerleader to the end. She was always there to listen no matter how far apart we

are. Also, she's lives in a great place to visit when things get really rough.
From the New Jersey arena: I would like to thank Bernice Anglin for her timely wisdom and directness. Bernice has been and continues to be a close friend and confidant no matter how far away.

Also, Doug McCormick for his feedback concerning this project and his ability to calm me down. Doug's sense of humor always helped me put things in perspective.

From the Danville area:

I would like to thank my cousin, Ramona Haraway, for always being there for me and having faith in me. She always allowed me to be the best judge of what I could do and what decisions I should make.

Finally, and most importantly, I would like to thank my parents. Their support played a major part in making this dissertation a reality. Nothing I could say could really express the true depth of my gratitude to my parents. Further, my mother served as my unofficial research assistant and she deserves credit for helping me call my subjects and using her connections to get me into the construction companies I used in my studies.

Surprisingly, I would like to thank my dog Aloysius for being a good companion for the past five years. He put up with many late, restless nights and short walks. Thanks Ali.

In conclusion, I would like to dedicate this to my late grandmothers, Eunice Davis and Cora Marshall. I hope you both will be proud of this accomplishment.

Table of Contents

LIST OF TABLES

The Role of Personal Prototypes, Level of Knowledge, and Competing Tasks in Performance Appraisal.

Introduction

Impression formation is an important topic area in social psychology, social cognition, and applied psychology. When forming impressions, people utilize psychologically meaningful categories (i.e. traits) to organize their memories (Asch, 1946; Hamilton, 1981; Ostrom, Pryor, & Simpson, 1981). Impression formation is influenced also by the goals the perceiver has for encoding and retrieving information (Hastie, Park, & Weber, 1984; Srull & Wyer, 1986, 1989; Wyer & Srull, 1980, 1986). These impressions allow the perceiver to test hypotheses about others (Slowiaczek, Klayman, Sherman, & Skov, 1989; Snyder, Campbell, & Preston, 1982; Snyder & Swann, 1978). Recently, literature in the performance appraisal area of applied psychology has begun to focus on how individuals (raters) form impressions of others by categorizing ratees based on their personal construct system (Borman, 1987; Jackson, 1972; Rothstein & Jackson, 1975; Siess & Jackson, 1970). This introduction ties together the literature from social psychology, social cognition, and applied psychology.

In the past, performance appraisal literature has focused on the 'mechanics' of performance appraisal, that is, the development of formal evaluation instruments and their susceptibility to various systematic errors (for example, halo, leniency/severity, central tendency) (Cooper, 1981; Feldman, 1981). According to Brown (1968), past research has failed to consider performance appraisal as a dynamic process but rather has oversimplified the performance appraisal process and its associated validity problems. Since the 1970's, there has been a shift in the performance appraisal literature towards a focus on the 'process' of performance appraisal.

Performance appraisal has been construed as a function of three interacting systems: the organizational context within which the appraisal takes place, the information processing system of the appraiser, and the behavioral system of the appraisee (Landy & Farr, 1983). Recently, research on the process of performance appraisal has centered on performance appraisal which involves information search, organization, and retrieval (DeNisi, Cafferty, & Meglino, 1984; Mount & Thompson, 1987; Murphy & Balzer, 1986; Smither, Reilly, & Buda, 1988). To manage the various processing demands placed on them, raters tend to utilize numerous cognitive management strategies in the appraisal situation. One cognitive management strategy a rater may use is his/her personal construct theory. Kelly (1955) stated that individuals form personal construct systems which they use to judge events and make predictions about future events. These strategies may lead to various errors in encoding, in retrieval and in subsequent judgments of ratees. Some examples of the factors which may influence the use of various cognitive management strategies are competing tasks, rater knowledge, and the raters' ability to categorize and define their environment.

The purpose of this study is to explore the possible constraints on the rater's performance appraisal. Raters with more job knowledge may process, recall, and evaluate others differently than raters with less job knowledge. How a rater categorizes information may depend on that rater's "mini-theory" of performance levels. The amount of information which raters encounter in the observation stages of performance appraisal may also affect their strategies for dealing with information.

Prototypes, Personal Prototypes and Information Overload

Previous research has indicated that people are limited in the amount of information they can process (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Bruner (1958) pointed out that categorization is basic to perception, information storage and organization. Despite built-in storage limitations, people use categories to process large quantities of information (Smith, Adams, & Schorr, 1978). Two methods of processing large quantities of information are schemata and prototypes.

A prototype resembles an instance in which all known attributes are filled in even if all the known characteristics are not directly relevant to category membership (Anderson, 1980). Prototypes are used to organize the most common or 'average' characteristic associated with an instance. A prototype of a jock, for example, contains a large and varied set of features that through experience have come to be associated with very athletic types. Prototypes allow individuals to distinguish traits, categories, and/or stimulus objects.

According to Markus (1977), people vary in their schemata or, specifically, prototypes about themselves, and use these self-schemata to organize self-relevant information and to process information about others (Markus Smith, & Moreland, 1985). Hastorf, Schneider, and Polefka (1970) point out that categories and their associated 'personal' prototypes have been referred to as stereotypes or implicit personality theories depending on whether they are based on attributes such as race or labels such as aggressive. Individuals tend to sort and classify the people they meet into categories based on the target person's similarity to that individual's personal system of categorization. These implicit personality theories affect all stages of information processing. Previous research has shown that personal prototypes and categorization affect initial impressions (Fiske & Taylor, 1991; McCloskey & Glucksberg, 1978); encoding (Zadny & Gerard, 1974); and retrieval (Cantor & Mischel, 1977).

The following study proposes that individuals have their own personal idea of what constitutes a good performer versus a poor performer. Each individual possesses schemata and prototypes of performance. Past research (Jackson, 1975; Whitely & Doyle, 1976; 1978) dealt with the notion that individuals possess implicit personal theories of performance levels. A student's implicit theories of performance levels were identified by extracting a set of latent categories that accounted for the communality of categorization patterns between subjects (Whitely & Doyle, 1978). The consequences of these implicit theories are that the rater's implicit theories may represent both a true component and an error component. The true component may correspond to the part of the implicit theory which reflects the actual pattern of covariation in ratees; the error component may correspond to the part of the implicit theory which is invalid or determined by individual response patterns of the student rater.

Previous studies have focused on average prototypes (i.e. averaging across individuals) in investigating schemata and prototype utilization (Rosch, 1978; Cantor & Mischel, 1979; Lord, Foti, & DeVader, 1984). The problem with using average prototypes is that average prototypes mask important differences in personal prototypes. Individuals have their own idea of what constitutes a good versus a poor performer. Taking each individual's prototype and averaging across individuals would serve to mask important differences between people. Further, the personal experience and knowledge of the rater may influence the prototype of the rater. A person with considerable job knowledge will be more likely to construct a different prototype than that of a person with less job experience and knowledge. The individual with more job knowledge will also encode, recognize, and recall different aspects of performance (Hauenstein & Walker, 1989). In other words, the level of rater experience and knowledge will greatly affect a rater's prototype.

Further, raters who have a competing task to perform while observing performance will be less able to encode, store and later recall all the nuances of performance than raters without a competing task to perform while observing performance. Raters with a competing task are more likely to categorize ratee's level of performance and then use their 'personal idea' of a good, average, and poor performer to generate behaviors of a particular ratee than raters without a competing task.

Individuals form and construct their own "individual" construct systems and then use these systems to judge events (Kelly, 1955). These systems aid individuals in organizing and simplifying information. Borman (1987) noted that it appeared logical that there might be significant individual differences in work-related constructs that affect what a rater looks for in observing ratee work behavior. For example, if one rater has an important construct, "sharing new job information with others on the job", and a second rater doesn't share that construct or anything similar to it, the first rater may be more likely than the second rater to concentrate on work behavior related directly to that aspect of performance. Borman (1987) applied this personal construct theory to Army officers and found that these officers generated a core set of constructs which differentiated effective from ineffective non-commissioned officers and that different officers emphasized different combinations of this set of constructs.

Raters often perform many tasks; supervision and performance appraisal may be just two of the tasks performed on the job. There have been a few studies dealing with the increasing processing demands of raters. Foti and Hauenstein (1990) investigated the effects of processing demands and prior impressions on subsequent judgments of raters. They found that when processing demands were minimal or typical, the rater's prior impressions either had no effect on judgments or biased processing toward inconsistent behaviors. However, as processing demands increased, rater's judgments were biased toward prior impressions. Raters tended to use information that would confirm their initial impression. These results are consistent with the idea that as processing demands increase, raters tend to exhibit biased encoding in the direction of their original impressions of the ratee (assimilation encoding) (Bargh & Thein, 1985; Graesser, Gordon, & Sawyer, 1979; Hastie & Kumar, 1979; Woll & Graesser, 1982; Wyer & Gordon, 1982).

Rater Expertise and Knowledge

An area of research which has recently received attention in the performance appraisal literature is rater expertise and knowledge. A rater's level of experience appears to play a critical role in performance appraisal. Kozlowski, Kirsch, and Chao (1986) found that raters with high job knowledge were more sensitive to true performance covariation, whereas low job knowledge subjects were more sensitive to their internal conceptual similarity schemata when rating a familiar ratee. Kozlowski et al also indicated that raters high in job knowledge tended to rely on actual performance covariation when rating familiar ratees, but they relied on their conceptual similarity ratings when rating unfamiliar ratees. Raters with less job knowledge and with less ratee familiarity exhibited more halo error than high knowledge raters. Hauenstein and Kovach (1988) found that raters with more interview experience (high knowledge) recalled more judgments and fewer behaviors of target applicants than less experienced raters. Despite the fact that job knowledgeable raters used categorization information, high job knowledgeable raters tended to encode and store behavioral information as well as low knowledge raters did and tended to be more efficient processors of information than were low knowledge raters.

Summary

The present study highlighted past research dealing with information processing but also added to the literature by combining variables in a manner which simulated more realistically the cognitive demands which individuals encounter by focusing on individual differences. Previous research has explored the rating process by focusing on group differences in information processing while ignoring individual differences in information processing. These studies have ignored the fact that individuals have their own idiosyncratic view of the world. Perhaps, the demands on individuals may directly demonstrate the influence of individual differences in the rating process.

The focus of this study is the influence of rater knowledge, competing tasks, and personal prototypes on performance appraisal. Specifically, personal prototypes for one occupation were generated and a family resemblance score computed for each subject. The sample consisted of carpenters, carpenter apprentices, and laborers in the construction industry in Part I of this study. Subjects also completed a job knowledge test. Part II was designed to test two stages of the performance appraisal process: 1) recognition of stimulus information and 2) accuracy of performance judgements.

Every organization needs some method of distinguishing between good and poor performance and every employee desires information about the effectiveness of his or her performance. There has been considerable dissatisfaction with the development and progression of performance appraisal systems (Borman, 1978; Kane and Lawler; 1978; Landy and Farr, 1980). Researchers have noted the importance of recognizing the rater or evaluator as an information gatherer operating in a complex environment with many demands (Feldman, 1981). Also, researchers have noted the need for a more integrated approach to performance appraisal research. And, finally, all this must be embedded in an organizational context that serves to promote and constrain behavior in various ways.

Literature Review

The following presentation reviews six areas of literature. First, a review of schema theory is presented including the role of schemata in encoding, retrieval, and judgement. Second, cognitive categorization literature along with category development literature is reviewed. Third, prototype and personal prototype literature is presented. Fourth, information overload and divided attention literature is presented. Fifth, the effect of task complexity and competing tasks in performance appraisal is presented. Finally, a brief review of the literature focusing on the role of knowledge and expertise in the rating process is presented.

Schema Theory

Bartlett's (1932) schema theory was an early attempt to explain the relations between an abstract structure and the recall of specific facts related to that abstract structure. To Bartlett, memory was not primarily reduplicative or reproductive. Bartlett proposed that perceived events are assimilated into mental schemata that have been formed by similar events which have occurred in the past in the perceiver's world. Bartlett asserted that memory for complex materials consisted of a general impression (schema) together with abstracted details which can be used to reconstruct the original message presented. Thus, the perceiver has a general impression (schema) which is less likely to contain specific, word-for-word productions and more likely to be used in reinterpretation, integration of originally separate concepts, and inferences about the material presented.

A schema is defined as a cognitive structure that represents knowledge about a given concept or type of stimulus (Brewer & Nakamura, 1984; Fiske & Linville, 1980; Fiske & Taylor, 1991; Hastie, 1981; Rumelhart & Ortony, 1977; Taylor & Crocker, 1981). A schema also consists of a framework for tying together the information about any given concept or event with specifications concerning the types of interrelations and restrictions placed on the way information fits together (Norman & Bobrow, 1976). Schemata are organized according to specifiable a priori criteria (Wyer & Gordon, 1984). Schemata are concerned with top-down processing which is heavily reliant on one's prior experience in contrast to bottom-up processing which is more data-driven (Abelson, 1976; Bobrow & Norman, 1975; Rumelhart & Ortony, 1977).

Schemata develop through experience (Anderson, 1981; Anderson, Kline, & Beasley, 1979). The more a person encounters schema-relevant examples the more abstract the schema becomes (Abelson, 1976; Abelson, 1982; Anderson, Kline & Beasley, 1979). One approach to understanding the impact of schemata on the interpretation of events and stimuli is to look at the amount, accuracy, speed, and focus of memory. Smith, Adams, and Schorr (1978) demonstrated that story details are remembered more rapidly when they have a unifying theme than when they do not. Subjects had difficulty rejecting distractors that were thematically related to integrated facts.

Tesser (1978) stated that a schema is essentially a naive theory of a stimulus domain and the individual user, a naive scientist. A schema tells us what stimulus attributes we should attend to and what ones to ignore. Second, a schema contains a network of associations which tie the stimulus attributes, together, thereby, providing rules for thinking about the stimulus. Schemata may vary in terms of abstractness, complexity, and function. Schemata also exhibit a perseverence effect: schemata stubbornly exist even in light of contradictory evidence (Fiske & Taylor, 1991). According to Abelson and Reich (1969), one of the simplest schemata may be a set of logically or 'psychologically' related propositions which combine to represent a generalization about one's social world. For example, "people with similar values usually like one another," and "people generally get what they deserve".

The use of schemata allows an individual to identify stimuli quickly, to fill in information missing from the stimulus configuration, and, possibly, to select a strategy for obtaining further information, solving a particular problem or reaching a goal. Higgins, Herman, and Zanna (1981) organized the functions of schemata into two general categories: (1) the encoding and representation function and (2) interpretative or inferential function. The encoding and representation function consists of ways in which schemata enable the individual to acquire information from the environment and represent this acquired information in short term memory. The interpretative or inferential function serves to aid the perceiver to solve problems, set goals or select an appropriate response.

There are various types of schemata in social cognition: person schemata, self-schemata, role schemata, event schemata, and content-free or procedural social schemata (Fiske & Taylor, 1991). Person schemata, which are composed of traits or goals, are a person's understanding of the psychology of typical or specific individuals. Person schemata aid in categorizing other people and remembering schema relevant behavior. Self-schemata are the schemata which we possess about ourselves. These schemata serve to guide the processing of information about the self. Role schemata describe the appropriate norms and behavior for broad social categories, such as age, race, sex, and occupation. Event schemata are cognitive structures about sequences of events or standard social occasions (Abelson, 1981; Schank & Abelson, 1977). Event schemata help observers to understand ambiguous information and to fill in missing information. Finally, content free or procedural social schemata consist entirely of rules for linking elements but not much elaborate content; they guide information processing toward schema-relevant information.

Bartlett (1932) originally proposed that the recall of text information which was related to the underlying theme of a passage was more likely to be recalled than was information not connected to the theme. Schema consistent information will be recalled more readily than schema inconsistent information particularly when making complex judgments (Bodenhausen & Lichtenstein, 1987; Bodenhausen & Wyer, 1985; Brewer & Treyans, 1981; Rumelhart, 1977; Thorndike, 1977). Hastie (1981) suggests that schema-inconsistent information is remembered more easily since discrepant information receives added attention. This memory advantage for inconsistent information holds true if the goal for the subjects is to form an impression but it does not hold true if the goal is remembering the material (Srull, 1981; Srull & Wyer, 1989; Wyer, Bodenhausen, & Srull, 1984; Wyer & Gordon, 1982). Brewer, Dull, and Lui (1981) found that inconsistent information required longer encoding time than did consistent or irrelevant information.

Rater expectations also influence encoding. Previous research indicates that raters have a tendency to encode probable behaviors at a grosser and more general level than inconsistent behaviors. However, when ratee behaviors are unpredictable, the behaviors of the target individual will be perceived at a more specific level (Newtson, 1973; Wilder, 1978). For example, over time a supervisor who observes an employee working in a consistent and predictable manner will most likely encode that employee's behavior at the grossest level since close monitoring by the supervisor is most likely unnecessary and unwarranted. The supervisor will come to expect consistent performance of this type of employee and thus find it unnecessary to encode behaviors at a discrete level. On the other hand, an employee who performs inconsistently and unpredictably while performing identical work would have their behavior coded at a more discrete level (e.g. proofreading their work, setting the correct margins). It is apparent that rater expectations can affect the number and type of critical behavior incidents extracted from the behavior stream (Glover, Plake, Roberts, Zimmer, & Palmere, 1981; Glover, Plake, & Zimmer,1982).

Summary

A schema provides a framework for interpreting and categorizing new information. Each person forms his or her schemata from past and present experiences. Fiske and Taylor (1991) point that schemata have a perseverance effect; schemata persist even in the presence of contradictory evidence. Because individuals are limited in their information processing capacity, schemata aid the subject in being able to handle the extra processing demands.

Cognitive Categorization

Another cognitive mechanism which people utilize in tying together information is categories. Categories are fuzzy sets; membership is not based on fixed, well-defined formal attributes. Individuals identify category members by assessing their similarity to the category prototype. Bruner (1958) noted that categorization is basic to perception, information storage, and organization. Categorization allows people to process large quantities of data despite inherent limitations of storage capacity. Individuals and events can be classified into categories (Rosch,1977; Rosch, Mervis, Gray, Johnson, and Boyes-Braehm, 1976).

Evidence shows that people learn and remember information by actively categorizing or coding input according to well learned conceptual schema (e.g. Schank and Abelson, 1977). The categorization process fits into the schematic process by explaining how a given schema might be applied to a particular instance (Fiske & Taylor, 1991). For example, categorization research might tell you how to decide that a futon might be a bed and a schematic process describes the effects of the label on subsequent perceptions, memories, and inferences about beds.

As Wyer and Gordon (1984) point out, categories are represented by a set of features that serve as a bases for inferring membership in it. Schemata are cognitive representations whose features are organized according to specifiable a priori criteria. An example of a category would be 'lawyer' which may be characterized as well-educated, middle-class, and articulate. Some of these attributes may be more typical of some lawyers than others. On the other hand, an example of a spatial schema, in contrast, is a face with its associated nose, eyes, and mouth.

Research has indicated that category utilization is governed in part by how recently the category has been activated. For example, a person who has his or her car broken into would probably view anyone lurking near his or her car as a potential thief. In one study, Higgins, Rholes, and Jones (1977) first exposed subjects to positive and negative traits of a target, adventurous versus reckless. These are two words with similar meanings but different connotations. All subjects read about Donald, who shot rapids, drove in a demolition derby, and planned to learn skydiving. In this study, subjects who had previously been primed with the relevant positive trait "adventurous" evaluated Donald more positively than did people who had been primed with "reckless". The use of an inapplicable priming trait with the control group showed no priming effect. This study suggests that the priming of a relevant trait makes that trait more accessible in future interactions. Srull and Wyer (1979) and Wyer and Srull (1980, 1981) suggest that this priming effect is due to the fact that activating a category places it at the top of the mental heap or "storage bin".

Mount and Thompson (1987) looked at the effects of cognitive categorization on a rater's accuracy, leniency, and halo of performance evaluations. Subordinates rated the performance of managers on three performance dimensions. Rating accuracy was higher when the subordinate perceived the manager as performing in a manner congruent with the subordinate's expectations of how the manager should behave. Zadny and Gerard (1974) stated that behaviors which are expected are more salient and as a result are noticed and recalled more easily than unexpected behaviors. The Mount et al study also indicated that subordinates who perceive their managers behaving congruently with expectations demonstrated greater halo on performance dimensions than did subordinates who perceived their managers behaving incongruently. Further, subjects who perceived their managers' behaviors to be congruent exhibit greater leniency in the three performance dimensions than did subordinates who perceived their managers' behaviors to be incongruent.

In summary, categories represent a set of features that serve as bases for inferring membership in it. Category features are not organized or interrelated except by their presence in the set that is common to category members. There is no necessary relationship between category characteristics. Categories and schemata are conceptually different types of cognitive representations; however, they are nevertheless related. For example, the variables of a schema may frequently be categories. Therefore, the concrete representation of schema variables in terms of a given set of information is essentially a process of assigning features of this information to categories specified in the schema.

Schemata are cognitive structures which are primarily focused on the necessary essentials of category membership and not the ideal instance. Prototypes are a type of schemata. People use schemata and prototypes to cognitively simplify information, organize incoming information, and later fill in information about a target person by recalling their schema for that person and filling in information associated with their particular schema.

Category Development. The more experience an individual has with a stimulus, the more dimensions of that stimulus become salient to the individual. Category or schema development occurs through increasing experience with some stimulus or situation. The more experience an individual has with some stimulus or situation, the more differentiated and complex his or her schema becomes. Well-developed schemata contain more information (Crocker, Fiske, and Taylor, 1984). As schemata develop they become more abstract (Anderson, 1981; Anderson, Kline, & Beasley; 1979; Johnson, 1981), more tightly organized (Smith, Adams, and Schorr, 1978), more cognitively compact (Burnstein & Schul, 1982, 1983; Fiske & Dyer, 1985; Schul, 1983; Schul, Burnstein, & Martinez, 1983; Sentis & Burnstein, 1979), and more resilient to inconsistency (Fiske & Neuberg, 1990; Higgins & Bargh, 1987; Ruble & Stangor, 1986).

When individuals are experts about a domain they have more information about that particular domain and they have more information of different dimensions of that specific domain area (Chase & Simon, 1973; Linville, 1982; Linville & Jones, 1980). Well-developed schemata contain more schema-consistent information. Thus, for an expert, any single piece of incongruent information confronts a vast store of congruent information in memory.

Prototypes. Prototypes have been defined as: (1) the mean value of the set of stimulus objects on each relevant feature dimension (Posner & Keele, 1968; Reed, 1972); (2) an abstract set of features commonly associated with members of a category (Cantor, 1981); and (3) an abstract set of features commonly associated with members of a category, with each feature assigned a weight according to degree of association with the category (Cantor & Mischel, 1979; Rosch & Mervis, 1975; Smith & Medin, 1981). Cantor and Mischel (1977) hypothesized that "average" attributes become the focus of attention in prototypicality judgments made on the basis of social information. For example, a prototype of a "criminal type" might have 8.0 years of education, have 2.3 children, be .75 black and .75 male and with .35 twitch on the face (Cantor, 1981). Judgments of similarity to a prototype are based on the family resemblance criterion. Specifically, no single set of features defines the category but rather any of several features contribute to the perception that an object resembles the category prototype. A trait prototype contains a list of specific traits that are considered to be highly related to a more abstract superordinate trait that functions as a unifying category label (e.g., introversion: timid) and the prototype is comprised of the average value of each feature picture. The difference between categories and prototypes lies in the "fuzziness" of categories (Fiske & Taylor, 1991). People identify category members by assessing their similarity to the category prototype. As Fiske and Taylor (1991) note, categorization operates by family resemblance within a fuzzy set, not by a rigid set of rules. Categories, being organized hierarchically, vary in terms of levels of abstraction with prototypes being less abstract than categories and schemata.

Posner and Keele (1968) and Reed (1972) demonstrated the tendency for subjects to abstract a prototypic visual pattern representing the central tendency of a set of dot patterns. This abstracted schema was then used as the standard against which to compare new patterns. The abstraction of the visual prototype also guides subjects to incorrectly recognize items in a recognition memory task when the items closely resemble the abstracted prototype (Franks & Bransford, 1971; Posner & Keele, 1970; Reitman & Bower, 1973). Prototypes, again, are the 'average' of a set of feature dimensions of a stimulus object.

Personal Prototypes

Individuals develop and form idiosyncratic 'theories' of job performance. Prototypes, as mentioned previously, can be seen as good examples of a schema. Schneider, Hastorf, and Ellsworth (1979) stated that these implicit personality theories describe assumptions individuals make about relationships between traits in people. Individuals possess their own personal idea of what are important constructs of performance and these constructs may not be shared by another rater.

Borman (1987) investigated personal constructs, performance schemata, and folk theories of subordinate effectiveness. U.S. Army officers recorded the names of nine non-commissioned officers (NCOs) that they considered to be effective in their jobs and 9 NCOs that they considered to be ineffective in their jobs. Six triad combinations of these 18 role persons were presented with three triads consisting of two effective vs. one ineffective NCO and the other three triads consisting of two ineffective vs. one effective NCO. These subjects were asked to record how the effective NCOs were different from the ineffective NCOs.

A label and a definition were provided by the subjects for each of these differentiated constructs. One hundred eight-nine constructs were generated by the officers and these constructs were grouped into 49 reference dimensions covering the following areas: (1) personal characteristics and personal traits; (2) cognitive and physical abilities; (3) performance constructs relevant to most or all army enlisted jobs; and (4) military leadership constructs. The results indicated an eight factor solution with the data indicating 123 of the 189 personal work constructs generated (65.1%) substantial loading on a content factor. Of the remaining 66 constructs, 21 (11.1%) loaded on subject factors and 45 (23.8%) had mixed loadings or low communalities. The results of this study demonstrated that army officers have a common set of constructs which differentiate effective from ineffective officers but that different officers emphasize different combinations of this set.

Individual also form their own theories about themselves. Markus (1977) proposed that each individual has a self-schemata which serves "to form a cognitive representation about the self drawn from past experience, which organizes and guides the processing of self-related information contained in the individual's social experiences" (p. 64). People then use these self-schemata to organize information about others in schema-related domains particularly when drawing inferences beyond the information given (Catrambone & Markus, 1987). The purpose of self-schemata is to direct an individual's selection and organization of information about the self such that the person who views himself as independent will most likely require shorter processing time for words concerned with independence than to other types of words (Markus, 1977). Finally, subjects who view themselves as independent would endorse a significant number of adjectives associated with the concept of independence than do subjects who do not characterize themselves as independent. These self-schemata include cognitive representations derived from specific events, general situations, and general representation derived from repeated categorization and subsequent evaluation of the person's behavior by himself.

Markus, Smith and Moreland (1985) looked at the role of self-schemata in two experiments. In the first experiment, they hypothesized that schematics would see more coherence and meaning among activities of a film showing a target individual exhibiting self-schema relevant behavior. The authors used a task to compare individuals with relatively high levels of involvement and self-knowledge in masculinity, masculine schematics, with individuals with low levels of involvements in this area, aschematics. A schematic is a person who possesses a particular schema whereas a aschematic does not possess this particular schema but others (Markus, 1977). Markus (1977) pointed out that individuals are self-schematic on traits which are important to them and on which they are sure the opposite does not hold true.

All subjects were told to break two film segments into units, one irrelevant to masculinity and one relevant to masculinity. The authors found that schematics unitized the schema-relevant films into larger segments than the aschematics. The two groups did not differ in their unitizing of schema-irrelevant segments. These findings are consistent with the notion that schematics were able to use their self-schemata to organize behavior portrayed in the film into larger chunks.

The second experiment investigated the relationship between the self-schema and the perception of others. This study used the same task but modified the reason for focusing on the other person. There were three conditions in this study: (1) Replication of Study 1 where subjects were told to unitize the film into segments; (2) Detail condition where subjects were told to divide the film into the "smallest possible action units" and to "concentrate on the details of the film, and (3) Impression condition where subjects were given no instructions concerning the size of their units but were told to pay attention since they would be asked to make judgments later about the actor in the video. The results from Study 2 mirror the results in Study 1. In each condition, schematics described the actor with more masculine traits than did aschematics. The results, however, were not significant. The schematics when told to pay attention to the actor's behavior produced units smaller than the units produced by the aschematics. Markus and colleagues interpreted this as showing that when told to pay attention, schematics would extract more schema relevant behavior and that schematics were not just vague or global thinkers.

Information Overload and Divided Attention

Many complex situations require that information be processed about people and/or events. This requirement can create cognitive overload because the individual's attention must be divided or withdrawn from some aspect of his or her environment to focus on other aspects of the stimulus environment (White & Carlston, 1983). When an individual's attention is withdrawn or diverted from the target person, that individual is more likely to base his or her judgement of the target person on a schema. Alternatively, individuals who aren't forced to shift their attention and can thus focus on the target person are more likely to utilize the target person's actual behavior.

Bargh and Thein (1985) looked at the effect of increased processing demands, and their effect on how prior social information is used in impression formation. Bargh and Thein (1985) manipulated two levels of information overload, overload vs. non-overload, and two levels of chronic category accessibility, nonchronics vs. chronics. Drawing from the work of Higgins, King, and Mavin (1982), the authors defined 'chronics' as subjects who were familiar with the specified trait whereas 'nonchronics" were unfamiliar with a specified trait. In this study, subjects read either about a target who behaved in an basically honest manner but occasionally displayed dishonest behavior or a target who behaved in a basically dishonest manner but occasionally displayed honest behavior. Some subjects were given control over the presentation rate of information about the target. Other subjects had no control over the rate of information presented and thus experienced information overload. The former subjects gave more attention to and had better recall of the minority behaviors than of the majority behaviors. However, nonchronic accessors took more time in processing information and had increased recall of minority behaviors compared to chronic accessors. The 'nonchronics' free recall of impression consistent (majority) behaviors was not significantly affected by the information load manipulation. The chronic accessors did show a decline in free recall of majority behaviors under information overload conditions. Furthermore, in the information overload condition, chronic accessors had greater recall for minority behaviors than majority behaviors while nonchronic accessors did not have better recall for minority behaviors than majority behaviors.

These findings support Bargh and Thein's hypothesis that individuals will be able to develop impressions and engage in elaborative processing if they possess cognitive structures for which the presented information is relevant. In other words, perhaps chronic accessors have the attentional capacity necessary to form and maintain an impression of the target person during information overload and thus their patterns of free recall in the information overload condition would be similar to those in the self-paced information condition.

In White and Carlston (1983), subjects received information concerning the target actors in order to create an impression or schema of that person's behavior. Subjects then listened to an audiotape of two actors with the primed actor being in the foreground or background. A second condition allowed subjects to switch the audiotape so that the target person was in the foreground or the background. In the audiotapes, the target subjects either acted in a neutral manner, exhibiting positive or negative aspects of the relevant trait, or the target subject acted in a stimulus inconsistent manner. The results indicated that subjects shifted their attention back to the target person as soon as that person began behaving in a schema inconsistent manner. When the actors displayed behavior consistent with the rater's "primed" impression, they would shift their attention elsewhere, but when the actor behaved in a manner inconsistent with their "primed" impression, they shifted their attention back to the actor.

Summary and Conclusions

In summary, processing demands on the rater influence how the rater handles or simplifies the rating process. Raters utilize various cognitive simplification strategies such as social schemata in order to operate in a system with inherent processing constraints and demands. Other factors, such as self-schemata and prototypes, also influence the probability of recall for information. Unfortunately, these heuristics can lead to rating errors and biases. Furthermore, the amount of information which a rater must process will influence the level of information that he or she can recall in the rating process.

Task Complexity and Competing Tasks

Task complexity and competing tasks also influence performance appraisal. There has been a dearth of research in the applied literature on task complexity and competing tasks. Bodenhausen and Lichtenstein (1987) examined the effect that task complexity has on social stereotypes and information processing strategies. They manipulated type of judgment made (trait/guilt), the ethnicity of the target (Hispanic/nondescriptive), expected processing load (large/small) and the nature of the evidence presented (predominantly favorable/predominantly unfavorable information). It was hypothesized that stereotypes would have a greater impact on judgment tasks that are more complex. The results demonstrated that subjects with a comparatively simple goal of judging the defendant's aggression invariably showed no appreciable bias attributable to the defendant's ethnicity in their subsequent judgments of his aggressiveness, his guilt, or future behavior. However, subjects in the complex condition of judging the defendant's guilt consistently demonstrated biases in the judgments of aggressiveness, guilt or future behavior. These results demonstrated that subjects who have a complex processing objective will be more likely to adopt a heuristic strategy to achieve their goal. These results are in agreement with assertion made by Abelson (1976) and Wyer and Carlston (1979).

In experiment 2 of Bodenhausen et al (1987), subjects received two complexity manipulations: a guilt judgment objective (high complexity) vs. a judgment objective (low complexity) and an anticipated processing load manipulation (high vs. low processing load). The results indicated that subjects perceived the experimental task as more difficult when they were expected to make a guilt judgment than when they were expecting to make an aggressiveness (trait) judgment. However, the anticipated processing load manipulation did not significantly affect perceptions of task difficulty. Furthermore, the interaction between processing load and processing objectives also failed to reach significance. Perhaps, as the authors suggest, the expectancy of a large amount of information did not evoke the stereotyping strategy.

Foti and Hauenstein (1990) studied the effects of information overload on the rating process. They hypothesized that as the processing demands of raters increased, accuracy in recognizing performance incidents would decrease for all raters. Decreases in accuracy would be greatest in conditions where behaviors were consistent with prior performance information. Raters, in the 'primed' condition, would make more biased judgments in the direction of the primed cue and would demonstrate decreasing differentiation across performance dimensions. Subjects were given either a good, poor or no performance cue of the ratee and then viewed a videotape of an instructor presenting a lecture. Subjects were assigned to either a minimal, typical or taxed processing condition. Raters with a positive impression (good performance cue) were less accurate than other raters in their recognition of good performance incidents of the ratee. Rater's memory for poor performance incidents was not affected by the rater's prior impressions. As processing demands increased, the rater's prior impression of the ratee influenced the subsequent ratee behaviors to which the raters attended.

Hauenstein and Walker (1989) also explored the issue of job familiarity, task complexity, and performance appraisal. They found that when a judgment task was low in complexity, the correlations between the number of behaviors recalled and rating accuracy, for high job knowledge and low job knowledge, were positive and significant. However, when the raters were responsible for a highly complex task, the correlations between the recall of behaviors and rating accuracy for both types of raters were significant and negative. The data also demonstrated marginal significance for number of judgments recalled and rating accuracy among high and low job knowledge raters. It was concluded from this study that job knowledge and judgment complexity moderates the relationship between memory and accuracy of performance appraisal.

The type and number of behaviors an individual attends to and encodes is affected by the performance variability of the behavior which the rater views. Research has demonstrated known limitations in attentional processes and memory capacity (Hogarth, 1980; Klatzky, 1980). These limitations affect the number of behaviors reported and their rapid or overlapping occurrence affect the number and type of critical incidents reported by raters. Ebbesen and Allen (1979) showed that with an increase in the rate of exposure to actions there was produced a reduction in the accuracy of recognition memory for behavioral details.

In an everyday work environment, rating responsibility is embedded in a myriad of responsibilities. The centrality of the rating task also influences performance appraisal. Balzer (1986) examined the effects of initial impression (positive/negative) and centrality of the appraisal tasks (high/low) during performance appraisal.

Task centrality was concerned with whether or not the rating task was the sole activity required during the performance appraisal task. In the high centrality manipulation, subjects were told there would be a brief quiz on the material covered by the lecturer and that it would not be important to do well on the quiz. The main responsibility of the subjects was to complete a behavioral diary task. Subjects in the low rating task centrality condition were told it was critical to pay attention to and understand the content of the lecture. Subjects were also informed they would be given an 'exam' on the videotaped lecture material. Then the subject raters were shown a set of very difficult sample questions which were supposedly representative of the exam questions. Results showed that subjects who received good or poor initial performance information collected different types of critical behaviors from an identical second lecture. Specifically, the subjects who viewed good performance initially collected fewer behaviors than those who initially viewed poor initial performance. This was reversed in the high rating task centrality condition where subjects in the high rating task centrality condition/negative impression condition collected fewer incidents than subjects in the high task centrality/positive condition. On the other hand, subjects in the low rating task centrality condition collected fewer behaviors in the positive initial impression condition as compared with the negative initial impression condition.

<u>Summary</u>

In summary, task complexity and the presence of a competing task influences rater accuracy, perceived task difficulty and the use of cognitive heuristics (Bodenhausen & Lichtenstein, 1987; Foti & Hauenstein, 1990). Furthermore, the results from the rating task centrality/initial impression interaction suggests that different levels of rater centrality lead to different processing strategies depending on the initial impression raters hold for a ratee. In the Balzer study, individuals in high rating task centrality condition/negative impression condition collected fewer incidents than subjects in the high task centrality/positive condition. Also, overall, individuals in the positive initial impression condition recorded fewer positive behaviors than those who initially viewed poor initial performance condition.

Rater Knowledge and Expertise

Recently, researchers in performance appraisal have focused on the process of performance appraisal. Performance appraisal is a complex function for an employee. It requires him or her to observe, assimilate, and evaluate large-amounts of information about another employee. Studies in performance appraisal research have recently focused on the process of performance appraisal and some of the factors that may influence the evaluations of a rater.

The experience level of the rater may affect performance ratings. Kozlowski, Kirsch, and Chao (1986) studied the relationship between job knowledge, ratee familiarity, conceptual similarity, and halo error. Subjects were divided into groups of high and low job knowledge and high and low ratee familiarity. Ratee familiarity was operationally defined in terms of knowledge of ratees. Specifically, raters were told to select two baseball players from a list of major league players, one well-known and one not well-known. Level of knowledge was based on self-reports of the extent to which the raters were active observers of baseball. Subjects were asked to judge the degree of similarity for all possible pairings of seven performance dimensions unique to baseball. Kozlowski et al (1986) showed that raters with high job knowledge were more sensitive to true performance covariation. Raters with with low job knowledge were more sensitive to their internal conceptual similarity schemata when rating a familiar ratee. Further, raters high in job knowledge tended to rely on actual performance covariation when rating familiar ratees but on conceptual similarity ratings when rating unfamiliar ratees. More halo error was seen when raters were either unfamiliar with the content domain or with their ratees.

Kozlowski and Kirsch (1987) asked subjects to provide conceptual similarity (CS) judgments for all unique pairings of 10 performance dimensions for the job of baseball player. Each subject also rated the performance of 20 baseball players on each of the 10 performance dimensions. Individual level intercorrelation matrices were computed from these ratings. The evidence indicated that when raters evaluated familiar ratees, results were consistent with the systematic distortion hypothesis. The systematic distortion hypothesis states that a rater's observation, storage, and recall processes are biased in the direction of his or her implicit covariance schemata, that is their mini-theories about how events covary. Raters with less knowledge about baseball showed significantly larger CS rating covariation than did raters with relatively greater knowledge about baseball. However, when raters with low job knowledge rated unfamiliar ratees, they exhibited significant declines in CS rating covariation. This contradicts the systematic distortion hypothesis. In general, the magnitude of the CS rating covariance in this study was not high.

Kozlowski and Ford (1991) investigated the effects of ratee familiarity, information constraint, and memory on rater information acquisition strategies. Subjects were exposed to prior ratee information and then received varying amounts of incident information for the six ratees. There were four levels of delay: 0 or no delay, 2 days, 4 days, and 7 days. The subject raters were instructed to familiarize themselves with each ratee's performance record. The search procedure consisted of subjects being exposed to a 10-minute computer controlled training program with each subject rater assigned to different amounts of information they could access via the computer training information board. When raters felt that they had sufficient information to judge the ratee or when they assessed their information limit, they were instructed to produce an overall effectiveness rating for each of the ratees. The results indicated that ratee familiarity had an impact on search because raters accessed more performance-relevant information for ratees for whom there was less prior information. Raters under no information constraint searched for more information than those under high constraint. Raters tended to search for more performance information for poor performers than good performers when the level of ratee familiarity was high. Raters in the no delay condition tended to search for more information for low prior information ratees relative to delay raters. Finally, Hauenstein and Kovach (1988) found that familiar raters recalled fewer specific behaviors than did unfamiliar raters, made more dispositional inferences, and were slightly less accurate at recognizing behavioral incidents than unfamiliar raters.

Summary

In summary, some of these studies support the idea that rater knowledge and expertise increase accuracy. Kozlowski and colleagues found that experienced raters tend to be more acutely aware of true performance covariation with familiar ratees but use conceptual similarity ratings when rating unfamiliar ratees. Low job knowledge raters utilize conceptual similarity covariation whether familiar or unfamiliar with the ratee. Also, rates exhibited more halo when raters were either unfamiliar with the content domain or their ratees.

Hypotheses

This study investigated the effect of a competing task, level of knowledge, and personal prototypes on rater accuracy and behavior recognition. A field study was conducted to test these three factors using a 2 (high family resemblance prototype/low family resemblance prototype) by 2 (presence/absence of competing task) between subjects design with level of rater knowledge as a covariate. Subjects were asked to generate their idea of a good, average, and poor carpenter performing four specific tasks. Second, subjects completed a carpentry knowledge questionnaire. Third, subjects were asked to view three short videotaped performances of a woodworking student portraying a carpenter while they worked on a task. Fourth and finally, subjects completed a recognition form and a ratee performance appraisal measure.

Hunt and Lansman (1982) pointed out that there are limited attentional resources to be divided among the multiple tasks required of raters. The attention given to behaviors during their actual occurrence may affect the later recall of behavioral incidents. Raters thus can not attend to all available stimuli but must be selective in their perception (Klatzky, 1980). Schneider, Hastorf, and Ellsworth (1979) point out that different attentional strategies between raters will most likely result in the differential salience of certain stimuli. Ebbesen (1980) stresses that behavior perception may not rely on a passive process but observers may selectively and actively attend to different features and aspects of the ongoing behavior.

Bodenhausen and Lichtenstein (1987) found that when subjects have a complex task to perform, they tend to adopt a heuristic strategy to handle large amounts of information. Also, if attention and information processing capabilities are limited as pointed out by previous literature (Crowder, 1976; Hunt & Lansman, 1982; Klatsky, 1980), we would predict that individuals who are exposed to multiple tasks (therefore not being able to direct all their attention to the task) would collect fewer critical incidents from videotapes than individuals who could direct all their attention to the videotapes.

Previous research has shown that raters have limitations in attention processes and memory capacity (Hogarth, 1980; Klatzy, 1980). Specifically, how many behaviors and the type of behaviors a rater can attend to are influenced by the variability of the behaviors which the rater views (Crowder, 1976; Hunt & Lansman, 1982; Klatzy, 1980). The first factor explored is information processing capacity. Based on this literature, the following hypothesis is made:

> Hypothesis 1: Raters, regardless of the level of knowledge, who have a competing task to complete while observing the videotaped lectures and prior to performance rating will make less accurate ratings than raters without a task to complete.

The knowledge level of the rater may affect performance ratings. Previous research has indicated tha high job knowledge raters were more responsive to observed performance covariation whereas low job knowledge raters were more responsive to their internal conceptual similarity schemata when rating a familiar ratee (Kozlowski & Ford, 1991; Kozlowski, Kirsch, & Chao, 1986; Kozlowski & Kirsch, 1987). Also, high job knowledge raters tended to rely on acutal perfomrance covariation when rating familiar ratees but conceptual similarity ratings when rating unfamiliar ratees.

Do raters use specific information in making a rating or do they form an evaluation based on categorization? It has been previously thought that larger samples of performance relevant ratee behavior inherent in familiar ratees would lead to more reliable ratings (e.g. Ferguson, 1949; Hollander, 1954, 1957; Wherrry & Bartlett, 1982). Borman (1978) also supported the assumption that raters are thought to observe, store, and recall specific ratee behaviors with judgments resulting from an integration among the recalled behavioral items. In contrast, Ilgen and Feldman (1983) put forth a categorization model which asserts that raters assign ratees to cognitive categories based on salient ratee features prototypic of the category. The raters access behaviors consistent with the prototypic level of performance for that ratee and not the specific behavioral information of that particular ratee.

Kozlowski and Kirsch (1987) suggest that both models may be operative in a performance appraisal situation. Initially, ratrs need specific behavioral information in order to assign the ratee to a category. After assignment, the category and its associated behaviors are accessible. The rater then accesses behaviors consistent with the category. Kozlowski and Ford (1988) examined the effects of familiarity, constraint, and memory on rater information acquisition strategies. They hypothesized that with an increasing delay between raters' exposure to prior information and rating, memory for specific information would weaken. Thus, when raters attempt to seek specific information for integration beofre judgment, they would search more and search with an increasing delay. Kozlowski et al (1988) found that ratee familiarity had an impact on search. Their results indicated that raters search for more ratee information when search was unconstrained. Second, raters searched more with low performing ratees when a great deal of prior ratee information had been provided when the level of ratee familiarity was high. Further, raters exhibited less search when there was more prior information available, and finally the no delay raters tended to search for more information for low prior information ratees relative to the delay raters.

Finally, raters possess their own heuristics which aid in processing information quickly . People are limited in the amount of information they can process (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Bruner (1958) pointed out that categories are basic to perception, information storage and organization. Despite built-in storage limitations, people use categories to process large quantities of information (Smith, Adams, & Schorr, 1978). Individuals tend to sort and classify the people they meet into categories or groups based on their own similarity to that individual's personal system of categorization. These implicit theories affect all stages of information processing. Previous research has shown that prototypes and categorization affect initial impressions (Fiske & Taylor, 1991; McCloskey & Glucksberg, 1978), encoding (Zadny & Gerard, 1974); and retrieval (Cantor & Mischel, 1977).

Cantor and Mischel (1977) hypothesized that central category attributes become the focus on attention in judgments made on the basis of social information. Judgments of similarity to a prototype are based on the family resemblance criterion. Specifically, no single set of features defines the category but rather any of several features contribute to the perception that an object resembles the category prototype. An index of family resemblance can be used to indicate an individual's personal prototype involving how much a specific attribute or characteristic has in common with a ctegory structure. The obtained family resemblance index can be used to indicate the normativeness (i.e. 'standardness') of an individual's category system (Lord, Foti, & DeVader, 1984; Rosch, 1977, 1978;

Rosch & Mervis, 1975). Based on the preceding literature, the following hypothesis is made:

Hypothesis 2: A rater's personal prototypes and level of knowledge will influence the level of recognition accuracy.

Hypothesis 2 (a): Raters with higher job knowledge will recognize more behaviors consistent with their personal prototype and falsely recognize more behaviors from their personal prototype than lower job knowledge raters.

2 (b): Lower job knowledge raters will be more accurate in their recognition of ratee behaviors than would higher job knowledge raters.

Hauenstein and Kovach (1988) showed that familiar raters recalled fewer specific ratee behaviorss, made more dispositional inferences, and were slightly less accurate at recognizing behavioral incidents than did unfamiliar raters. Familiarity in this study referred to the possession of a set of richly defined, well-differentiated cognitive structures associated with a stimulus domain (Lord & Maher, 1990; Markus, Smith, & Moreland, 1985). Familiar raters used a categorizational process whereby memories of ratee performance was organized around trait inferences generated from ratee behaviors. Unfamiliar raters, on the other hand, used a behavioral level processing strategy in which memories were organized around what the ratee looked like and what the ratee did.

Foti and Hauenstein (1990) looked at the effect of increasing processing demands on the rater in the performance appraisal setting. They found that in the minimal demands condition, although raters' memories were biased toward recognizing behaviors consistent with their initial impression, their performance ratings were behviorally based. However, as processing demands increased, subjects tended to focus more on schema inconsistent behaviors resulting in better memory for inconsistent behviors, but not to the extent that they changed their impression significantly. Perhaps, as raters become more taxed during observation, the less impact the schema inconsistent behviors appear to have on the final impression (White & Carston, 1983).

One final study focusing on processing demands is Hauenstein and Walker (1988). They looked at the effects of judgment complexity and job knowledge. The authors hypothesized that when the judgment tasks was low in complexity, job knowledge will not moderate the memory-appraisal accuracy relations, and that memory will be independent of appraisal accuracy. However, when the judgments task is high in complexity, the relationship between memory and judgment will vary as a function of job knowledge. Their data confirmed these hypotheses. From this research, the following hypothesis is made:

Hypothesis 3 (a): Raters with higher job knowledge will falsely recognize behaviors consistent with personal prototype when there is a competing task than when there is no competing task.

Hypothesis 3 (b): Higher job knowledge raters will make less accurate performance appraisal ratings because of their reliance on their personal prototypes under increased processing conditions, whereas they will be no such difference for lower job knowledge raters.

Method

Part 1

Subjects

There were 80 subjects in this study. The higher knowledge subjects were primarily experienced carpenters, architectural millworkers, and supervisors in construction. The experienced carpenters had worked in rough and finish carpentry in industrial and residential settings. The architectural millwork subjects were employed in a lumber company and were involved in design, production, and finish of millwork. There were a total of 43 higher knowledge subjects with four architectural millworkers, 28 carpenters, and six carpentry supervisors (carpenters who were at the supervisory level). The five remaining higher knowledge subjects were employed in the construction business but not as carpenters. The 37 lower knowledge subjects were employed as drivers and/or laborers in the construction industry. All subjects were employed in Southern Virginia. The knowledge level of the subjects was determined via a 22-item multiple choice job knowledge test on carpentry.

Free Response

Previous studies (Lord, Foti, & DeVader, 1984) have proposed that free responses are useful in generating schema information. Subjects were individually questioned concerning their idea of a good, average, and poor carpentry performer in relation to four tasks, sanding, sawing, staining, and hammering. Each subject was questioned, by task, to indicate how a good vs. average vs. poor performer would complete these tasks. The performance level order was counterbalanced to control for order effects. The responses of the subjects were audiotaped and later transcribed.

Level of Knowledge Exam

Level of knowledge for all subjects was determined by a carpentry exam developed by Walker (1989). A 22-item multiple choice questionnaire inquiring about different aspects of trim, rough, and finish carpentry was administered to all subjects (see Appendix A for level of knowledge questionnaire). A subject matter expert aided in the development of the test items in the Walker (1989) knowledge questionnaire. Test-retest reliability was used to assess the reliability of the carpentry knowledge exam. Subjects were given the knowledge exam twice: (1) when the free recall data was generated and (2) when the videotape of carpentry performance was viewed.

Design and Procedure

All subjects were administered a consent form and were told by the experimenter that their participation was strictly voluntary (see Appendix B for consent form). Upon completion, the experimenter collected the consent forms and then either administered the knowledge test or the free response interview. The order of completion for the level of knowledge exam and the free response interview was counterbalanced.

Method

Part II

Subjects

The subjects in Part II were the same subjects as in Part I.

Stimulus Materials

The occupation of carpenter was utilized as the prototypic profession. The performance levels chosen were used to maximize the differences between the subject's personal prototype. The carpentry tapes were a subset of those used by

Williams, DeNisi, Meglino, and Cafferty (1986) and consisted of twelve individual performance segments showing three workers performing four tasks (sawing, sanding, hammering, and staining). Each ratee's performance segment lasted approximately two minutes, and the total length of the carpentry tape was approximately ten minutes. The carpenters in this tape consisted of paid male volunteers from an advanced woodworking class at a technical school in upstate New York (Williams et al, 1986).

Three levels of worker proficiency were established in these tapes. The different levels of performance consisted of one worker performing 3 out of 4 tasks correctly (75% proficient), one worker performing 2 out of 4 tasks correctly (50% proficient), and one worker performing 1 out of 4 tasks correctly (25% proficient). The videotapes were pretested by Williams et al (1986). Subjects in the Williams et al study rated the 75% proficient worker significantly higher than both the 50% proficient worker and the 25% proficient worker, but the latter two did not differ significantly. However, these ratings were made two days after observation of ratee performance. Using a sample of carpenters, Walker (1989) found that mean ratings for the average carpenter were not significantly higher than that of the good carpenter using the same carpentry videotapes. However, the differences in the means were in the correct direction. The carpenters did correctly differentiate between the average and poor carpenters. For the present study, information was presented blocked by person. Subjects viewed, in order, a 50% proficient worker, and then either a 25% proficient worker followed by a 75% proficient worker, or a 75% proficient worker followed by a 25% proficient worker performing all four woodworking tasks. This order was used to establish an anchoring effect.

Independent Variables

Competing Tasks. The purpose of the task was to divide the subjects' attention, not occupy the subject's full attention. Higher knowledge subjects in the competing task condition were given a task which required them to solve three applied mathematical problems similar to what carpenters might encounter on the job (see Appendix C). Lower knowledge subjects were given a task in which they had to think of other uses of two common objects (see Appendix D). The rationale for two competing tasks was that higher knowledge task would have more face validity. In contrast, the lower knowledge subjects would find this carpentry relevant task too difficult and thus be unable to attend to the videotape. Lower knowledge subjects were thus given a task which allowed viewing of the tape but also divided the subjects' attention. The competing task group were allowed to read the task instructions prior to viewing the videotape. Subjects then viewed the videotape while completing the task. Subjects were allowed to read the instructions prior to the videotape viewing because reading during the videotape might have led to the subjects' full attention being directed toward the scenario, thus not allowing the subjects to encode any stimulus information from the videotape.

Subjects were told that the information from the task would be used to see how they go about solving tasks similar to ones they encounter on the job.

Personal Prototype. The free response interview was used to obtain each subject's personal prototype for good, average, and poor carpentry performance and this information was used to compute each subject's family resemblance score. A trait was considered a personal prototype for a subject if it appeared in one of the three descriptions of the subject's free response generation of their idea of a good, average, and poor carpenter performance. As stated by Rosch and Mervis (1975), if the 'categorical' of these categories is to be explained, it appears most likely to reside in family resemblance between members. Rosch and Mervis (1975) found evidence which supports the hypothesis that the more an item has attributes in common with other members of the category, the more it will be considered a good and representative member of that particular category.

Based on the procedure from Rosch (1978) and Lord, Foti, and DeVader (1984), family resemblance scores were computed for each subject. Previous researchers (Lord et al, 1984; Rosch, 1978) used the family resemblance scoring procedure to average across individuals whereas in this study the family resemblance scoring procedure was used to generate a family resemblance score for each individual. The basis for the computation of family resemblance scores indicating personal prototypes involved relating how much a specific attribute or characteristic had in common with a category structure (i.e. performance levels). This was then used to examine how standard individuals were in rating subordinates by focusing on the normativeness of their personal categorization system. Family resemblance scores were calculated for each trait and behavior using the following procedure. First, each characteristic or trait was listed along with the frequency counts for each characteristic or trait. All synonyms in the list were checked and then grouped under the attribute heading. Second, to determine whether an attribute or behavior from the free recall list was to be included, it must have been mentioned by at least 1 subject out of 10 (i.e., 10% of the subjects would have to have mentioned it). Third, each attribute received a family resemblance score. This score was the number of behaviors mentioned by a particular subject for a specific behavior (hammering, sanding, sawing, or staining) at a particular performance level divided by the total number of behaviors across all subjects for that particular behavior and performance level which met the above criteria of being mentioned by at least 10% of the subjects. Fourth and finally each individual's family resemblance item scores across all behaviors and performance levels were added and divided by the total number of items to obtain an overall family resemblance score per subject. An example of the computation of family resemblance score follows: subjects X lists 'sands against the grain' and 'hits nail on first attempt' as the two behaviors in his/her personal prototype. Sands against the grain was listed under this subject's prototype of poor performance-sanding and received a family resemblance score of 33 (1/3 = .33) since there were three poor behaviors for sanding across all subjects which met the criteria for inclusion and this particular subject mentioned only one behavior which was also one of the three poor sanding behaviors. Hits nail on first attempt was listed as an example of this subject's prototype of good performance-hammering and received a family resemblance score of .25 (1/4 = .25) since there were four behaviors which met the criteria for inclusion and this subject only mentioned one behavior for good-hammering. The family resemblance score for Subject X was thus the average of these two items mentioned in the individual's prototype (((.33 + .25)/2) = .29). This scoring procedure was similar to the past research of Lord, Foti, and DeVader (1975), Rosch (1978), and Rosch and Mervis (1975).

Level of Knowledge Exam. All subjects completed a level of knowledge questionnaire in Part I of the study. The score on this questionnaire was used to generate the level of carpentry knowledge for that particular subject rater.

Dependent Variables

Critical Behavior Recognition Form. A recognition form was constructed from a content analysis of the free recall material obtained from Part I and the carpentry performance videotape (see Appendix E). The construction of the recognition form consisted of four sets of items: (1) items present in the videotape and present in the free recall interview; (2) items present in the videotape and absent in the free recall interview; (3) items absent in the videotape but present in the free recall interview; and (4) items absent in the videotape and in free recall interview. All items were written at a high level of behavioral specificity. Subjects responded by answering "yes" for occurring behaviors and "no" for behaviors that did not occur (see Appendix E for Critical Behavior Recognition Form). The scoring of the recognition form was as follows: (1) hit: a behavior which was present in the videotape for a particular performance level and to which the subject responded 'yes' on the recognition form and was present in their free recall of personal prototypes; (2) miss: a behavior which was present in the videotape for a particular performance level and to which the subject responded 'no' on the recognition form and was absent in their free recall of personal prototypes; (3) correct rejection: a behavior which was absent in the videotape for a particular performance level and to which the subject responded 'no' on the recognition form and was absent from the subject's free recall of personal prototypes; (4) false alarm: a behavior which was absent in the videotape for a particular performance level and to which the subject responded 'yes' on the recognition form and was present in the subject's free recall of personal prototypes; and (5) the total for each of these categories was obtained by summing the number of hits, misses, correct rejections, and false alarms.

Accuracy was determined using signal detection theory. Signal detection theory (Srull, 1984) states that familiarity values associated with each item are determined by recency and frequency of past exposure. Subjects can indicate a signal is present when there is actually a signal (hit), indicate a signal is present when there is actually none (false alarm), indicate that no signal is present when there is none (correct rejection), or indicate that there is no signal when one is actually present (miss). The advantage of this method is that a subject's overall performance can be summarized by plotting the probability of making a hit as a function of the probability of making a false alarm. Perfect accuracy is when a subject demonstrates a hit rate of one and a false positive rate of zero. The lowest level of accuracy is chance and takes place when the subject's true hit rate is less than or equal to his/her false positive rate. Values for A' range from .5 (recognition at chance) to 1.00 (perfect recognition accuracy). The probability of a hit was determined by dividing the total number of hits by the total number of possible hits for each individual. The total number of hits (the total number of behaviors which were present in the subject's personal prototype, present in the videotape for a particular performance level, and to which the subject responded "yes" on the recognition form) were divided by the total number of possible hits (the total number of behaviors present in the subject's personal prototype, present in the videotaped performance, and present in the recognition form). For example, subject A had his/her total number of hits for average performance level as three (total number of behaviors present in the subject's personal prototype, present in the videotape and to which the subject responded "yes" on the recognition form) and subject A had a total of four possible hits (behaviors which were present in the personal prototype, present in the videotaped performance of the 50% (average) proficient worker, and present in the recognition form). The probability of a hit for subject A was 3/4 or .75. The rationale for using this probability was each subject's free recall prototype responses were linked to the recognition form and thus there was a direct connection between the recognition form, the free recall prototype for each performance level, and the videotaped performance. A similar computation scheme was used for correct rejection probability. Each total number of correct rejections (the total number of items which were absent in the subject's free recall of their personal prototype, absent in the videotaped performance and to which the subject responded "no" to the occurrence of the behavior on the recognition form) were divided by the total number of possible correct rejections (behaviors absent in the subject's personal prototype and absent in the videotaped performance but present in the recognition form). The computation of false alarm rate was simply 1 - the probability of a correct rejection. The computation of the probability of a miss was 1 - probability of a hit. Each of these probabilities were computed per performance level (good, average, and poor ratee).

Performance Rating. Raters provided a performance appraisal of each ratee on five performance dimensions using a 5-point rating scale (see Appendix F for Performance Rating Form). The ratee's performance measures were matched to each of the true score ratings.

Manipulation Check. There were a total of three manipulation checks. First, a manipulation check was provided for the order effects of the free recall interview and the knowledge questionnaire on family resemblance scores. Second, a manipulation check was provided for the order effects of the videotaped performances (Average/Good/Poor or Average/Poor/Good) on recognition accuracy, false alarm rate and rating accuracy. Third, and finally, a manipulation check for the extent to which the subject felt their attention was divided due to the task was provided. This manipulation check was included in the second session after they completed the recognition form and graphic rating scale form (see Appendix G for manipulation check items).

Generation of True Scores. The true scores for the judgments of ratee performance were obtained from Walker (1989) (see Table 9). In this study, as noted in Borman (1977), mean ratings computed over a number of expert judges produces a true score measure of a specific ratee's performance. Murphy and Balzer (1986) assert that this mean approximates the expected value of the rating obtained from an expert observer under optimal conditions.

Carpentry subject matter experts were briefed about their rating task (Walker, 1989) in compliance with the guidelines of Sulsky and Balzer (1986). The subject matter experts were shown the videotapes without exposure to the rating scale. The chief investigator (Walker, 1989) began conversations concerning what each individual carpenter thought were the most salient and representative dimensions of carpentry performance seen on the tape. The subject matter experts were then run separately due to scheduling constraints with efforts taken to ensure consistency across experts on the instructions about how to observe the ratees and use the rating scales. The subject matter experts were also informed about the different types of rating errors (e.g., leniency, severity, central tendency, halo) and how these errors may influence their ratings. Subjects were also given opportunity to ask questions concerning rating scales and possible rating errors. The videotapes were then shown to the carpenters a second time with the knowledge that they would have to make more behavioral and judgmental tasks afterwards.

<u>Procedure</u>.

Level of knowledge of the subjects was generated from the score obtained on the knowledge questionnaire. Each group member was randomly assigned to either a competing task or no competing task condition and were run in groups of 1-4 individuals. Subjects were told that the experimenter was attempting to obtain an idea of how carpenters and construction workers rate carpentry performance and what behaviors they consider important. Subjects in the competing task condition were told that it was necessary that they work on a task while viewing the performance of the videotaped lectures in order to simulate their work experience.

After viewing the videotape, subjects completed a recognition form consisting of the critical behavior list,a performance appraisal questionnaire, and the knowledge questionnaire for reliability purposes.

Results

Manipulation Checks

In order to assess the order effects due to the free recall interview and knowledge questionnaire, a manipulation check was conducted with family resemblance scores as the dependent variable. Results of a one-way analysis of variance (ANOVA) with family resemblance scores as the dependent variable was provided with no significant differences in family resemblance scores found due to order effects (see Table 1).

Further, a manipulation check was provided for the possible order effects due to the videotape performance order. Subjects either viewed a performance triad of Average/Good/Poor or Average/Poor/Good. The average performance was presented first as an anchoring effect. Four analyses of variance (ANOVA), one for each component of accuracy (elevation, differential elevation, stereotype accuracy, and differential accuracy), were computed with the components of accuracy as the dependent variables and the performance order as the independent variable. The results are contained in Table 2 with associated non-significant findings due to order. Analyses of variance (ANOVA) were also computed with A' and false alarm rate as the dependent variables and videotaped performance order as the independent variable. The data indicated non-significant effects due to order (see Table 3).

Table 4 indicates the means and standard deviations for the two items assessing the competing task manipulation. Item 1 asked subjects "how well could you pay attention to the carpenter's behaviors in the videotape?". Responses ranged from 1, "not well" to 9, "very well". Results of a one-way analysis of variance (ANOVA) showed significant differences between the two task conditions $F(1,78)=6.09$, $p<.01$ (see Table 5). Thus, subjects in the competing task condition

(M=4.51) reported having their attention divided as compared with the subjects with no competing task (M=5.61) to complete with lower numbers indicating more division of attention.

Item 2 asked subjects on a 9-point likert scale, "how difficult was it to remember what each carpenter did in the videotape?". Results of a one-way analysis of variance (ANOVA) indicated that the subjects with the competing task (M=3.95) differed in their reported ability to recall behaviors in the videotape as compared with subjects without a competing task (M=5.76), $F(1,78)=20.06$, $p<.0001$ (see Table 5).

These findings thus demonstrate that subjects, with a competing task to complete while viewing the videotape, reported having difficulty in paying attention to the carpenter's behaviors in the videotape and reported finding it difficult to remember to remember what each carpenter did in the videotape.

Family Resemblance

Family resemblance scores were calculated for each trait and behavior using the following procedure. First, each characteristic or trait was listed along with the frequency counts for each. All synonyms in the list were checked and then grouped under the attribute heading. Second, to determine whether an attribute or behavior was to be included, it had to be mentioned by at least 1 subject out of 10 (i.e., 10%). Third, each attribute received a family resemblance score which was the number of behaviors mentioned by a particular subject for a specific behavior (hammering, sanding, sawing, or staining) at a particular performance level divided by the total number of behaviors across all subjects for that particular behavior and performance level which met the above criteria of being mentioned by at least 10% of the subjects. Fourth and finally, each individual's family resemblance item scores across all behaviors and performance levels were added and divided by the total number of items to obtain an overall family resemblance score per subject. An example of the computation of a family resemblance score is in the Method Section, Part II, Independent Variables, Personal Prototype. Table 6 contains the carpentry behaviors and their corresponding frequency counts. Table 7 contains descriptive statistics for family resemblance scores with level of knowledge level being treated as a dichotomous variable with high knowledge subjects being those scoring 15 and above and low knowledge scoring 14 and below. Table 7 also contains class intervals for knowledge scores. A correlation analysis was computed between family resemblance scores and knowledge scores. The obtained Pearson Product Moment Correlation was .530 (see Table 8). Family resemblance scores ranged from .21 to .97.

Reliability of Level of Knowledge Questionnaire

Level of knowledge was assessed by a 22-item questionnaire (see Appendix A for Level of Knowledge Questionnaire). The highest possible score for the exam was 22 with the range of scores being 3-22. A test-retest reliability analysis resulted in reliability coefficient of .97.

Hypotheses

Performance Ratings. Hypotheses 1 stated that raters, regardless of the level of knowledge, who have a competing task to complete while observing the videotaped lectures and prior to performance rating would make less accurate performance ratings than raters without a task to complete with the dependent variable being the performance rating. The true scores, by behavior and ratee performance level for the graphic ratings of carpentry performance are shown in

Table 9. The means for the behavior rating for each behavior and each performance level are contained in Table 10.

The variance components of rating accuracy (elevation, stereotype accuracy, differential elevation and differential accuracy) were computed (see Appendix H for component formulations). Sulsky and Balzer (1986) recommended the variance measures since these measures actually take into account the distance between subjects' ratings and true score ratings. Table 11 contains descriptive statistics for the entire sample. Table 12 contains descriptive statistics for the different components of accuracy by level of knowledge. Table 13 contains the descriptive statistics for the different components of accuracy by presence/absence of competing task.

Four one-way analyses of variances, one for each component of accuracy with presence or absence of competing task as the independent variable and each component of accuracy as the dependent variable were computed to test hypothesis 1 with level of knowledge as the covariate. A significant main effect for competing task would have supported the hypothesis. Overall, none of the analyses of variance reached an adequate significance level (see Tables 14-17). Thus, no main effect for presence/absence of competing task was found.

Summary: Performance Ratings.

As noted previously, the manipulation check revealed the presence of a competing task appeared to divide the subject's attention as compared with the group of subjects who did not receive the competing task. Also, subjects with a competing task appeared to have more difficulty recalling the behaviors seen in the videotape. However, the hypothesis that subjects with a competing task would make less accurate ratings than subjects without a competing task was not supported.

Recognition Accuracy

Hypothesis 2 (a) stated that raters with higher job knowledge would falsely recognize more behaviors consistent with their personal prototype than lower knowledge raters. Hypothesis 2 (a) was examined using false alarm rate. The method for computation of false alarm rate was discussed previously in the Method section (see Method Section Part II, Dependent Variables, Critical Behavior Recognition Form). Descriptive statistics for false alarm rate and correct rejections are contained in Table 18. Descriptive statistics for false alarm rate with level of knowledge dichotomized are presented in Table 19. Descriptive statistics for false alarm rate with the subjects sorted by presence/absence of competing task are contained in Table 20. A multiple regression with level of knowledge as the covariate and family resemblance scores as the predictor and false alarm rate as the dependent variable was computed to test hypotheses 2 (a). Hypothesis 2 (a) stated that raters with a higher job knowledge would falsely recognize more behaviors consistent from their personal prototype than lower knowledge raters. A significant main effect for family resemblance scores would support hypothesis 2 (a). The results with average performance level were significant with $F(2,77) = 8.011$, $p<.007$. The data indicated a significant main effect due to family resemblance scores (see Table 21). The results with poor and good performance levels were however non significant (see Tables 22-23).

Hypothesis 2 (b) stated that lower job knowledge raters would be more accurate in their recognition of actual ratee behaviors than would higher knowledge raters. Specifically, lower job knowledge raters would be more accurate in their recognition of ratee behaviors than would higher job knowledge raters. The nonparametric accuracy index, A', was computed. The computing expression for this index is given by Grier (1971) (see Appendix H). Grier asserts the advantage of using A' is that it preserves the sensitivity and bias inherent in data. These indices are derived from signal detection theory based on each subject's memory operating curve (MOC). Larson, Lingle, and Scerbo (1984) stated than an MOC curve is a theoretical curve derived on the basis of a subject's hit and false alarm rates. A' is a recognition memory accuracy index and false alarm rate is an indication of subject's tendency to mark a behavior as occurring in the videotape after viewing the videotape when actually the behavior had not occurred in the videotape. Descriptive statistics for A' and misses are contained in Table 24. Descriptive statistics for A' with subjects sorted into higher and lower knowledge subjects are presented in Table 25. Descriptive statistics for A' with the subjects sorted according to presence/absence of competing task are contained in Table 26. A multiple regression with family resemblance scores as the predictor and level of knowledge as the covariate was computed with the dependent variable being A'. The results did not reach significance (see Tables 27-29). The F value was significant with the average performance level but the corresponding regression weight for family resemblance scores was non-significant.

Summary: Recognition Accuracy

The results indicated that when rating average performers, high knowledge raters falsely recognized more behaviors consistent with their personal prototype than lower knowledge raters when rating average performers. This hypothesis was not supported when rating good and poor performers. Also, the hypothesis that lower knowledge raters would be more accurate in their recognition of ratee behaviors than higher knowledge raters was not supported.

Memory and Appraisal Accuracy

Hypothesis 3 (a) stated that higher knowledge raters would falsely recognize behaviors consistent with their personal prototype when there was a competing task than when there was no competing task. A multiple regression with presence/absence of competing task and family resemblance scores as the predictors and level of knowledge as the covariate was computed to test this hypothesis with false alarm rate as the dependent variable. A significant three-way interaction among competing task, level of knowledge, and family resemblance score would have supported Hypothesis 3 (a). This hypothesis was not supported (see Tables 30-32). The F associated with viewing average performance level was significant but the associated three-way interaction with competing task, family resemblance scores, and level of knowledge was not significant (see Table 30).

Hypothesis 3 (b) stated that higher job knowledge raters would make less accurate performance appraisal ratings whereas there should be no such difference for lower job knowledge raters. Four multiple regressions were computed, one for each component of accuracy as the dependent variable with family resemblance scores and presence/absence of a competing task as the predictors and level of knowledge as the covariate. A significant three-way interaction between job knowledge, family resemblance scores, and presence/absence of competing task would have supported hypothesis 3 (b). The three-way interactions did not reach significance (see Tables 33-36).

Summary: Memory and Appraisal Accuracy

These results suggest that high knowledge raters did not make more false alarms from their personal prototype under increased processing conditions. Further, high knowledge raters do not rely more on their personal prototypes and thus make less accurate ratings under increased processing conditions as compared with low knowledge raters.

Discussion

The present study tested the hypotheses that rater knowledge, personal prototypes and presence of a competing task would influence rating accuracy and recognition accuracy. Hypotheses for these three variables were based on both empirical and theoretical work in the areas of applied psychology, information processing and social cognition. Little support for the hypothesized effects were found. First, it was predicted that raters, regardless of their level of knowledge who have a competing task to complete while observing a videotape of carpentry performance would make less accurate rating than raters without a task to complete. This was not supported by the data. Further, it was hypothesized that higher job knowledge raters would falsely recognize more behaviors consistent with their personal prototype than lower knowledge raters. This hypothesis was partially supported by the data with false alarm rate as the dependent variable when viewing average performance level. Further, it was hypothesized that lower job knowledge raters would be more accurate in their recognition of ratee behaviors than higher job knowledge raters. Specifically, lower job knowledge raters would be more accurate in their recognition of ratee behaviors than would higher job knowledge raters. The results did not support this hypothesis. Further, it was hypothesized that higher job knowledge raters would falsely recognize behaviors consistent with their personal prototype when there was a competing task than when there was no competing task. This hypothesis was not supported by the data. Finally, it was hypothesized that higher job knowledge raters would make less accurate performance appraisal ratings because of their reliance on personal prototypes under increased processing conditions whereas there would be no difference for lower knowledge raters. This was also not supported.

Manipulation Effectiveness.

The manipulation of competing task was successful. The presence/absence of the task appeared to divide the subjects' attention to the videotaped performances. Subjects in the competing task group reported having difficulty paying attention to the videotape and recalling what each carpenter did in the videotape. The subjects were selectively given one of two competing tasks based on their level of knowledge. Subjects who obtained high knowledge scores and had experience in carpentry were given a carpentry task similar to a task which they might encounter on the job whereas subjects who were less experienced and less knowledgeable in carpentry were given a simple task which the subjects were able to complete while still being able to attend to the videotape. The manipulation effectiveness of the competing task was indicated by significant ANOVAs. However, the absolute difference in the means between the competing task/no competing task was small. Perhaps the subjects had difficulty recalling behaviors and difficulty paying attention since the corresponding means were not that disparate for the competing task to influence rating accuracy levels.

The following five sections will focus on the possible reasons and implications for the results of the aforementioned hypotheses.

Performance Ratings

Hypothesis 1 focused on the effect a competing task would have on rating accuracy. The hypothesis that a competing task could make a meaningful difference in rating accuracy was not supported by a main effect for the presence of a competing task. The manipulation check did demonstrate significant differences between the competing task and no competing task group in terms of reported ability to attend to the videotape and reported ability to recall behaviors in the videotape. The absence of an effect due to the presence/absence of competing task has several possible explanations. One, the nature of the variability of the distribution of the four components of accuracy may have contributed to the non-significant results due to range restriction in the components of accuracy. As shown in Table 13 it can be seen the majority of subjects, regardless of the presence/absence of competing task, were fairly accurate across components of accuracy except for differential accuracy. Since a smaller number denotes higher levels of rating accuracy, it should be noted that the differences between the competing task and no competing task groups were in the correct direction with subjects without a competing task exhibiting higher levels of rating accuracy.

Perhaps another explanation for the lack of a difference between the two groups was the nature of the videotape and the occupation of carpenter. The videotaped performances of sanding, sawing, hammering and staining consisted of four discrete behaviors blocked by subjects. In the field, raters would have more of an opportunity to see ratees perform over many occasions and be more likely to form an impression based on several exposures to the ratee. However, in the field, raters would not necessarily see information blocked by person. Typically, supervisors would see one ratee perform one behavior and then another ratee performan another behavior and so on. In this experiment, subjects were given a limited amount of information in a short amount of time and then given a rating task based on this videotaped performance. One of the differences between this study and the Kozlowski, Kirsch, and Chao (1986) study was that ratee familiarity was the same for all subjects in this study whereas in the Kirsch et al study (1986), ratee familiarity was either high or low. Perhaps, having different levels of job knowledge but with similar levels of ratee familiarity served to mask some of the differences present in the raters.

Recognition Accuracy

Hypothesis 2 was concerned with the rater's personal prototype and level of knowledge and the subsequent effect that these variables would have on recognition accuracy. Hypothesis 2 (a) stated that raters with higher job knowledge would falsely recognize more behaviors consistent with their personal prototype than lower knowledge raters. False alarm rate was the dependent variable used to test hypothesis 2 (a). The data for hypothesis 2 (a) demonstrated mixed findings. Hypothesis 2 (a) was supported when the subjects viewed the average performer. This may have been due to the difficulty in remembering the behaviors of the average performer. The differences between the behaviors of a good and poor performer are fairly salient whereas the average performer exhibits both good and poor instances of job behavior and therefore making recall and recognition more difficult. Hypothesis 2 (b) stated that raters with lower job knowledge would be more accurate in their recognition of ratee behaviors than higher knowledge raters. A' was the dependent variable for hypothesis 2 (b). This hypothesis was not supported.

Memory and Appraisal Accuracy

Hypotheses 3 (a) and (b) dealt with memory sensitivity and appraisal accuracy and the effect knowledge level and personal prototypes would have on rating accuracy and memory sensitivity. Hypothesis 3 (a) stated that higher job knowledge raters would falsely recognize more behaviors from their personal prototype when there was a competing task than when there was no competing task.

This hypothesis was not supported by the data. It should be noted that false alarm rate was higher in the competing task group (see Table 20).

Hypothesis 3 (b) also was not supported in that higher job knowledge did not make less accurate performance appraisal ratings because of their reliance on personal prototypes under increased processing conditions. This non-significant finding also may be related back to the absolute differences in the means in the competing task manipulation.

Limitations and Future Implications.

One possible explanation for the lack of significant results may have been due to not controlling for intelligence level of the raters. As indicated by Smither and Reilly (1987) and Hauenstein and Alexander (1991), rater intelligence plays an important role in rater accuracy. Perhaps intelligence could have been measured and controlled in this study.

What might have occurred in this study is that the majority of the subjects were ones who possessed similar personal prototypes of performance levels which masks any differences inherent in their rating ability and memory sensitivity

Although, the findings of this study were largely non-supportive of the hypotheses, implications for future research can be derived by focusing on a three part conceptualization of rater's personal prototypes, rater intelligence, and rater job knowledge. First, perhaps the measure of personal prototypes could have been altered. This study used free recall interviews to access personal prototypes and then used family resemblance scores as an indicator of these personal construct systems. Future studies could use the Borman (1987) trait implication procedure to indicate personal construct systems. This method involves having subjects rate the similarity between each of his/her constructs and a number of reference concepts.

The subjects take the reference constructs and rate on a 5-point scale the similarity between each of their own personal work constructs and each of the reference constructs. There are four categories of reference constructs in Borman's procedure: personal characteristics and personal traits, cognitive/physical abilities, performance constructs for enlisted soldiers, and military leadership constructs. An example of each of the respective categories are: energy level, intelligence, staying out of trouble, and supporting other unit members. Similarity judgments along with reference concepts make up a numerical definition of that construct.

Another method of studying personal prototypes lies with the research of Jackson and colleagues. These researchers (Jackson, 1972; Reed & Jackson, 1975; Rothstein & Jackson, 1980) posit that the rationale for studying personal prototypes lies with the relationship between individual differences in person categories and rating ability. Jackson and colleagues have pointed out that the inferential accuracy model research has over time demonstrated reliable individual differences in implicit theories of raters. The inferential accuracy model states that there are two distinct processes which underlie the implicit conceptions of behavioral covariation and these processes are sensitivity and threshold. Simply stated, inferential accuracy is defined by a person's (rater's) ability with limited information available about the target (ratee) to judge other relevant characteristics about that person correctly and to identify behavioral exemplars as part of a pattern of behavioral consistencies (good performer). Sensitivity is the individual differences in the awareness of shared implicit network of behavioral consistencies. Threshold is defined by individual differences in terms of readiness to attribute behaviors to others based on the implied relations among behaviors. Perhaps this study could have looked at threshold (determined by the estimation of the average rating given each ratee by each rater) and sensitivity (estimation by Pearson Product Moment correlation between individual judgments of the inferential relations between behaviors with regard to ratee and covariation that also assessed the target's characteristics).

Second, future studies could measure intelligence and look at its effect on the content of rater's personal prototypes. Smither and Reilly (1987) hypothesized that rater ability or intelligence would be positively related to rater accuracy. These findings supported a curvilinear relationship between intelligence and accuracy in that the most intelligent raters tended to be somewhat less accurate than more moderately intelligence raters who were in turn more accurate than the least intelligent raters.

This role of intelligence in the process of rating accuracy was further studied by Hauenstein and Alexander (1990). They found that raters with moderate intelligence exhibited greatest accuracy for elevation,the differences between normative and idiosyncratic raters occurred only for raters with high or low intelligence. Also, only idiosyncratic raters exhibited the classic rating accuracy intelligence relationship. In addition, normative raters were more accurate than idiosyncratic raters at only low and moderate levels of intelligence. Finally, when controlling for intelligence, normative raters demonstrated stronger halo effects than idiosyncratic raters.

Also, referring back to previous literature (Hauenstein & Alexander, 1990; Jackson, 1972; Reed & Jackson, 1975; Rothstein & Jackson, 1980; Smither & Reilly, 1987), perhaps an examination of these variables, personal prototypes, rater intelligence, and rater job knowledge could be combined in an experiment in order to examine their multiple and combined effects on rater accuracy and rater memory sensitivity.

The above findings warrant some discussion since they do not support previous findings dealing with rater familiarity such as Kozlowski, Kirsch and Chao (1986), Kozlowski and Kirsch (1987), and Kozlowski and Ford (1988) and findings dealing with information processing such as Hauenstein and Kovach (1988), Foti and Hauenstein (1990) and Hauenstein and Walker (1988). Future studies in the laboratory and field settings should investigate other task characteristics which may influence the encoding and recognition of behaviors. More importantly, level of knowledge should be looked at as not simply knowledge of the job in question but intelligence level of the rater. Perhaps, it is not simply being familiar with the job which a rater appraises but also the basic ability level of the rater which influences their ability to rate accurately and to accurately recall behaviors performed by the ratee.

Summary and Conclusions.

Taking the above limitations aside, this study had several strong points. First, all ratees rated the performance of the ratees at the same time. In a typical appraisal context, raters complete appraisal during the same time of the year or month depending on the organization. Second, the rater population in this study was familiar with the profession which they rated. Third, this study examined personal construct systems in a method which had never been used before. Finally, this study was conducted in a field setting which typically has the usual distractions and interruptions present in a performance appraisal context. This study used a field sample with data collection in the field with the typical distractions present.

The main conclusion to be drawn from study data is that performance appraisal is a multifaceted process with many factors influencing the process.

References

Abelson, R.P. (1976). Script processing in attitude formation and decision making. In J.S. Carroll & J.N. Payne (Eds.), Cognition and Social Behavior. Hillsdale, NJ: Erlbaum.

Abelson, R.P. (1981). The psychological status of the script concept. American Psychologist, 36, 715-729.

Abelson, R.P., & Reich, C.M. (1969). Implicational molecules: A method for extracting meaning from input sentences. In D.E. Walker & L.M. Norton (Eds.), Proceedings of the International Joint Conference on Artificial Intelligence, May, 1969.

Anderson, J.R. (1980). Cognitive psychology and its implications. San Francisco: Freeman.

Anderson, J.R. (1981). Effects of prior knowledge on memory for new information. Memory and Cognition, 9, 249-277.

Anderson, J.R., Kline, P.J., & Beasley, C.M., Jr. (1979). A general learning theory and its application to schemna abstraction. In G.H. Bower (Ed.), The psychology of learning and motivation. New York: Academic Press.

Asch, S.E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 41, 1230-1240.

Balzer, W.K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. Organizational Behavior and Human Decision Processes, 37, 329-347.

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.

Bartlett, F. (1932). A study in experimental and social psychology. New York: Cambridge Press.

Bobrow, D.G., & Norman, D.A. (1975). Some principles of memory schemata. In D.G. Bobrow & A.G. Collins (Eds.), Representation and understanding: Studies in cognitive science (pp. 131-150). New York: Academic Press.

Bodenhausen, G.V. & Lichtenstein, M. (1987). Social stereotypes and information processing strategies: The impact of task complexity. Journal of Personality and Social Psychology, 52, 871-880.

Bodenhausen, G.V., & Wyer, R.S. (1985). Effects of stereotypes on decision making and information-processing strategies. Journal of Personality and Social Psychology, 48, 267-282.

Borman, W.C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20, 238-252.

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

Borman, W.C. (1987). Personal constructs, performance schemata, and "folk theories" ofsubordinate effectiveness: Explorations in an Army officer sample. Organizational Behavior and Human Decision Processes, 40, 307-322.

Brewer, M.B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as as prototypes. Journal of Personality and Social Psychology, 8, 393-400.

Brewer, W.F., & Nakamura, G.V. (1984). The nature and functions of schemas. In R. S. Wyer, Jr., & T.K. Srull (Eds.), Handbook of social cognition (Volume 1, pp. 119-160). Hillsdale, NJ: Erlbaum.

Brewer, W.F., & Treyens, J.C. (1981). Role of schemata in memory for places. Cognitive Psychology, 13, 207-230.

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

Bruner, J.S. (1958). Social psychology and perception. In E.Macoby, T. Newcomb, & E. Hartley (eds.), Readings in social psychology. New York: Holt, Rinehart & Winston.

Burnstein, E., & Schul, Y. (1982). The informational basis of social judgments: Operations in forming impressions of other persons. Journal of Experimental Social Psychology, 18, 217-234.

Burnstein, E., & Schul, Y. (1983) The informational basis of social judgments: Memory for integrated and nonintegrated trait descriptions. Journal of Experimental Social Psychology, 19, 49-57.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 35, 38-48.

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

Catrambone, R., & Markus, H. (1987). The role of self-schemas in going beyond the information given. Social Cognition, 5, 349-368.

Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.) Visual information processing. New York: Academic Press.

Cooper, W.H. (1981). Ubiquitous halo. Psychological Bulletin, 90, 218-244.

Crocker, J., Fiske, S.T., & Taylor, S.E. (1984). Schematic bases of belief change in R. Eiser (Ed.), Attitudinal judgement. New York: Springer-Verlag.

Crowder, R.G. (1976). Principles of memory and learning. Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Ebbesen, E. (1980). Cognitive process in understanding ongoing behavior in R. Hastie, T.M. Ostrom, E.B. Ebbesen, R. S. Wyer, D.L. Hamilton, & D.E. Carlston (Eds.) Person Memory: The cognitive basis of social perception. Hillsdale, NJ: Lawrence Erlbaum.

Ebbesen, E., & Allen, R.B. (1979). Cognitive processes in implicit personality trait inferences. Journal of Personality and Social Psychology, 37, 471-488.

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

Ferguson, L.W. (1949). The value of acquaintance ratings in criterion research. Personnel Psychology, 2, 93-102.

Fiske, S.T., & Dyer, L.M. (1985). Structure and development of social schemata: Evidence from positive and negative transfer effects. Journal of Personality and Social Psychology, 48, 839-852.

Fiske, S.T., & Linville, P.W. (1980). What does the schema concept buy us? Personality and Social Psychology Bulletin, 6, 545-557.

Fiske, S.T., & Neuberg, S.L. (1990). A continuum of impression formation, from category based to individuating processes: Influences of information and motivation on attention and interpretation. In M.P. Zanna (Ed.), Advances in experimental social psychology (Vol. 23, pp. 1-74). New York: Academic Press.

Fiske, S.T., & Taylor, S.E. (1991). Social Cognition. New York: Random House.

Foti, R.J., & Hauenstein, N.M.A. (1990). Processing demands and the effects of prior impressions on subsequent judgments: Clarifying the assimilation/contrast debate. Paper under review.

Franks, J.J., & Bransford, J.D. (1971). Abstraction of visual patterns. Journal of Experimental Social Psychology, 90, 65-74.

Glover, J., Plake, B., & Zimmer, J. (1982). Distinctiveness of encoding and memory for learning tasks. Journal of Education Psychology, 74, 189-198.

Glover, J., Plake, B., Zimmer, J., & Palmere, M. (1981). Distinctiveness of encoding: The effects of paraphrasing and drawing inferences on memory from prose. Journal of Educational Psychology, 73, 736-744.

Graesser, H.L., Gordon, S.E., & Sawyer, J.D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer and tag hypothesis. Journal of Verbal Learning and Verbal Behavior, 18, 319 -332.

Grier, J.B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. Psychological Bulletin, 75, 424-429.

Hamilton, D.L. (1981). Organizational processes in impression formation. In E. T. Higgins, C.P. Herman, & M.P. Zanna (Eds.), Social cognition: The Ontario Symposium (Vol. 1, pp. 135-160). Hillsdale, NJ: Erlbaum.

Hastie, R. (1981). Schematic principles in human memory. In E.T. Higgins, C.P. Herman, & M.P. Zanna (Eds.), Social Cognition: the Ontario symposium. Hillsdale, N.J.: Erlbaum.

Hastie, R., & Kumar, P.A. (1979). Person Memory: Personality traits as organizing principles in memory for behaviors. Journal of Personality and Social

Psychology, 37, 25-38.

Hastie, R., Park, B., & Weber, R. (1984). Social Memory. In R.S. Wyer, Jr., & T.K. Srull (Eds.), Handbook of social cognition (Vol. 2, pp. 151-212). Hillsdale, NJ: Erlbaum.

Hastorf, A.H., Schneider, D.J., & Polefka, J. (1970). Person perception. Reading, Mass.: Addison-Wesley.

Hauenstein, N.M.A., & Alexander, R.A. (1990). Rating ability in performance judgments: The joint influence of implicit theories and intelligence. Paper under review.

Hauenstein, N.M.A., & Kovach, R.C. (1988). The effects of category familiarity on the rating process. Paper presented at the 96th annual Convention of the American Psychological Association, Atlanta, GA.

Hauenstein, N.M.A., & Walker, S.E. (1989, April). The relationship between memory and appraisal accuracy: The moderating effects of job knowledge and judgment complexity. In D.J. Woehr (Chair), The Relationship between Memory and Judgment: Implications for Performance Appraisal. Symposium presented at the annual meeting for the Society for Industrial Psychology, Boston, Massachusetts..

Higgins, E.T., & Bargh, J.A. (1987). Social cognition and social perception. In M.R. Rosenweig & L.W. Porter (Eds.), Annual Review of psychology (Vol. 38, pp. 369-425). Palo Alto, CA: Annual Reviews.

Higgins E.T., Herman, C.P., & Zanna, M.P. (1981). Social cognition: The Ontario Symposium (Vol 1.). Hillsdale, NJ: Erlbaum, 1981.

Higgins, E.T., King, G.A., & Mavin, G.H. (1982). Individual construct accessibility and subjective impressions and recall. Journal of Personality and Social Psychology, 43, 35-47.

Higgins, E.T., Rholes, W., & Jones, C. (1977). Category accessibility and impression formation. Journal of Experimental Social Psychology, 13, 141-154.

Hogarth, R.M. (1980). Judgment and choice: The psychology of decision. New York: Wiley.

Hollander, E.P. (1957). The reliability of peer nominations under various conditions of administration. Journal of Applied Psychology, 41, 211-225.

Hunt, E., & Lansman, A. (1982). Individual differences in attention. In R.Sternberg (Ed.), Advances in psychology of human intelligence. Hillsdale, NJ: ERlbaum.

Ilgen, D., & Feldman, J. (1983). Performance appraisal: A process focus. In B. Staw & T. Cummings (Eds.). Research in organizational behavior (Vol. 5, pp. 141-197), New York: JAI Press.

Jackson, D.N. (1972). A model for inferential accuracy. The Canadian Psychologist, 13, 185-194.

Kane, J.S., & Lawler, E.E., III. (1978) Methods of peer assessment. Psychological Bulletin, 85, 555-586.

Kavanaugh, M.J., MacKinney, A.C., & Wolins, A.C. (1971). Issues in managagerial performance: Multitrait-multimethod analysis in ratings. Psychological Bulletin, 75, 34-49.

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

Klatzy, R.C. (1980). Human memory: Structures and Processes. San Francisco, CA: Freeman.

Kozlowski, S.W.J., & Ford, J.J. (1991). Effects of familiarity, performance, constraint and memory on rater information acquisition strategies. In A.S. DeNisi (Chair), Memory issues in the performance appraisal process. Symposium presented at the Annual Meeting of the Academy of Management, Anaheim, CA.

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

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

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

Landy, F., & Farr, J. (1983). The measurement of work performance: Methods, theory, and applications. New York : Academic Press.

Larson, J.R., Lingle, J.H., & Scerbo, M.M. (1984). The impact of performance cues on leader-behavior ratings: The role of selective information availabilty and probablistic response bias. Organizational Behaviors and Human Performance, 33, 323-349.

Linville, P.W. (1982). The complexity extremity effect and age based stereotyping. Journal of Personality and Social Psychology, 42, 193-211.

Linville, P.W., & Jones, E.E. (1980). Polarized appraisals of outgroup members. Journal of Personality and Social Psychology, 38, 689-703.

Lord, R.G., Foti, R.J., & DeVader, C.L. (1984). A test of leadership categorization theory: Internal structure, information processing and leadership perceptions. Organizational Behavior and Human Performance, 34, 343-378.

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

Markus, H. (1977). Self-schemata and processing information about the self. Journal of Personality and Social Psychology, 35, 63-78.

Markus, R.W., Smith, J., & Moreland, R.L. (1985). Role of the self-concept in the perception of others. Journal of Personality and Social Psychology, 49, 1494-1512.

McCloskey, M.E., & Glucksberg, S. (1978). Natural categories: well-defined or fuzzy sets. Memory and Cognition, 6, 462-472.

Mintzberg, H. (1973). The nature of managerial work. New York: Harper & Row.

Mount, M.K., & Thompson, D.E. (1987). Cognitive categorization and quality of performance ratings. Journal of Applied Psychology, 69, 147-156.

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

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

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. Journal of Personality and Social Psychology, 28, 28-38.

Norman, D.A., & Bobrow, D.G. (1976). On the role of active memory processes in perception and cognitive. In C.N.Cofer (Ed). The structure of human memory. San Francisco: Freeman.

Ostrom, T.M., Pryor, J.B., & Simpson, D.D. (1981). The organization of social information. In E.T. Higgins, C.P. Herman, & M.P. Zanna (Eds.), Social cognition: The Ontario Symposium (Vol. 1, pp.3-38). Hillsdale, NJ: Erlbaum.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.

Reed, P.L., & Jackson, D.N. (1975). Clinical judgment of psychopathology: A model for inferential accuracy. Journal of Abnormal Psychology, 84, 475-482.

Reed, S.K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.

Reitman, J.S., & Bower, G.H. (1973). Storage and later rememberance of exemplars of a concept. Cognitive Psychology, 4, 194-206.

Rosch, E. (1977). Human categorization in E. Rosch & B.B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, NJ: Lawrence Erlbam.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), Cognition and categorization, Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in neutral categories. Cognitive Psychology, 8, 382-439.

Rothstein, M., & Jackson, D.N. (1980). Decision making in the employment interview: An experimental approach. Journal of Applied Psychology, 65, 271-283.

Ruble, D.N., & Stangor, C. (1986). Stalking the elusive schema: Insights from developmental and social psychological analyses of gender schemas. Social

Cognition, 4, 227-261.

Rumelhart, D.E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), Schooling and the acquisition of knowledge (pp. 99-136). Hillsdale, NJ: Erlbaum.

Rumelhart, E.E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S.J. Samuels Eds.). Basic processes in reading: Perception and comprehension. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R.C., & Abelson, R.P. (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Erlbaum.

Schneider, W., Hastorf, A.H., & Ellsworth, P.C. (1979). Person Perception. Reading, MA: Addison-Welsey.

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

Schul, Y. (1983). Integration and asbstraction in impression formation. Journal of Personality and Social Psychology, 44, 45-54.

Schul, Y., & Burnstein, E. (1985). The informational basis of social judgments: Using past impression rather than the trait description in forming a new impression. Journal of Exerpimental Social Psychology, 21, 421-439.

Schul, Y., Burnstein, E., & Martinez. J. (1983). The informational basis of social judgments: Under what conditions are inconsistent trait descriptions processed as easily as consistent ones? European Journal of Social Psychology, 13,143-151.

Sentis, K. P., & Burnstein, E. (1979). Remembering schema consistent information: Effects of a balance schema on recognition memory. Journal of Personality and Social Psychology, 37, 2200-2211.

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

Siess, T.F., & Jackson, D.N. (1970). Vocational interests and personality: An empirical investigation. Journal of Counseling Psychology, 17, 27-35.

Skowiaczek, L.M., Klayman, J., Sherman, S.J., & Skov, B. (1989). Information selection and use in hypothesis testing: What is a good question, what is a good answer. Unpublished manuscript.

Smith, E.E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. Cognitive Psychology, 10, 438-464.

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

Smith, J.W., & Reilly, R.R. (1987). True intercorrelations among job components, time delays in rating, and rater intelligence as determinants of accuracy in performance ratings. Orgnizational Behviors and Human Decision Processes, 40, 369-391.

Smither, J.W., Reilly, R.R., & Buda, R. (1988). Effects of prior performance information on ratings of present performance: Contrast versus assimilation revisited. Journal of Applied Psychology, 73, 487-496.

Snyder, M., & Campbell, B.H., & Preston, E. (1982). Testing hypothesis about human nature: Assessing the accuracy of social stereotypes. Social Cognition, 1, 256-272.

Snyder, M. & Swann, W.B. (1978). Hypothesis testing processes in social interaction. Journal of Personality and Social Psychology, 34, 1034-1042.

Spiro, R., & Tirre, W. (1980). Individual differences in schema utilization during discouse processing. Journal of Educational Psychology, 72, 204-208.

Srull, T.K. (1984). Methodological techniques for the study of person memory and social cognition. In R.S. Wyer & T. K. Srull (Eds.). Handbook of Social Cognition, (Vol. 2, pp. 1-72). Hillsdale, NJ: Lawrence Erlbaum Associates.

Srull, T.K., & Wyer, R.S., Jr. (1979). The role of category accessibiltiy in the interpretation of information about persons: Some determinants and implications. Journal of Personality and Social Psychology, 37, 1660-1672.

Srull, T.K., & Wyer, R.S., Jr. (1980). Category accessability and social perception: Some implications of the study of human memory and interpersonal judgements. Journal of Personality and Social Psychology, 38, 841-856.

Srull, T. K., & Wyer, R.S., Jr. (1989). Person memory and judgment. Psychological Review, 96, 58-83.

Sulsky, L.M., & Balzer, W.W. (1985). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73, 497-506.

Taylor, S.E., & Crocker, J. (1981). Schematic bases of social information processing. In E.T. Higgins, C.P Herman, & M.P. Zanna (Eds.), Social cognition: The Ontario Symposium (Vol, 1 pp89-134). Hillsdale, NJ: Erlbaum.

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.). Advances in experimental social psychology (Vol 11.). New York: Academic Press.

Thorndike, P.W. (1977). Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 9, 77-110.

Walker, S.E. (1989). Empirical test of the assumptions of processing invariance in laboratory studies of performance appraisal. Dissertation.

Wherry, R.J., & Bartlett, C.J. (1982). The control of bias in ratings: A theory of rating. Personnel Psychology, 35, 521-552.

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

Whitely, S.E. , & Doyle, K.O. (1976). Implicit theories in student ratings. American Educational Research Journal, 13, 214-252.

Whitely, S.E., & Doyle, K.O. (1978). Dimensions of effective teaching: Factors or artifacts. Educational and Psychological Measurement, 38, 107-117.

Whitely, S.E. , & Doyle, K.O. (1978). Dimensions of effective teaching: Factors or artifacts. Educational and Psychological Measurement, 38, 107-117.

Wilder, D. (1978) Effect of predictability on units of perception and attribution. Personality and Social Psychology Bulletin, 4, 281-284.

Williams, K.J., DeNisi, A.S., Meglino, B.M., & Cafferty, T.P. (1986). Initial decisions and subsequent performance ratings. Journal of Applied Psychology, 71, 189-195.

Woll, S.B., & Graeser, A.C. (1982). Memory discrimination for information typical or atypical of person schemas. Social Cognition, I, 287-310.

Wyer, R.S., Bodenhausen, G.V., & Srull, T.K. (1984). The cognitive representation of persons and groups and its effect on recall and recognition memory. Journal of Experimental Social Psychology, 20, 445-469.

Wyer, R.S., & Carlston, D.E. (1979). Social cognition, interference, and attribution. Hillsdale, NJ: Erlbaum.

Wyer, R.S., & Gordon, S.E. (1982). The recall of information about persons and groups. Journal of Experimental Social Psychology, 18, 128-164.

Wyer, R.S., & Gordon, S.E. (1984). The cognitive representation of social information. In R. S. Wyer, Jr. & T.K. Srull (Eds.). Handbook of Social Cognition, Volume 2. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wyer, R.S., & Srull, T.K. (1980). The processing of social stimulus information: A conceptual integration. In R.Hastie, T.M. Ostrom, E.B., Ebbesen, R.S. Wyer, D.L.Hamilton, D.E. Carlston (Eds.). Person Memory: The cognitive basis of social perception. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Wyer, R.S., & Srull, T.K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E.T. Higgins, C. P., Herman, & M.P. Zanna (Eds.). Social cognition: The Ontario Symposium on personality and social psychology. Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

Zadny, J., & Gerard, M. (1974). Attributed intentions and informational selectivity. Journal of Experimental Social Psychology, 10, 34-52.

Table 1

Summary Table of ANOVA for Order Effects due to Free Response Interview and Knowledge Questionnaire

Source of Variance	DF	Sum of Squares	Mean Square	F
Family Resemblance Scores	1	.0302	.0302	1.13
Error	78	2.082	.0267	

Table 2

Summary Table of ANOVA for Order Effects due to Videotape Performance Order with the Components of Accuracy as the Dependent Variables

Source of Variance	DF	Sum of Squares	Mean Square	F
Components of Accuracy				
Elevation	1	.0892	.0892	1.01
Error	78	6.895	.0884	
Differential Elevation	1	.2153	.2153	2.0
Error	78	8.406	.1077	
Stereotype Accuracy	1	.0406	.0406	1.24
Error	78	2.547	.0326	
Differential Accuracy	1	.0068	.0068	.06
Error	78	4.04	.0517	

Table 3

Summary Table of ANOVA for Order Effects due to Videotape Performance Order with A' and False Alarm Rate as the Dependent Variables

Source of Variance	DF	Sum of Squares	Mean Square	F
Recognition Accuracy				
A' (Average Performance)	1	.0348	.0348	.87
Error	74	2.98	.0402	
A' (Good Performance)	1	.0877	.0877	3.03
Error	77	2.232	.0289	
A' (Poor Performance)	1	.0056	.0056	.16
Error	78	2.802	.0359	
False Alarm Rate (Average)	1	.0687	.0687	1.44
Error	78	3.713	.0476	
False Alarm Rate (Good)	1	.0676	.0676	1.48
Error	78	3.574	.0458	
False Alarm Rate (Poor)	1	.0290	.0290	.76
Error	78	2.967	.0380	

Table 4

Means and Standard Deviations for Manipulation Checks of Presence/Absence of Competing Task

	Competing Task	
	Present	Absent
Q#1	4.51 (2.23)	5.61 (1.73)
Q#2	3.95 (1.82)	5.76 (1.81)

Note: Q#1 = degree that subject felt he/she could pay attention to behaviors in the tape.
Q#2 = ability of subject to recall videotape behaviors.

Note: Numbers in parentheses denote standard deviations.
Higher numbers denote more division of attention.

Table 5

Summary Table of ANOVA for Competing Task Manipulation

Source of Variance	Sum of Squares	DF	F
Q#1	24.325	1	6.09**
Error	335.8	78	
Q#2	66.062	1	20.06***
Error	256.82	78	

Note:

**p < .01.

***p< .001.

Table 6

Frequency of Behaviors mentioned in the Free Recall Interview

Hammering

	Good	Performance Level Average	Poor
1. Bends Nails		2	10
2. Chokes up on hammer			33
3. Correct Size Hammer	9	5	
4. Uses wrist in hammering	33	32	
5. Damages wood (marks on wood, cat's face).		3	41
6. No damage to wood (no cat's face, no marks).	28	16	3
7. Misses Nails		2	13
8. Nail Set Used	7	3	
9. Hits nail first try	9	5	

Sanding

	Good	Performance Level Average	Poor
1. Sands with Grain	50	4	2
2. Not smooth (Rough cuts, cuts too deep, uneven cuts, gouges wood).		1	28
3. Sands in wrong direction			55
4. Wrong paper		2	10
5. Inconsistent strokes (partial strokes, short uneven strokes).		3	5
6. Correct grid sandpaper	12	4	
7. No pits or dips	26	15	1
8. Holds sander in one place too long	2	10	

Sawing.

	Good	Performance Level Average	Poor
1. Crooked	1	10	37
2. Cuts on line (straight cut)	68	54	1
3. Out of Square			17
4. Corners fit	11	6	
5. Practices unsafe behavior			16
6. Safe (safety glasses used, uses	21	9	

guard)			
7. Marks off line (misses line, watches mark)	11	8	2
8. Square Cut	11	4	
9. No mark			8

Staining

	Performance Level		
	Good	Average	Poor
1. Even (smooth, accurate)	90	51	
2. Too much stain (heavy, thick)		4	20
3. Blotchy, runny			16
4. Smooth, even strokes	7	6	
5. Consistent, even, uniform stain	47	29	
6. Sloppy, uneven			9
7. Cleans up drippings	9	3	
8. No wiping it down		3	8
9. Correct amt. of stain	3	4	

Table 7

Descriptive Statistics for Family Resemblance Scores by Knowledge Level

Subjects with higher scores (Scores 15-22).

Variable	N	Mean	Standard Deviation	Min	Max
Family Resemblance Scores	43	.5978	.145	.207	.937

Subjects with lower scores (Scores 0-14).

Variable	N	Mean	Standard Deviation	Min	Max
Family Resemblance Scores	37	.464	.155	.211	.974

Class Intervals for Knowledge Scores

Score	Frequency
0-2	0
3-4	2
5-6	4
7-8	6
9-10	12
11-12	5
13-14	8
15-16	8
17-18	7
19-20	19
21-22	9

Table 8

Results of Correlation Analysis Between Family Resemblance Scores and Knowledge Level Scores

Pearson Product Moment Correlation Coefficient

H_0: Rho = 0 / N=80

	Family Resemblance Scores
Knowledge Level	.530

Table 9

Means and Standard Deviations for True Scores by Performance Level for Behavioral and Graphic Ratings for Carpentry Performance

Variable	Performance		
	Good	Average	Poor
JUDGMENTS			
1. Saw	3.80 (0.45)	1.60 (0.89)	3.40 (0.89)
2. Hammer	2.00 (0.71)	4.60 (0.55)	1.60 (0.55)
3. Stain	3.20 (1.48)	4.00 (0.99)	2.00 (0.71)
4. Sand	3.80 (0.84)	1.40 (0.55)	1.80 (0.84)
5. Overall Rating	3.40 (0.55)	2.60 (0.55)	1.80 (0.45)

Note: numbers in parentheses denote standard deviations.

Table 10

Means and Standard Deviations for Observed Ratings by Performance Level for Behavioral and Graphic Ratings for Carpentry Performance

Variable	Performance: Good	Average	Poor
JUDGMENTS			
1. Saw	3.07 (0.96)	3.32 (0.09)	2.12 (1.01)
2. Hammer	2.87 (0.93)	2.98 (1.07)	2.69 (1.34)
3. Stain	2.84 (0.95)	2.75 (0.99)	2.79 (1.02)
4. Sand	3.20 (0.01)	2.85 (1.10)	2.30 (1.02)
5. Overall Rating	3.05 (0.83)	3.05 (0.83)	2.45 (0.88)

Note: numbers in parentheses denote standard deviations.

Table 11

Descriptive Statistics for Components of Accuracy

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Elevation	80	.358	.297	.003	1.396
Differential Elevation	80	.521	.328	.038	.294
Stereotype Accuracy	80	.411	.181	.110	.957
Differential Accuracy	80	1.067	.225	.682	1.73

Note: Smaller values denote higher levels of rating accuracy.

Table 12

Descriptive Statistics for the different components of accuracy by knowledge level

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Subjects with higher knowledge scores (15-22)					
Elevation	43	.335	.322	.0033	1.39
Differential Elevation	43	.446	.310	.037	1.30
Stereotype Accuracy	43	.416	.180	.111	.957
Differential Accuracy	43	1.05	.228	.682	1.73
Subjects with lower knowledge scores (0-14)					
Elevation	37	.385	.268	.003	1.070
Differential Elevation	37	.608	.331	.131	1.275
Stereotype Accuracy	37	.406	.184	.145	.849
Differential Accuracy	37	1.092	.223	.700	1.56

Note: Smaller numbers denote higher levels of rating accuracy.

Table 13

Descriptive Statistics for Components of Accuracy by Presence/Absence of Competing Task

Presence of Competing Task

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Elevation	41	.37	.282	.003	1.07
Differential Elevation	41	.489	.322	.037	1.27
Stereotype Accuracy	41	.400	.76	.11	.849
Differential Accuracy	41	1.05	.180	.700	1.51

Absence of Competing Task

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Elevation	39	.345	.315	.003	1.40
Differential Elevation	39	.555	.336	.038	1.30
Stereotype Accuracy	39	.423	.187	.145	.957
Differential Accuracy	39	1.089	.264	.682	1.73

Note: Smaller values denote higher levels of rating accuracy.

Table 14

ANOVA Results for Elevation Component of Accuracy for Hypothesis 1

Source of Variation	DF	Sum of Squares	Mean Squares	F Value
Model	1	.0119	.0119	.13
Residual	78	6.473	.0894	

Table 15

ANOVA Results for Differential Elevation for Hypothesis 1

Source of Variation	DF	Sum of Squares	Mean Squares	F Value
Model	1	.0858	.0858	1.96
Residual	78	8.450	.1083	

Table 16

ANOVA Results for Stereotype Accuracy for Hypothesis 1

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	1	.0103	.0103	.311
Residual	78	2.578	.0331	

Table 17

ANOVA Results for Differential Accuracy for Hypothesis 1

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	1	.0364	.0364	.72
Residual	78	3.966	.0508	

Table 18

Descriptive Statistics for False Alarm Rate and Correct Rejection Rate

Performance Level	N	Mean	Standard Deviation	Minimum	Maximum
Average	80	.267	.218	0	.875
Good	80	.238	.214	0	.875
Poor	80	.213	.194	0	.833

Note: Lower numbers denote lower probability of False Alarms.

Performance Level	N	Mean	Standard Deviation	Minimum	Maximum
Average	80	.733	.218	.125	1.0
Good	80	.761	.214	.125	1.0
Poor	80	.786	.194	.166	1.0

Note: Lower numbers denote lower probability of Correct Rejections.

Table 19

Descriptive Statistics for False Alarm Rate by knowledge level

	N	Mean	Standard Deviation	Minimum	Maximum
Subjects with higher knowledge scores (15-22)					
Average Performance	43	.177	.157	0	.625
Good Performance	43	.182	.198	0	.857
Poor Performance	43	.172	.153	0	.50
Subjects with lower knowledge scores (0-14)					
	N	Mean	Standard Deviation	Minimum	Maximum
Average Performance	37	.370	.235	0	.875
Good Performance	37	.304	.216	0	.875
Poor Performance	37	.261	.226	0	.833

Table 20

Descriptive Statistics for False Alarm Rate by Presence/Absence of Competing Task

	N	Mean	Standard Deviation	Minimum	Maximum
Presence of Competing Task					
Avg	41	.318	.232	0	.875
Good	41	.270	.210	0	.875
Poor	41	.256	.210	0	.833
Absence of Competing Task					
Avg	39	.211	.191	0	.750
Good	39	.205	.217	0	.857
Poor	39	.167	.167	0	.500

Table 21

Regression Results of Hypothesis 2 (a) for False Alarm Rate (Average Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	2	.6515	.3257	8.011***
Error	77	3.131	.0407	

R-SQUARE .1722

Variables in the Equation	b	t	p>t
Knowledge Level	-.0148	-2.95	.0042
Family Resemblance	-.1196	-.731	.468

Note:

***p<.0007

Table 22

Regression Results of Hypothesis 2 (a) for False Alarm Rate (Good Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	2	2.588	.1294	2.946
Error	77	3.382	.0440	

R-SQUARE .0711

Variables in the Equation	b	t	p>t
Knowledge Level	-.0117	-2.242	.0278
Family Resemblance	-.0678	-.399	.691

Table 23

Regression Results of Hypothesis 2 (a) for False Alarm Rate (Poor Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	2	.1527	.0764	2.068
Error	77	2.843	.0369	

R-SQUARE .0510

Variables in the Equation	b	t	p>t
Knowledge Level	-.0031	.659	.512
Family Resemblance	-.199	-1.283	.204

Table 24

Descriptive Statistics for A' and Misses

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Average	76	.815	.201	.50	1.0
Good	79	.829	.172	.50	1.0
Poor	80	.797	.189	.50	1.0

Note: Higher number denote better recognition accuracy.

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Average	80	.283	.398	0	1.0
Good	80	.285	.365	0	1.0
Poor	80	.347	.376	0	1.0

Note: Lower numbers indicate lower probability of misses.

Table 25

Descriptive Statistics for A' by Knowledge Level

	Subjects with higher knowledge scores (15-22)				
	N	Mean	Standard Deviation	Minimum	Maximum
Average Performance	43	.883	.163	.50	1.0
Good Performance	43	.840	.182	.50	1.0
Poor Performance	43	.816	.189	.50	1.0
	Subjects with lower knowledge scores (0-14)				
	N	Mean	Standard Deviation	Minimum	Maximum
Average Performance	33	.725	.211	.50	1.0
Good Performance	36	.815	.161	.50	.97
Poor Performance	37	.775	.187	.50	1.0

Table 26

Desriptive Statistics for A' by Presence/Absence of Competing Task

Presence of Competing Task

	N	Mean	Standard Deviation	Minimum	Maximum
Avg	38	.764	.213	.50	1.0
Good	41	.794	.181	.50	1.0
Poor	41	.797	.183	.50	1.0

Absence of Competing Task

	N	Mean	Standard Deviation	Minimum	Maximum
Avg	38	.864	.175	.50	1.0
Good	38	.865	.151	.50	1.0
Poor	39	.798	.196	.50	1.0

Table 27

Regression Results for A' for hypothesis 2 (b) (Average Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	2	.5010	.2505	7.271***
Error	73	2.515	.0345	

R-SQUARE .1661

Variables in the Equation	b	t	p > t
Knowledge Level	.0110	2.336	.0223
Family Resemblance	.2127	1.389	.1689

Note: ***$p < .0013$

Table 28

Regression Results for A' for Hypothesis 2 (b) (Good Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	2	.0733	.0366	1.239
Error	76	2.247	.0296	

R-SQUARE .0316

Variables in the Equation	b	t	p > t
Knowledge Level	-.0039	-.091	.9276
Family Resemblance	.1971	1.390	.1686

Table 29

Regression Results for A' for Hypothesis 2 (b) (Poor Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	2	.101	.0505	1.436
Error	77	2.706	.0351	

R-SQUARE .0360

Variables in the Equation	b	t	p > t
Knowledge Level	-.0010	-.211	.833
Family Resemblance	.2339	1.535	.1282

Table 30

Regression Results for False Alarm Rate for Hypothesis 3 (a) (Average Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	.7686	.1098	3.322***
Error	68	2.246	.0331	

R-SQUARE .2548

Variables in the Equation	b	t	p>t
Knowledge Level	.0023	.091	.9280
Competing Task	-.5595	-1.125	.2647
Family Resemblance	.1047	.137	.8913
Competing Task X Family Resemblance	.578	.561	.5763
Competing Task X Knowledge Level	.0311	.858	.3939
Family Resemblance X Knowledge Level	.0030	.062	.9507
Competing Task X Family Resemblance X Knowledge Level	-.0356	-.521	.6041

Note: ***p<.0042

Table 31

Regression Results for False Alarm Rate for Hypothesis 3 (a) (Good Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	.2758	.0394	1.368
Error	71	2.044	.0288	

R-SQUARE .1189

Variables in the Equation	b	t	p>t
Knowledge Level	-.0043	-.185	.8539
Competing Task	-.482	-1.047	.2985
Family Resemblance	-.2141	-.300	.7648
Competing Task X Family Resemblance	1.131	1.184	.2402
Competing Task X Knowledge Level	.0137	.408	.6848
Family Resemblance X Knowledge Level	.01459	.319	.7505
Competing Task X Family Resemblance X Knowledge Level	-.0478	-.755	.4529

Table 32

Regression Results for False Alarm Rate for Hypothesis 3 (a) (Poor Performance Level)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	.1911	.0273	.751
Error	72	2.616	.0363	

R-SQUARE .0681

Variables in the Equation	b	t	p>t
Knowledge Level	.0207	.831	.4089
Competing Task	.0017	.003	.9973
Family Resemblance	.9946	1.33	.186
Competing Task X Family Resemblance	-.1003	-.097	.9229
Competing Task X Knowledge Level	.0090	.245	.8068
Family Resemblance X Knowledge Level	-.0462	.946	.3473
Competing Task X Family Resemblance X Knowledge Level	-.0878	-.127	.8996

Table 33

Regression Results for Elevation Component of Accuracy for Hypothesis 3 (b)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	.2835	.0415	.435
Error	72	6.701	.0931	

R-SQUARE .0406

Variables in the Equation	b	t	p > t
Knowledge Level	.0115	.288	.7744
Competing Task	.2945	.373	.7160
Family Resemblance	.9376	.787	.4341
Competing Task X Family Resemblance	-.6872	-.416	.6787
Competing Task X Knowledge Level	-.2005	-.341	.7339
Family Resemblance X Knowledge Level	-.5023	-.643	.5226
Competing Task X Family Resemblance X Class Level	.0460	.414	.6801

Table 34

Regression Results for Differential Elevation Component of Accuracy for Hypothesis 3 (b)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	1.014	.1449	1.387
Error	72	7.521	.1045	

R-SQUARE .1188

Variables in the Equation	b	t	p > t
Knowledge Level	.0490	1.159	.2504
Competing Task	1.109	1.323	.1899
Family Resemblance	1.7977	1.423	.1589
Competing Task X Family Resemblance	-1.89	-1.081	.2833
Competing Task X Knowledge Level	-.0814	-1.308	.195
Family Resemblance X Knowledge Level	-.1107	-1.337	.1854
Competing Task X Family Resemblance X Knowledge Level	.1229	1.045	.2996

Table 35

Regression Results for Stereotype Accuracy Component of Accuracy for Hypothesis 3 (b)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	.1833	.0262	.784
Error	72	2.404	.0334	

R-SQUARE .0708

Variables in the Equation	b	t	p > t
Knowledge Level	.0017	.069	.9453
Competing Task	-.1174	-.248	.8050
Family Resemblance	-.4460	-.625	.5342
Competing Task X Family Resemblance	.4533	.458	.6483
Competing Task X Knowledge Level	.0083	.230	.8189
Family Resemblance X Knowledge Level	.0116	.247	.8053
Competing Task X Family Resemblance X Knowledge Level	-.0314	-.472	.5382

Table 36

Regression Results for Differential Accuracy Component of Accuracy for Hypothesis 3 (a)

Source of Variation	DF	Sum of Squares	Mean Squares	F
Model	7	.1186	.017	.315
Error	72	3.8836	.0540	

R-SQUARE .0297

Variables in the Equation	b	t	p > t
Knowledge Level	.0257	.847	.3997
Competing Task	.2232	.371	.7118
Family Resemblance	.0389	.924	.3583
Competing Task X Family Resemblance	-.7525	-.598	.515
Competing Task X Knowledge Level	-.0149	-.333	.7404
Family Resemblance X Knowledge Level	-.5910	-.993	.3241
Competing Task X Family Resemblance X Knowledge Level	.0436	.516	.6075

Appendix A

CARPENTRY QUESTIONNAIRE

Name:
Age:
Years of Carpentry Experience:

Instructions

This questionnaire contains 22 multiple-choice items covering various aspects of carpentry. Please answer each question to the best of your knowledge.
Each question has four (4) possible answers: a, b, c, or d. On your questionnaire, circle the letter corresponding to the correct answer.

EXAMPLE: What tool is a "chuck-key" used with?

a. Saw.
b. Hammer.
c. Drill.
d. Vise.

The correct answer is "c". You would circle the letter "c" on your questionnaire.

1. What size level would you use to check if a wall is plumb?
a. 9 inch.
b. 2 feet.
c. 4 feet.
d. 5 feet.

2. What is a "catspaw" used for?
a. extracting nails.
b. determining bevel cuts.
c. countersinking.
d. measuring angles.

3. What type of saw is most often used for cutting base molding?
a. miter.
b. saber.
c. circular.
d. hacksaw.

4. How many inches are studs usually centered on?
a. 14.
b. 16.
c. 18.
d. 20.

5. What does a "2 X 4" actually measure?
a. 1 1/4 X 3 1/4.
b. 1 1/2 X 3 1/2.
c. 1 3/4 X 3 3/4.
d. 2 X 4.

6. What tool is used to determine lengths of common, hip, and valley rafters?
a. framing square.
b. tape rule.
c. hand square.
d. bevel square.

7. What size nail would you use to install shoe molding?
a. 2d.
b. 4d.
c. 8d.
d. 10d.

8. What tool has a rafter table stamped on its face?
a. circular saw.
b. bevel square.
c. miter saw.
d. framing square.

9. What grade of lumber is used for framing members such as studs, rafters, and joists?
a. number 1 common.
b. number 2 common.
c. number 3 common.
d. number 4 common.

10. What handsaw is used for general cutting?
a. backsaw.
b. ripsaw.
c. hacksaw.
d. crosscut saw.

11. What is another word for framing carpentry?
a. finish carpentry.
b. trim carpentry.
c. rough carpentry.
d. cabinet building.

12. What kind of interior trim goes around doors and windows?
a. baseboard.
b. lattice.
c. molding.
d. casing.

13. What is the exterior trim called at the point where roof projections and side walls meet?
a. cornice.
b. crown molding.
c. rough sill.
d. wall apron.

14. What saw is best for cutting curved lines?
a. coping saw.
b. circular saw.
c. hacksaw.
d. crosscut saw.

15. What is the trim at the lower part of a window opening called?
a. apron.
b. casing.
c. sill.
d. jamb.

16. Which of the following are shims used for?
a. bracing a wall.
b. padding out a door frame.
c. scabbing wood together.
d. anchoring a stud wall.

17. Which of the following has a clutch?
a. miter saw.
b. drill.
c. circular saw.
d. screwshooter.

18. What kind of tool is best for planing door edges?
a. block plane.
b. smooth plane.
c. jack plane.
d. joiner plane.

19. What are tongue and groove boards typically used for?
a. ceilings.
b. roofs.
c. floors.
d. walls.

20. What is a "lintel"?
a. a horizontal support used over doors and windows.
b. a saw used for making bevel cuts.
c. a drill used to bore into hardwoods.
d. a beam which runs across the center of a roof.

21. What do you call pieces of wood placed between floor joists or wall studs for reinforcement?
a. braces.
b. nailers.
c. bridges.
d. supports.

22. How do you dull the tip of a nail?
a. file the tip down.
b. snip the tip off.
c. hit the tip with a hammer.
d. bend the tip.

ANSWER KEY: Carpentry Test

1.	c	12.	d
2.	a	13.	a
3.	b	14.	a
4.	b	15.	a
5.	b	16.	b
6.	a	17.	d
7.	b	18.	d
8.	d	19.	c
9.	b	20.	a
10.	d	21.	c
11.	c	22.	c

Appendix B

CONSENT FORM

TO ALL RESEARCH PARTICIPANTS:

This experiment is being conducted under the supervision of Patricia L. Marshall, graduate student and Roseanne J. Foti, faculty supervisor; it is an investigation of the way in which persons view the behavior of other persons in a work setting. We will be asking you to complete a job knowledge test and then view several different persons on videotape and then rate their performance. Below are some of the items of information that you should know when deciding to participate in this study.

(1) No psychological or physical harm is expected to result from your participation in the experiment.

(2) The experiment requires approximately 2 hours of your time. Your cooperation is greatly appreciated.

(3) Your agreement to participate should be voluntary and thus can be withdrawn at any time by you without penalty.

(4) All information gathered from your responses is intended for research purposes only. Therefore, your responses will remain completely confidential.

(5) This research project has been approved by the Human Subjects Committee of the Virginia Tech Psychology department. Any questions you have can be answered by contacting one of the individuals below:

Dr. Helen Crawford: (703-231-6520): Human Subjects Committee
Dr. Roseanne Foti: (office: 703-231-5814)
Pat L. Marshall: (home: 804-799-4280)
Ernest Stout

(6) There is a copy of this consent form available if you should wish to retain a copy for your personal records, BUT PLEASE FILL OUT THE COPY THAT IS ATTACHED TO THE SURVEY FOR OUR RECORDS.

If you consent voluntarily and have an understanding of the conditions outlined above to participate in the experiment, please sign you name below. Thank you very much for your assistance.

Signature________________________________

Name (please print):________________________

ID #:____

Appendix C

Competing Task for Higher Knowledge Subjects

1. How do you square up a building?

2. How would you lay off for a set of steps with a finished floor to floor height of 8 feet 9 inches?

3. What is the standard spacing for floor joists?

4. What is the standard frame opening for a 3046 window?

5. How do you get rise for a set of steps with a finish to finish floor of 8'?

Appendix D

Competing Task for Lower Knowledge Subjects

This task involves considering some common objects. Each object has a common use which will be given to you. You are to list ten other uses for which the object or parts of the object could serve.
For example:

Given: a newspaper (used for reading)

You might list the following uses:

1. to start a fire.
2. to wrap garbage in.
3. to swat flies.
4. stuffing to pack boxes.
5. to line drawers or shelves.

Note that all of the uses listed are different from each other and different from the primary use of a newspaper, which is for reading.

Remember, each use must be different from the others and different from the most common use which is given to you. In addition, do not use the same unusual use as a response to more than one object. In other words, none of your responses should occur more than once in the entire set.

LIST TEN POSSIBLE USES FOR EACH OF THE FOLLOWING OBJECTS:

a wire hanger (used for holding clothes)

1.______________________
2.______________________
3.______________________
4.______________________
5.______________________
6.______________________
7.______________________
8.______________________
9.______________________
10._____________________

a ruler (used for measuring length)

1.______________________
2.______________________
3.______________________
4.______________________
5.______________________
6.______________________
7.______________________
8.______________________
9.______________________
10._____________________

Appendix E

Critical Behavior Recognition Form

In this questionnaire, you should circle the behaviors that you saw in the videotape performed by the carpenter.

There are three response sheets in which you circle whether you saw the behaviors performed or not performed. If you did not see the carpenter perform a behavior, simply do not circle the number. It is important to fill out the sheets in the order in which you saw the different performers (for example sheet 1 for performer 1, sheet 2 for performer 2, and sheet 3 for performer 3).

Carpenter 1.

1. Carpenter sanded with the grain.
2. Carpenter sawed off the line
3. Carpenter used a rag to wipe off stain.
4. Carpenter used a nail set.
5. Carpenter sanded against the grain.
6. Carpenter choked up on the hammer.
7. Carpenter marked a line before sawing.
8. Carpenter sawed on the line.
9. Carpenter sawed a crooked line.
10. Carpenter used correct safety procedures when sawing.
11. Carpenter used a saw horse when sawing.
12. Carpenter sanded too deeply.
13. Carpenter drove nails in at angle.
14. Carpenter marked an x before sawing.
15. Carpenter used a block sander.
16. Carpenter brushed on too much stain.

Carpenter 2.

1. Carpenter sanded with the grain.
2. Carpenter sawed off the line
3. Carpenter used a rag to wipe off stain.
4. Carpenter used a nail set.
5. Carpenter sanded against the grain.
6. Carpenter choked up on the hammer.
7. Carpenter marked a line before sawing.
8. Carpenter sawed on the line.
9. Carpenter sawed a crooked line.
10. Carpenter used correct safety procedures when sawing.
11. Carpenter used a saw horse when sawing.
12. Carpenter sanded too deeply.
13. Carpenter drove nails in at angle.
14. Carpenter marked an x before sawing.
15. Carpenter used a block sander.
16. Carpenter brushed on too much stain.

Carpenter 3.

1. Carpenter sanded with the grain.
2. Carpenter sawed off the line
3. Carpenter used a rag to wipe off stain.
4. Carpenter used a nail set.
5. Carpenter sanded against the grain.
6. Carpenter choked up on the hammer.
7. Carpenter marked a line before sawing.
8. Carpenter sawed on the line.
9. Carpenter sawed a crooked line.
10. Carpenter used correct safety procedures when sawing.
11. Carpenter used a saw horse when sawing.
12. Carpenter sanded too deeply.
13. Carpenter drove nails in at angle.
14. Carpenter marked an x before sawing.
15. Carpenter used a block sander.
16. Carpenter brushed on too much stain.

Appendix F

Rating Scale: Carpentry Performance.

Rate the performance of each carpenter in order of appearance on the tape using the following scale. Mark all of your answers in the spaces provided directly on this sheet.

Carpenter #1.

1	2	3	4	5
Very Poor	Poor	Average	Good	Very Good

1. How well did carpenter #1 saw? ______
2. How well did carpenter #1 hammer? ______
3. How well did carpenter #1 stain? ______
4. How well did carpenter #1 sand? ______
5. How would you rate carpenter #1 on his OVERALL performance? ______

Carpenter #2.

1	2	3	4	5
Very Poor	Poor	Average	Good	Very Good

1. How well did carpenter #2 saw? ______

2. How well did carpenter #2 hammer? ______

3. How well did carpenter #2 stain? ______

4. How well did carpenter #2 sand? ______

5. How would you rate carpenter #2 on his OVERALL performance? ______

Carpenter #3.

1	2	3	4	5
Very Poor	Poor	Average	Good	Very Good

1. How well did carpenter #3 saw? ______
2. How well did carpenter #3 hammer? ______
3. How well did carpenter #3 stain? ______
4. How well did carpenter #3 sand? ______
5. How would you rate carpenter #3 on his OVERALL performance? ______

Appendix G

Manipulation Check Items for Competing Task

Please answer the following questions:

1. How well could you pay attention to the carpenter's behaviors in the tape?

1	3	5	7	9
not well	a little well	somewhat well	well	very well

2. How difficult was it to remember what each carpenter did in the videotape?

1	3	5	7	9
very difficult	a little difficult	somewhat difficult	not difficult	easy

Appendix H

Computational Formulae for Variance and Correlational Measures of Rating Accuracy.

1. Algebraic Difference Score Formulae for the Four Components of Accuracy. For a raters who evaluates n ratees on k items or dimensions, scores on elevation (E), differential elevation (DE), stereotype accuracy (SA), and differential accuracy (DA) are given by the square root of the following terms:

$$E^2 = (x.. - t..)^2$$

$$DE2 = 1/n\Sigma[(xi. - x..) - (ti. - t..)]^2$$

$$SA2 = 1/k\Sigma[(x.j - x..) - (t.j - t..)]^2$$

$$DA2 = 1/kn\Sigma\Sigma[(xij = xi. - x.j + x..) - (tij - ti. tij + t..)]^2$$

where xij and tij = rating and true score for ratee i on item j; xi. ti. = mean rating and mean true score for ratee i; x.j and t.j = mean rating and mean true score for item j; and x.. and t.. = mean rating and mean true score over all ratees and items.

From:

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

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

Appendix I

Computing Expression for A'

$$A' = 1/2 + \frac{(y - x)(1 + y - x)}{4y(1 - x)}$$

Note: y = probability of a hit.
x = probability of a false alarm.

From:

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. Psychological Bulletin, 75, 424-429.

VITA

Patricia Lane Marshall

School Address
Department of Psychology
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061-0436

Home Address
Rt. 3 Box 1650
Danville, VA 24540
(804) 799-4280

Education

1985-1992

Ph.D. Industrial/Organizational Psychology
Virginia Polytechnic Institute
and State University
Blacksburg, VA

Dissertation: The Effect of Competing Tasks, Personal Prototypes and Rater Knowledge on Performance Appraisal.

1983-1985

M.A. Psychology
University of Richmond
Richmond, Virginia
Major field of study: Psychology

1976-1980

B.S. Psychology
Virginia Commonwealth University
Richmond, Virginia

Professional Experience

1990-1990

Adjunct Faculty
Averett College
Danville, VA

-Taught graduate course in personnel management in M.B.A. program.

1989-1990

Consultant
American Telephone and Telegraph
Selection and Testing, Corporate Headquarters
Morristown, NJ

-Validated phone simulation instrument used in Selection of Sales Representatives.
-Conducted a job analysis and developed a written selection test for drafting jobs.

1986-1989	Coordinator of Undergraduate Advising Virginia Polytechnic Institute and State University Department of Psychology -Advised undergraduates of course work. -Maintained confidential computer (Dbase) files of all students enrolled in psychology.
1985-1986	Graduate Teaching Assistant Virginia Polytechnic Institute and State University Department of Psychology Blacksburg, VA -Taught laboratories in Introductory to Psychology. -Tested and evaluated students.
1983-1984	Social Science Research Analyst Federal Bureau of Prisons, Justice Department Federal Correctional Institute Petersburg, VA -Collected, retrieved, stored, and analyzed federal corrections social science data. -Interviewed inmates for various research projects.

Professional Pressentations

Marshall, P.L., & Foti, R.J. (March 1989). <u>The relationship of self-monitoring and task experience on leader emergence</u>. Poster presented at the Southeastern Psychological Association, Washington, D.C.

References

Dr. Roseanne J. Foti
Associate Professor of Psychology
Department of Psychology
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061
(703) 231-5814

Dr. Douglas J. McCormick
A T and T
Room 402-O
1 Speedwell Avenue
Morristown, NJ 07961
(201) 898-3475

Dr. Jerard F. Kehoe
District Manager
A T and T
Room 88W-426
1 Speedwell Avenue
Morristown, NJ 07961
(201) 898-3457

Dr. Joseph A. Sgro
Department Head
Department of Psychology
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061
(703) 231-6581

Professional Organizations

Academy of Management

American Psychological Association, Division 14

Southeastern Psychological Association

Relevant Coursework

Industrial and Organizational Psychology

Psych 5110	Personnel Psychology
Psych 5330	Psychology of Leadership
Psych 5400	Work and Motivation
Psych 5980	Special Study: Job Analysis
Psych 6004	Contemporary Literature in Applied Psychology
Psych 6010	Seminar in Applied Behavioral Science
Psych 6030	Criterion Development and Evaluation

Management

Mgt 5310	Dynamics of Organizational Behavior
Mgt 5370	Wage and Incentive Systems
Mgt 5480	Human Resource Planning and Development

Research Methods

Psych 5130	Advanced Test Theory
Psych 5311	Research Methods I
Psych 5312	Research Methods II
Psych 5313	Research Methods III
Edae 6340	Multivariate Statistics for Applications to Education Problems
Stat 5052	Statistics in Research

Industrial Engineering and Operations Research

IEOR 5311 Design and Evaluation of Man-Machine Systems

Ccomputer Skills

SAS-Mainframe
SAS-PC
SPSS-X