THE EFFECT OF JOB KNOWLEDGE AND TASK COMPLEXITY
ON INFORMATION PROCESSING AND RATING ABILITY

by

Robert T. Brill

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute & State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in
Psychology

APPROVED:

Neil M. A. Hauenstein, Chairperson

Roseanne J. Foti

Joseph J. Franchina

Javid Kaiser

Joseph A. Sgro

October, 1992

Blacksburg, Virginia
THE EFFECT OF JOB KNOWLEDGE AND TASK COMPLEXITY ON INFORMATION PROCESSING AND RATING ABILITY

by

Robert T. Brill

Committee Chairperson: Neil M. A. Hauenstein

Psychology

(ABSTRACT)

Previous research exploring the assumptions of process invariance (Walker, 1989) found discrepancies between process and rating outcomes when rater and ratee populations were crossed. Divergent results were attributed to differences in the ratee stimulus performance tapes. The present study attempted to explore how levels of task complexity would moderate the relationship between job knowledge and both information processing and rating accuracy. 123 male subjects were measured on their knowledge of football, and viewed the performance of either offensive tackles (complex task) or running backs (simple task) under directions to either form an impression or remember as much detail as possible. It was expected that
observational purpose would moderate the type of information recalled, rating accuracy, and accuracy in recording behavior frequencies in the complex task condition only. Results supported hypotheses only for Cronbach's (1955) elevation measure pertaining to behavior frequency ratings, and consistent task complexity main effects among accuracy measures for behavior frequency ratings. Some possible limitations and explanations for the present results and some implications for future research are offered.
ACKNOWLEDGEMENTS

The completion of these acknowledgements, and obtaining the proper signatures will indicate the end of a long journey for myself. I think Jerry captured it best -- sometimes the light was shining on me and other times I could barely see. Regardless, "what a long strange trip it's been!" There are many people to thank for helping me reach my destination and for the successful completion of this project.

First, to my committee - endless thanks for their patience and support. I cannot say enough about the amount of respect I hold for, and debt of gratitude I owe to each one of them. Neil Hauenstein, the chair of this project, advisor, and friend, has been especially helpful and deserving of both my praise and gratitude. They say imitation is the best form of a compliment - I would certainly like to develop a research style and repertoire with my students similar to the Professor Hauenstein format, just with a more organized desk! I look forward to continued friendship and professional collaboration with my esteemed mentor. Also, special thanks to Roseanne Foti for her help on the project and in developing my knowledge and skills in the I/O Psych area in general. As head of the I/O area, Roseanne was the one I frequently pestered about administrative issues and her patience in that area was also much appreciated. A very big thank you is also extended to Joe Franchina also. His constructive feedback and personal
concern have made a lasting mark on my professional ability. Javad Kaiser demonstrated great adaptability and competence as a member of the committee. I greatly appreciate his time and effort in contributing his expertise to the improvement of the dissertation. Finally, many thanks to Dr. Joseph Sgro for his part in the project. Having poured through the "longer than necessary" draft, Dr. Sgro's input certainly refined the final version. In addition, he has been a constant source of support and friendship throughout my experience at Virginia Tech. His willingness to go the whole ten yards (and then some) for his students was very evident to me and I will always be indebted to him. I'll get the first three rounds when we get together in Bethlehem, and I have not relinquished my goal of getting into the hottub.

As I gleefully returned to Pennsylvania after 4 years, I wondered how I survived Blacksburg in some of the more difficult times. The reason was obvious - a bevy of friendships that I hope will last a lifetime. To those special people, my precious cohorts, my comrades in arms: Monnie, Marta, Trina, Lance, Stuart, Dean, Justin, and Bill. I thank each of you for the way you touched my life. The quality (and diversity) in the eight names reflected on that one line is amazing. To attempt to express my feelings for each of you individually is not possible, nor desirable since the words would only taint the strength of those feelings. In addition, I would like to make special thanks to Steve
Walker, a flat out great guy, whose own dissertation generated the idea-seed for the present effort. Also, many thanks to Diane Corrigan, my research assistant. Without her competence and effort this project would not have been realized. In the short time I have interacted with her I can see she will be a quality student. Also, much love and thanks to Gayle Kennedy for all her support. Finally from Blacksburg, I'd like to acknowledge Tim the cop, and another police officer whose name I don't know, but who I met at the Dutch Village pool. Many thanks.

The long and winding road has brought me back home close to Philadelphia and many of the people who constantly expressed their support for my venture. I am especially referring to the Moss troop, and LaSalle soccer team. Although several of them tried to get me thrown out of Blacksburg one long weekend in August, I am indebted to their constant friendship and refusal to let me forget my roots. I can't name them all, but I will mention Bob Altenbach. Not because he's the most important, but rather he's the only one who'll get pissed off if I don't mention him, the big baby. Just kiddin' Alty, you are the best!

To my parents, Dorothy and Joseph Brill, many thanks for your unending love, support, and encouragement. I said it all in the thesis and don't know what more to add. It seems like so long ago: What if Joey's not there to meet me at the firehouse?, what if I get into the wrong line?, what if I get
off the wrong exit on Interstate 81? (some things never change). I've come a long way baby! You two are the greatest parents one could hope for, and I treasure the friendship we've developed during these years. The lessons you've taught me are more valuable than any graduate course I could ever take in that I will always value what is most important - God and family. Thanks for perspective.

Tim, Joe and Mike Brill - thanks for making sure there was never a dull moment. You guys are the greatest. Please realize that the pride in me that you have expressed is indeed reciprocated for those unique strengths that each of you possess and exhibit. In addition, I need to make mention of two other very special families that have extended a great deal of support to me. Special thanks to the Christian family and the Tricoski (Inciardi) family. Also, thanks to Bono, George Bailey, Lee Iaococa, and S.E. Hinton for the inspiration, as well as Moravian College, especially Lori, Stacey, Art, Bill & Dana in the Psych department, another source of support (and incentive).

Lastly, all my love and thanks to my wife (that still sounds funny, but I like the implications). Joyce, I've said it many times before, I'll say it many times again, and it still won't be enough - I love you. I owe much of my success, and ALL my reason for striving higher, to your love, support, and companionship. We've finally begun, and I'm finding being with you a better place to be.
# TABLE OF CONTENTS

Abstract ....................................................... ii

Acknowledgements ........................................ iv

List of Tables .............................................. x

List of Figures ............................................. xiii

Introduction ............................................... 1

Literature Review ........................................ 14

Description of Study ....................................... 28

Hypotheses .................................................. 30

Methodology ............................................... 35

Results ...................................................... 47

Discussion ................................................. 88

References .................................................. 102
Tables .............................................. 107

Figures .............................................. 119

Appendices .......................................... 124

Author's Vita ...................................... 149
LIST OF TABLES

1. Means and Standard Deviations for Manipulation Checks Task Complexity and Observational Purpose by Experimental Conditions ...................... 107

2. Hierarchical Regression Analysis Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Perceptions of Complexity ...................... 108

3. ANOVA Table for Observational Purpose Manipulation Check .................................. 109


5. Hierarchical Regression Analysis Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Free Recall Measures ...................... 111
6. Correlations Between Job Knowledge and the Free Recall Measures Overall, and within each Task Condition .................. 112

7. Means and Standard Deviations for Accuracy Measures by Task Complexity Conditions and Observational Purpose Conditions .................. 113

8. Hierarchical Regression Analysis Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Accuracy Measures for Evaluative Ratings ............ 114

9. Hierarchical Regression Analysis Testing the Relationship Between Job Knowledge, Task Complexity and Observational Purpose on Accuracy Measures for Behavior Frequency Ratings (Elevation and Differential Elevation) ............. 115

10. Hierarchical Regression Analysis Testing the Relationship Between Job Knowledge, Task Complexity and Observational Purpose on Accuracy Measures for Behavior Frequency Ratings (Stereotypical Accuracy and Differential Accuracy) ............... 116
11. Correlations Between Job Knowledge and Accuracy Measures Overall, and Within each Experimental Condition .......................... 117

12. Hierarchical Regression Analysis Testing the Relationship Between Job Knowledge, and Perceptions of Complexity on all Process and Outcome Measures .......................... 118
LIST OF FIGURES

1. The Moderating Effect of Observational Purpose on the Relationship Between Job Knowledge and Elevation of Behavior Frequency Ratings for Simple and Complex Task Conditions .................. 119

2. Task Complexity by Observational Purpose Interaction for Differential Elevation Measure of Behavior Frequency Ratings .................... 120

3. The Moderating Effect of Task Complexity on the Relationship Between Job Knowledge and Differential Elevation for Behavior Frequency Ratings ......................................................... 121

4. The Moderating Effect of Observational Purpose on the Relationship Between Job Knowledge and Differential Accuracy of Behavior Frequency Ratings for Simple and Complex Task Conditions ... 122

5. The Moderating Effect of Perceptions of Complexity on the Relationship Between Job Knowledge and Behaviors Recalled ..................... 123

xiii
The Effect of Job Knowledge and Task Complexity on Information Processing and Rating Ability

Introduction

Performance criteria in the workplace should ideally be based on objective indicators of work output. However, problems associated with such measures, such as their inability to be directly tied to the individual's work, have been well documented (Cascio, 1987). Dependence on supervisory ratings in lieu of objective measures has generally become an acceptable alternative. Supervisor ratings serve many important functions within organizations. Performance criteria for test validation, performance feedback, and information for promotion or bonus allocation are just a few areas where the role of supervisor evaluations are critical. However, supervisor ratings are also problematic due to their subjective nature.

A great deal of research effort has been expended to improve the accuracy and validity of performance appraisal ratings. Initially, the focus was on improving the format of the rating instrument, as in the development of more behavioral types of anchors designed to remove ambiguity and ease the evaluation process for the rater (Harari & Zedeck, 1973). Such attempts had minimal effects on improving ratings (Landy & Farr, 1980), and it became apparent that
research should move toward focusing on the role of the rater's cognitive processes (Feldman, 1981).

As a result of this transition, researchers' understanding of the rating process quickly evolved from a basic, traditional model (Borman, 1978) to a much more thorough, comprehensive perspective of the rating process, integrating principles from social cognition literature (DeNisi, Cafferty, & Meglino, 1984). Measures of rating accuracy began to be supplemented with assessment of memory-based ratings (Murphy & Balzer, 1986), free recall (Hauenstein & Kovach, 1986), conceptual similarity schemas (Kozlowski & Kirsch, 1987; Smither & Reilly, 1987), scripts and prototypes (Foti & Lord, 1987; Hauenstein & Alexander, 1991), as well as exploration of effects due to performance cues (Binning, Zaba, & Whatham, 1986) and cognitive categorization (Mount & Thompson, 1987). In general, the attempt was to better understand the rating process, and discern which variables were critical to ensuring high quality ratings. The tentative findings of this body of research suggest that despite raters' limited ability to handle the abundance of information provided in a typical work performance setting, accurate ratings can be obtained via effective information-processing strategies, such as classification or categorization (Feldman, 1981).

In exploring the role of information processing in
performance appraisal, Walker (1989) attempted to test the
generalizability of information processing strategies used
by college students in performance appraisal studies.
Walker had students and carpenters rate the performance of
lecturers and carpenters. By crossing rater populations, he
hoped to show that knowledgeable raters (e.g. students
rating lecturers, carpenters rating carpenters), would
display similar processing strategies and more accurate
ratings compared to raters evaluating unfamiliar performance
areas. Results were mixed as the hypotheses were supported
for the carpenter performance films only.

Walker (1989) discussed his results in terms of the
distinct differences in the nature of the two sets of
performance films. The lecture tapes appeared to provide an
overly simple rating task. The cognitive processes used to
evaluate carpenters and lecturers seemed to differ as a
function of task complexity. Levels of lecturer performance
were easily distinguished due to gross differences depicted
in the stimulus tapes compared to carpentry performance
levels. This assertion was based on the ease with which all
raters, students and carpenters, were able to discriminate
among the three lecture performance levels, as well as the
high interrater agreement among subject matter experts when
providing true scores. Apparently, rating the lecture
films provided a simpler task compared to the carpentry
tapes. This suggests that not only is the rater's knowledge of the job important, but so is the nature of the task being evaluated.

In order to fully understand the dynamics of information-processing strategies within the context of performance appraisal, the role of both variables, job knowledge and task complexity need to be explored simultaneously. Demonstration of the effects of job knowledge and task complexity on raters' processing of information, and their rating outcomes also has implications for many of the laboratory studies in the performance appraisal area. It explores the appropriateness of assumptions of processing invariance (i.e. information is processed the same way by different rater populations), as well as different performance domains.

The present study will explore the potential role of task complexity as an essential pre-condition to the demonstration of a relationship between job knowledge and different processing strategies, as well as the accuracy of the ratings. This will be achieved by the assessment of job knowledge, and the manipulation of the nature of the task (simple versus complex) within the same occupational domain.

The Importance of Job Knowledge

In discussing job knowledge, the first step is to clarify the variable's distinction from familiarity and
expertise. In some research studies (Kozlowski, Kirsch, & Chao, 1986), the operationalization of familiarity relates to the degree to which the rater is familiar with the ratee or incumbent of the job, as determined by the degree to which the supervisor interacts with, or is exposed to a particular individual. In other studies (Walker, 1988), familiarity relates to understanding of the facets or dimensions of the job per se, independent of the incumbent in that position. In contrast, expertise would encompass knowledge of the job, as well as proficiency in carrying out that particular job (Posner, 1986). For example, an individual who is knowledgeable about many tools and their functions, as well as the advantages and disadvantages of different types of woods may be knowledgeable about carpentry, but far from an expert. To qualify as an expert carpenter, an individual would have mastered the use of tools and the production of sufficient products from the wood.

In the context of the present study, job knowledge relates directly to understanding the components of the job, and not the particular individual in that job; but does not require actual proficiency in actual job performance. Thus, job knowledge would reflect a thorough understanding of the job being evaluated similar to that which would be expected of a supervisor who may never have performed the job, but
must evaluate his/her subordinates who do.

A recent social cognitive model of the person memory-judgment relationship (Srull & Wyer, 1989) describes the possible distinction between the information processing strategy used by knowledgeable raters, and those who lack relevant job knowledge. In general, knowledgeable observers seem to encode behavioral information into memory by organizing specific behaviors into trait clusters. This is a more effective means of processing behavioral information for the purpose of performance rating since the evaluative judgment is stored and then recalled, in contrast to storing each specific behavior individually, recalling those behaviors, and finally forming a judgment. The latter process is much more difficult and demanding. In order to adopt the more effective, trait-level processing strategy, the observer needs to be confident in their "trait-based expectancies" with regard to the target stimulus (Srull & Wyer, 1989, p.67). Evidence suggests that this type of processing is facilitated by greater job knowledge.

Within the performance appraisal research, investigation of the job knowledge variable supports the model discussed above. Smither and Reilly (1987) found that raters with high job knowledge had more reliable conceptual similarity schemas compared to low job knowledge subjects. Similarly, Kozlowski and Kirsch (1987) found that high
levels of job knowledge lead to more effective processing strategies using well developed prototypes, which in turn lead to improved rating accuracy. Finally, Hauenstein and Kovach (1988) demonstrated that a rater's category familiarity influences how information is processed in appraisal settings.

Without knowledge-based interpretive frameworks, information is processed at a more superficial level depending solely on the strength of interbehavior associations, links between each individual behavior and related behaviors (Srull & Wyer, 1989). This type of processing fails to integrate the information into more basic and meaningful trait categories. As a result, when the observer needs to form evaluative judgments in the future, they may lack specific behavioral detail due to the less efficient processing strategy at time of observation. On the other hand, those with knowledge-based, interpretive frameworks should be able to formulate accurate evaluative judgments, but due to the nature of their processing strategy, the specific behaviors on which those judgments are based may not be retrievable (Srull & Wyer, 1989).

Markus, Smith, and Moreland (1985) explored these propositions by assessing the effect of pre-existing cognitive structures for a particular domain (self-schema) on an individual's social perception. Performance of
subjects with masculinity self-schemas were compared to subjects without such schemas in a unitizing task. Subjects were asked to chunk sequences of a film depicting schema-relevant events into meaningful units. Those with masculinity self-schemas (schemas) organized information into larger units than those subjects without such self-schemas (aschemas). This was attributed to the subject's expertise in that domain as reflected in the self-schema. In addition, the same subjects were able to shift their processing strategies and produce smaller units when given instructions to attend to the details of the film.

Similar evidence was provided by a free recall task. When asked to recall as much detail as possible, schematic subjects recorded a greater number of observations and tended to group them into fewer "similarity" piles, analogous to the larger units chunked. The researchers were able to reveal a more detailed picture of the information processing strategies used by the two groups of subjects. When the subject's purpose of observation was manipulated, and the recall responses were divided into behaviors and traits, under conditions of no instruction or instruction to form an impression, schematic subjects tended to provide significantly more trait evaluations than aschematic subjects. In addition, the number of trait-based evaluations reported by schematics were greater than
behaviors reported. Under a condition of instruction to recall as much detail as possible, the latter finding was reversed. Schematics recalled more behaviors than traits; while observations recalled by aschematics were evenly distributed over the two categories.

These findings are relevant to research exploring the underlying cognitive processes by which impressions are formed. By manipulating observational purpose, impression formation versus memory for behaviors, distinctions in the processing strategies used by raters can be revealed. This manipulation is discussed further below.

**Job Knowledge and Observational Purpose.** Manipulation of observational purpose has been successfully conducted many times in a research setting (Hamilton, Katz, & Leirer, 1980; Srull, 1981; Srull, Lichtenstein, & Rothbart, 1985; Wyer & Gordon, 1982), and it is apparent that different observational objectives lead to different representation of performance information in memory (Srull, & Weyer, 1989). Typically, prior to observing a target individual, subjects are asked to form an impression of the target, or remember as much detail as possible. The purpose of this instructional manipulation is to vary subjects' perceptual set, the way in which subjects encode the information they observe. Under directions to form an impression, subjects will tend to process information with an evaluative goal,
viewing performance with a critical eye. Attending to detail is similar to rote-memory approaches to retaining content. The subject gives very little evaluative assessment of the information, but rather attempts to remember what was done and not how well it was done. Subsequently, subjects are given a free recall task and asked to evaluate the individual. The nature of the items recalled, judgments or behaviors, indicates the type of information on which the evaluation is based.

Different results on the recall task are predicted for raters depending on how knowledgeable they are of the performance domain. Under instructions to recall as much detail as possible, raters lacking knowledge will recall fewer behaviors compared to knowledgeable raters. On the other hand, when asked to form impressions raters lacking knowledge may, in fact, recall more behaviors than knowledgeable raters. However, they will be extremely inferior to knowledgeable raters in recalling judgments, as well as in the accuracy of those judgments.

The advantage of knowledgeable raters seems to be due to the pre-existing interpretive framework within which knowledgeable raters operate (Markus, et al., 1985), but why do such raters lack the behavioral detail to back up their ratings when forming impressions? As suggested above, when raters possess interpretive frameworks about the job, at the
time of observation behavioral information is more efficiently organized by integrating specific behaviors into more general trait information. This occurs at the cost of the demanding task of attending to, storing, and eventually recalling specific behaviors (Srull & Wyer, 1989). In the present study manipulation of observational purpose will be used to explore and identify distinctions in processing strategies.

Task Complexity

The issue of task complexity has been explored in a variety of research contexts (Campbell, 1986). Wood (1988) suggests that three factors contribute to describing task complexity: component complexity, coordinative complexity, and dynamic complexity. These factors are reflected in the amount and interrelatedness of products, acts, and information cues which make up the key elements of tasks. Component complexity concerns the number of acts performed or information cues processed in completing a task. Coordinative complexity involves the degree to which products, acts, and information cues are interrelated and rely on critical sequencing. Dynamic complexity reflects the amount of flexibility required to adjust to changing situations. Campbell (1988) emphasized the central role of increased cognitive demand in defining task complexity. Similar to Woods' factors, in Campbell's framework sources
of task complexity are identified based on their contribution to increasing information load, information diversity, and rate of information change.

From a rater's perspective, complex tasks would be characterized by multi-faceted elements involving less repetition and consistency of behavioral information, greater variety of relevant behaviors displayed, and more ongoing flexibility potential on the part of the performer due to changing demands. The evaluation difficulty associated with complex tasks is primarily due to the limited availability of highly diagnostic information to help determine performance level (Bodenhausen & Lichtenstein, 1987; Schul, 1985). Finally, Hauenstein and Walker (1989) posited that the degree of ambiguity between individual and overall group performance would also affect information diagnosticity. Thus, evaluating a complex task embedded in a work group structure is even more difficult when that particular task is only indirectly linked to the group performance outcome.

Bodenhausen and Lichtenstein (1987) demonstrated that different heuristics or processing strategies may be employed under conditions of increased task complexity. Subjects judging a complex situation demonstrated greater reliance on stereotypes to organize information. This resulted in subsequent erroneous recall affirming the
stereotype. The authors demonstrated that compared to complex tasks, such integrative processing strategies are not necessary for simple tasks since incoming information is highly diagnostic. For example, the typical behaviors that would be observed when viewing a running back are highly salient and unambiguous in terms of distinguishing good from poor performance. Thus highly diagnostic behaviors are informative to the observer attempting to form an impression.

In summary, a full understanding of information processing is incomplete without an investigation of how the complexity of the task or job interacts with levels of job knowledge to affect process and rating outcome measures. The present study will explore these effects to add to our understanding of the cognitive processes underlying ratings, and to provide some commentary on the generalizability of performance appraisal laboratory studies.
Literature Review

Familiarity and Job Knowledge Research

The research discussed below pertains to raters who vary in terms of the degree to which they are familiar or knowledgeable about the target job in question. These two factors are exclusive to ratee familiarity which represents the degree of exposure to and interpersonal contact with the job incumbent (Kozlowski, et al., 1988); and expertise, the display of proficiency in actual performance of a particular job (Posner, 1988). As defined, familiarity and knowledge of the job are closely related and may involve very similar dynamics among raters differentiated only by their operationalization. On the other hand, there presently does not seem to be any studies which explore the two orthogonally. If a distinction were to be made it would seem to be that familiarity is based on the level of prior exposure to the job, and length of observation of actual job performance. In contrast, a strict definition of job knowledge would entail the level of aptitude concerning the tasks, duties, rules, and regulations pertaining to a particular job. As previously stated, the two are probably very closely related.

Within social cognition literature, researchers have explored the influence of self-schemas (Markus et al.,
1985), on social perception. The self-schema was developed by Markus and her colleagues (1985) who asserted that schemas tend to have the greatest effect on information-processing when the information is relevant to one's self. Subsequently, the self concept is one particular type of self-schema in that it is defined as "a set of cognitive structures that provide for individual expertise in particular social domains" (Markus et al., 1985, p. 1404). In testing the effect of self concepts on social perception, subjects were divided into two groups, schematic or aschematic with regard to masculinity. Both groups were asked to perform a unitizing task while viewing a film containing masculinity relevant information, and to perform a free recall task after the film was viewed. Unitization requires subjects to segment the film into meaningful units of action by pressing a button each time the subject perceives such a completed unit. The researchers found that subjects with stable self-schemas relevant to the stimulus information did process and organize the information differently.

Attributes that are relevant to an individual's disposition influence one's social perception in that they develop a greater understanding of related characteristics and behavior manifestations that accompany such attributes. As a framework for organizing and understanding many of our
life experiences, the self-schema's influence on perception of the self largely parallels the influence familiarity has on perception of other's within a particular domain.

The results of the two experiments conducted by Markus and her colleagues have already been described within the introduction. The implications of these results, and reviews of other studies suggest that there are some critical differences between people who do and do not possess a relevant self-schema for any particular domain. Those who possess relevant self-schemas are better able to distinguish between relevant and irrelevant content in comparison to that domain, to organize content material into more definable and meaningful units, retrieve a greater amount of information more accurately, and shift their processing strategies as a function of the observational purpose given to them (Markus, et al., 1985). It is expected that similar processes may occur in organizational contexts for raters who are knowledgeable about a performance domain compared to raters who lack such knowledge.

A great deal of the research in this area is driven by Srull and Wyer's (1989) model of people's memory processes when forming impressions and making judgments. In general, the model suggests that individuals interpret and encode observed behaviors in terms of an applicable and accessible
personality trait which summarizes the behavioral information. Instead of a personality trait, it may become a general overall impression which guides the interpretation of observed behaviors. A corollary to this general idea is the notion that memory organization varies as a function of the trait expectancies of the perceiver. Having trait expectancies facilitates the evaluation of incoming information; whereas those without trait expectancies may need to process each behavior into different traits that may be inconsistent. The result is that no overall impression can be made, and that such inefficient processing leads to stronger behavior-behavior linkages, instead of behavior-trait linkages. This is related to judgment processing in that judgments tend to come from one's assessment of the traits available or inferred relevant to the target being judged. The implication of this aspect of the model, and what was tested in the studies discussed above is that high levels of familiarity and/or knowledge should provide the trait expectancies needed for a more efficient and judgment-oriented strategy of information processing. High levels of familiarity and knowledge may help develop the trait expectancies, but also seem to be good indicators of how frequently used, and therefore accessible trait expectancies are to the perceiver (Wyer & Srull, 1986).

Drawing from the empirical work of researchers in
social cognition, performance appraisal research began to look at similar phenomenon within their applied subject matter. Our understanding of the social cognitive issues in performance appraisal have developed a great deal since shifting focus from the appraisal instrument to the information processing by the rater (Landy & Farr, 1980; Feldman, 1981). This is a logical area of focus since so many organizations rely on judgment measures for performance evaluation. The goal of this line of research is to better understand how raters process information, and form their ratings so as to eventually improve the accuracy of those ratings. Below is a review of studies which have investigated the effects of job familiarity or knowledge on social cognitive or rating accuracy measures in the area of performance appraisal.

Using the performance domain of baseball, Kozlowski, Kirsoh, and Chao (1986) asked subjects to judge the degree of similarity between seven offensive performance dimensions: batting average, home runs, base hits, runs scored, strike outs, stolen bases, and runs created. Perceptions of similarity of each possible pairing were obtained and compared to objective interdimensional covariations obtained from actual statistical records. Raters with high job knowledge were significantly more sensitive to actual performance covariation than low
knowledge raters. In addition, when providing ratings of familiar players, high knowledgeable raters tended to show higher correlations between ratings and true scores, as well as less halo than low knowledge raters. In both knowledge conditions these rating and true score correlations tended to decrease and halo increase as the target ratees became less familiar to the raters. Unlike the present study, Kozlowski and his colleagues determined levels of knowledge by scales that measured the amount of exposure to the sport.

Using similar measures and methodology, Kozlowski and Kirsch (1987) introduced a three week interval between stimulus presentation (end of the baseball season) and ratings. It was hypothesized that this would result in more memory-based judgments and cause greater reliance on perceived conceptual similarities among dimensions to guide performance ratings. For high knowledgeable raters, the results supported a systematic distortion hypothesis (Shweder, 1982) which predicts that the relationship between conceptual similarity and actual ratings will increase as ratees become less familiar since the subjects have less information to draw on and must rely on their implicit schemas to arrive at their ratings. This was accompanied by a decrease in rating and true score covariation which was reflected in a significant decreases in all four accuracy components. The same dynamics were not observed for medium
and low knowledgeable raters. In fact, their conceptual similarity - ratings correlations dropped and rating-accuracy correlations virtually remained the same. The decrements in accuracy displayed by high knowledgeable raters also did not occur. However, high knowledge raters accuracy components did not drop below the other knowledge groups. These results were interpreted as an illustration of better prototype systems for knowledgeable raters enabling them to more effectively encode, and recall performance-relevant information. Prototypes are exemplar representations of particular categories which assist raters in making judgments that reflect cognitive economies by promoting the formation of impressions using a few salient features of the comparison referent (Cantor & Mischel, 1979). It was proposed that the prototypes facilitated easier assignment of behavioral information to the proper performance dimensions. Conceptual similarity schemas assist knowledgeable raters when they make evaluations, especially without clear or sufficient performance information concerning any particular ratee. Low knowledgeable raters, on the other hand, lacking well-developed prototypes were not influenced by this process, and therefore did not have a guide to efficiently handle the category assignment of behaviors, especially when the target ratees were unfamiliar.
The relationship between conceptual similarity and job knowledge was also investigated using two separate subject populations, students and market researchers, who provided their conceptual similarity for the job of market researchers (Smither & Reilly, 1989). The market researchers, operationalizing high knowledge, provided more reliable conceptual similarity profiles, and this finding was corroborated by looking at levels of job knowledge within each subject population.

Kozlowski and Ford (1988) investigated rater strategies for information acquisition. Some of the variables they manipulated were ratee familiarity, delay between exposure and rating, and ratee performance level. Increases in familiarity led to decreases in information search. In addition, two significant interactions suggested that 1) for ratees who were familiar to the rater, the magnitude of information search would increase if the ratee showed low performance levels, and 2) when the delay was brief, the search for low familiar ratees was greater. Taken together, the researchers felt that these findings indicate that when raters are generally uncertain or unfamiliar, specific ratee behavior is sought for and used as the basis for any general evaluative or categorical assignment that would be helpful in arriving at a judgment.

Using an interview setting, Hauenstein & Kovach (1988)
investigated the effects of category familiarity on raters' processing of applicant information. It was hypothesized that raters who were unfamiliar with relevant categories would process at a behavioral level, organizing information around specific concrete descriptors, such as appearance and actual behaviors. Raters who were familiar with relevant categories would process at a categorical level, organizing information around trait inferences generated from behaviors. The researchers used a self-report measure of familiarity, which focused on experience with, and knowledge of the interview process. After being shown videotaped staged interviews, subjects were asked to respond to free recall and behavioral recognition measures. Hauenstein and Kovach's hypotheses were supported in that raters with high categorical familiarity recalled fewer specific behaviors, but made more trait inferences than unfamiliar raters. Also, unfamiliar raters were more accurate on the behavioral recognition measures. Since interview familiarity was based on self-report data, the researchers reported some concern for biases. As a result, a second study addressed the potential self-report response bias for the familiarity operationalization by recruiting subjects on the basis of interview experience. Hypotheses were again supported. Category familiarity moderated the information processing strategy. Despite using an apparent categorization
strategy, familiar raters did not show a decline in behavior recognition accuracy compared to unfamiliar raters, suggesting that the familiar raters' style is much more efficient at processing applicant information.

Hauenstein and Walker (1988) found that level of job knowledge determined the direction of the correlation between recall of behaviors and rating accuracy. Raters with high knowledge demonstrated a significant positive correlation, $r = .63$, while raters with low knowledge demonstrated a significant negative correlation, $r = -.51$. For low knowledge subjects, a strong negative correlation was obtained between judgments and accuracy, however the high knowledge effect was not obtained with judgments. In addition, the moderating effect of job knowledge occurred only under complex conditions. Task complexity is discussed below.

**Task Complexity Research**

The complexity of a task, whether it be a judgment task or behavioral task, typically has many qualities or characteristics which contribute to its level of complexity. In the area of performance appraisal, complexity may concern the intricacy and/or amount (overload) of information processing on the part of the rater, or it may be reflected in the nature of the actual job being performed by the
target ratee. The two factors may even interact with each other. The operationalization employed in the present study does seem to involve both aspects. The following review will discuss some of the relevant literature pertaining to task complexity in general, and within the social cognitive as well as performance appraisal literature.

The hypotheses for Hauenstein and Walker's study were driven by the notion that simple tasks tend to provide highly diagnostic information, small doses of which can lead to accurate impressions. Therefore, a great number of judgments, or even behaviors may not be needed. Since complex rating tasks tend to involve behavioral information which is often inconsistent and contains content that provides little diagnosticity, accurate impressions are going to be dependent upon more performance information, which sometimes may not be available. As stated above, Hauenstein and Walker (1986) found that job knowledge moderated the memory-accuracy relationship, but only under complex task conditions. The correlations were nonsignificant under low complexity task conditions. The researcher's based the validity of their complexity manipulation on the assurance by subject-matter experts concerning the greater behavioral and intraindividual performance variation, as well as the ambiguity of linking target performance to overall performance. Other
performance appraisal research suggests that an judgment task becomes particularly difficult and complex when true interdimensional correlations are low (Smither & Reilly, 1987; Smither, Barry, & Reilly, 1989).

In some research contexts, complexity has been defined as information overload. Although not used to operationalize task complexity in the present study, research which has manipulated this variable may shed some light on the cognitive processes in demanding situations. Information overload may divide or cause the withdrawal of one's attention from the target person, the possibility increases that the observer may need to rely on inferences, perhaps drawn from schemas instead of direct behaviors (White & Carlston, 1983). Schemas often function to simplify processing by serving as a heuristic or short-cut method of handling information. White and Carlston (1983) primed subjects with information concerning a particular target ratee. This established a schema before subjects were actually exposed to the ratee. When put in control of focusing on two ratees, subjects tended to focus on the individual for whom they had a preconceived schema only when that individual displayed schema-inconsistent behavior.

Bargh and Thien (1985) manipulated information load, and assessed its effects on information processing for subjects with different levels of construct accessibility.
Construct accessibility refers to the availability of meaningful and relevant constructs for any particular stimulus domain. In their study, Bargh and Thien used information concerning the honesty of some target individuals. Construct accessibility is somewhat related to the present study's job knowledge manipulation. Information load was operationalized as either having information presented at a constant rate (high overload), or having control of the rate of presentation (low overload). The increase in information load did not significantly affect those subjects with highly accessible constructs in recalling incongruent information. The study demonstrated that those with construct accessibility (chronic subjects) were able to form impressions even under conditions of high processing demand compared to nonchronic subjects, those who were not familiar with the honesty construct.

In addition to manipulating processing load (high and low), Bodenhausen and Lichtenstein (1987) manipulated complexity levels of social judgments. After providing a description of a target individual and the context of some aggressive behavior performed by that individual, subjects were asked to either judge the trait applicable to that target individual (low complexity), or the individual's guilt (high complexity). Subjects in the high complexity condition often displayed biases in their judgments.
concerning the person's guilt, aggressiveness, or future behavior suggesting that in such condition of judgment complexity, subjects tended to rely on heuristics to achieve the purpose of their information processing. This phenomenon occurred independent of processing load.

Foti and Hauenstein (in press) manipulated three levels of information load: minimal, typical, and taxed conditions. Also, subjects were given either a good, poor, or no performance cue to manipulate expectations. After viewing a videotaped lecture, subjects were given a behavior recognition questionnaire. Raters in the positive impression condition were not as accurate at recognizing positive ratee behaviors, while negative ratee behaviors were not affected by any performance cue condition. Increases in processing demand led to greater influence from the performance cue in determining to which ratee behaviors the subjects would attend. Greater overload influenced information processing in that subjects seemed to rely on previous impressions in storing behavioral information. This suggests that the nature of the task may also lead to changes in information processing.
Description of Study

The effects of rater knowledge and task complexity on process and outcome measures was investigated. In addition, subjects instructions concerning the purpose for observing the performance was manipulated in order to draw out the different strategies of information processing used by raters as a function of job knowledge. The present study contributes to the growing body of research integrating the principles of social cognition with performance appraisal by simultaneously assessing levels of rater job knowledge and task complexity (simple versus complex), and attempting to demonstrate that job knowledge is a critical variable, particularly when judging complex tasks.

The performance domain of football was used to explore this issue. Levels of job knowledge were assessed by way of a general test of football knowledge. Task complexity (simple versus complex) was varied by the player position serving as the target ratee. Observational purpose (memory for behavior versus impression formation) was manipulated as the strategy adopted by the subject for viewing the film as instructed by the experimenter. The advantage of using football performance as the stimulus is that both independent variables, job knowledge and task complexity, can be manipulated without having to venture into totally different populations of raters, and the target stimulus
film will involve actual task performance. This latter advantage is an improvement over the "staged" videotaped performance popular in performance appraisal research (Murphy & Balzer, 1986; Williams, DeNisi, Meglino, & Cafferty, 1985).

Process measures and outcome measures were obtained. Information processing was assessed using a free recall task, specifically the proportion of judgments to behaviors recalled. Outcome measures consisted of both behavior frequency ratings (e.g. how many blocks thrown, how many times carrying the ball), and judgmental ratings (e.g. rate overall performance).
Hypotheses

To organize the predicted findings, hypotheses will be presented for process measures first, followed by outcome measures.

Hypotheses concerning process measures:

Hypothesis 1: More information will be recalled in the simple task condition than in the complex task condition.

Hypothesis 1 reflects the general difficulty of processing complex information (Bodenhausen & Lichtenstein, 1987). In subsequent hypotheses task complexity will be predicted to interact with other variables of interest.

Hypothesis 2: The effects of job knowledge on memory of information will occur in the complex task condition, and not in the simple task condition.

Hypothesis 2a: Within the complex task condition, the type of observations recalled will vary as a function of job knowledge such that subjects with greater job knowledge will recall more judgments than subjects with lesser job knowledge. Subjects with less job knowledge will recall more behaviors than subjects
with greater job knowledge. No such relationships will be demonstrated in the simple task condition.

Hypothesis 2b: Within the complex task condition, observational purpose will moderate the relationship between job knowledge and recall of behaviors such that job knowledge will be positively related to recall of behaviors in the memory for behavior condition, and job knowledge will be inversely related to recall of behaviors in the impression formation condition. No such relationships will be demonstrated in the simple task condition.

Hypothesis 2c: Within the complex task condition, observation purpose will moderate the relationship between job knowledge and free recall of judgments such that job knowledge will be positively related to recall of judgments in the impression formation condition. No such relationship will be demonstrated in the simple task condition.

Hypothesis 2 is based on Walker’s (1989) conclusion that the teaching tapes were too simple to demonstrate processing differences due to knowledge of the domain. Thus, it is predicted that the effects of job knowledge on
memory will be seen only in the complex task condition. Hypotheses 2a through 2c specify the predicted effects of job knowledge in the complex task condition. More specifically, it is expected that relative to less knowledgeable raters, knowledgeable raters will exhibit a greater bias to recall judgments of performance, regardless of observational purpose (hypothesis 2a). However, the effects of the general bias predicted in hypothesis 2a will be facilitated in the impression formation condition (hypothesis 2c) and this bias will be reversed in the memory for behavior condition (hypothesis 2b). Predictions 2a through 2c are based on the expectation that knowledgeable raters will change processing strategies according to the observational purpose, but less knowledgeable raters are limited to inefficient, behaviorally-oriented information processing.

Hypotheses concerning outcome measures:

Hypothesis 3: Raters will be more accurate under simple information conditions compared to complex information conditions.

Hypothesis 4: The effects of job knowledge on the measures of accuracy will occur only in the complex
condition.

Hypothesis 4a: Within the complex task condition, observation purpose will moderate the relationship between job knowledge and rating accuracy such that job knowledge will be positively related to rating accuracy when given the impression formation instructions. No such relationship will be demonstrated in the simple task condition.

Hypothesis 4b: Within the complex task condition, observation purpose will moderate the relationship between job knowledge and behavior frequency accuracy such that job knowledge will be positively related to behavior frequency accuracy when given the memory of behavior instructions, and inversely related when given the impression formation instructions. No such relationship will be demonstrated in the simple task condition.

As in hypothesis 1, the information processing decrements due to increased information complexity are predicted in hypothesis 3. Similar to the prediction for process measures, hypothesis 4 predicts that the effects of job knowledge on accuracy will occur only when raters judge a complex task. Hypothesis 4a and 4b make predictions
concerning the moderating effects of observation purpose on
the relationship between job knowledge and accuracy for
behavior frequencies and judgmental ratings. The predicted
pattern of these outcome measures are once again based on
the prediction that processing strategies will change as a
function of observation purpose only with greater job
knowledge. Although knowledgeable raters form more
judgments and display more accurate judgmental ratings, they
sacrifice behavioral accuracy under impression formation
instructions.
Method

Subjects

The subjects were 123 undergraduate male students from either a large southeastern university or a small northeastern college who participated in return for class credit. Subjects from each school were equally distributed among the four experimental conditions.

Design

The design of the study was a 2 (Observational Purpose) X 2 (Task Complexity) design with Job Knowledge treated as a continuous individual difference variable.

Independent Variables

Observational Purpose. To manipulate observational purpose subjects were instructed by the experimenter and within the task directions (See Appendix A) to "form an impression of the players' performance", or "remember as much information about the players as possible". The former instructions conveyed to the subjects a purpose of impression formation (i.e. to arrive at an overall assessment or evaluation of the players), while the latter involved a memory-set purpose (i.e. to remember as many observations concerning the players as possible). In addition, subjects were informed that they subsequently would be asked to evaluate (impression-set), or recall behaviors (memory-set) regarding the target players. This
manipulation has been successfully used in previous studies (Foti & Lord, 1987; Wyer, Bodenhausen & Srull, 1984; Hamilton, et al., 1980a).

**Task Complexity.** Task complexity was manipulated by the position of the players assigned to be evaluated. Half of the subjects were instructed to focus on the two offensive tackles in the film (complex task) and the others were told to focus on the two running backs (simple task). Hauenstein and Walker (1992) chose this manipulation based upon coaches' judgments of the relative difficulty of evaluating various positions. They reported that coaches clearly believed that, relative to running backs, the performances of offensive tackles had greater behavioral variation, greater intraindividual performance variation, and more ambiguous linkages to overall group performance. As an indication of the differences in performance complexity between the two positions, the coaches used different passing grades for determining the success of each. For tackles, above 70 was considered a passing grade, whereas, 80 was the passing grade for running backs. Prior to judging performance, raters were informed of the passing grade for success associated with the position they were evaluating. This was done in order to parallel the rating scale with the one used by the subject matter experts.
Job Knowledge. Job knowledge was measured using a multiple choice job (football) knowledge test developed by Hauenstein and Walker (1982). This test appears in Appendix B. They found a coefficient alpha of .95 for the 45-item test and that the test was highly correlated with the sum of six self-report items related to personal football experiences (r = .76, p < .001). These self-report items were concerned with subjects' experiences in playing football and the extent to which they actively follow football (i.e., watch football on television; attend football games).

Stimulus Materials

Performance. All subjects will view a 16mm., silent film of 14 separate football plays taken from a game played between two NCAA division I teams in 1987. Football was used as the stimulus domain because it allows easy operationalization of job knowledge and judgment complexity, and the players' graded performances are available from the coaching staff for use as standards in the evaluation of rater accuracy. In addition, it overcomes some of the stimulus-bound problems of rating research identified by Funder (1987) by manipulating variables within the same performance domain, football. Each play shows the same team on offense and the opposite team on defense. Each play lasts between five and ten seconds and the order of
presentation of the 14 plays will be held constant across rating sessions.

Procedure

Subjects were run in groups (4 to 12) seated in front of a movie screen. Subjects were first told that they were to take a football knowledge test, and afterwards, they would view a football film and then return the following day to respond to some questionnaires. After completing the test, forms were collected and instructions for observation were distributed (See Appendix A). The experimenter read the instructions aloud as the subjects read to themselves. Subjects were told that they would view 14 separate football plays, and that each play would be seen three times, once forward, once backward, and once again forward. The use of multiple observations of each play is consistent with the typical evaluation strategy used by the coaches. Subjects were then told the color of the offensive team they would be following, as well as the two players upon which to focus. The positions were identified by a diagram highlighting where the two players stand for each play. The players were also designated by their uniform number. Finally, subjects were told that the experimenter would point to the two players they were to watch before every play. This would ensure that the raters evaluated the correct players. After the subjects verified that they understood this process,
they were once again given the observational purpose
manipulation verbally by the experimenter. After subjects
confirmed that they understood the purpose, they began to
view the football film.

At the end of the viewing session, subjects were
instructed to return to the same lab at the same time the
following day. The twenty-four hour delay was critical to
ensure that recall and rating tasks are truly based on
memory that would require the full process of encoding,
storing, and retrieval. This is the type of memory required
in actual evaluation contexts in organizational settings, as
opposed to short-term memory. The delay was to ensure a
valid test of the cognitive processes. At the second
session, subjects were given paper to write down "anything
they could remember" about the two players they had seen.
Each rater was given two pieces of blank paper, one for each
player. The player's uniform number was noted at the top of
the page (See Appendix C), and subjects had ten minutes to
list any comments they had about the players' performances.
During the recall task, subjects were exposed to a still
shot of the first play formation to serve as a possible cue
to facilitate recall. After completing the free recall
task, responses were collected and the subjects were
instructed to rate the overall performance of their target
players (See Appendix D). This was followed by a behavior
frequency rating questionnaire (See Appendix E), as well as a manipulation check questionnaire (See Appendix F). Responses to each questionnaire were collected before the next questionnaire was administered. After all manipulation check questionnaires were collected the subjects were debriefed and thanked for their participation. The entire experiment should lasted approximately 1 hour long (30 minutes each day).

Pilot Study.

In designing the study, a major concern was whether or not the relationship between job knowledge and task complexity was truly orthogonal. A 2 (High/Low Job Knowledge) X 2 (Position: Running Back/Offensive Tackle) pilot study was conducted to test the independence of job knowledge and task complexity. Type of information was a within-subjects factor. Seventy-two subjects were given the job knowledge test, and then viewed the film under similar procedures as described above, except subjects were directed to attempt to form an impression of all four players. To facilitate this demanding task, each play was shown three times forward and twice backward. The extra exposure was provided to allow subjects to sufficiently view all four players. Immediately following the film, subjects were asked to respond to questionnaires assessing the complexity of each position (See questions 1, 2, and 3 in Appendix F).
Subjects responded to three items for each position, and order of positions evaluated by subjects was counterbalanced. Using seven point scales (1 - not at all difficult; 7 - extremely difficult), the items assessed the subject's: 1) ability to critically observe performance, 2) ability to evaluate the performance, and 3) overall difficulty of forming impressions for each position.

In order to test the independence of each of the variables, job knowledge and judgment complexity, it was predicted that offensive tackles would reflect greater complexity scores compared to running backs, and that this would not entail a job knowledge main effect, nor an interaction between players position and job knowledge. In this case, job knowledge was classified as high or low based on whether or not subjects fell above or below the median score on the football knowledge test (31.187). Results supported the independence of the two variables. ANOVA with repeated measures, complexity ratings of each position, indicated no main effect for job knowledge, F(1,68) = 1.18; or interaction effect, F(1,68) = .15. However, a main effect for position, was significant, F(1,68) = 10.00, p<.01. Greater complexity was perceived for offensive tackles (M = 11.1) compared to running backs (M = 9.3). Thus, the orthogonal relationship between job knowledge and judgment complexity was supported.
Dependent Measures

Free Recall. The free recall data were supplied by requesting subjects to list and number each of their comments. Qualitative data was classified by two trained undergraduate students blind to experimental conditions. Each recall statement was classified as either a judgment (e.g., "the tackle did a good job"), or a behavior (e.g., "the tackle was called for holding"). Those statements which were behaviors tagged with judgments (e.g., "the runner often ran sweeps showing good acceleration"), were classified as judgments. This scoring protocol for free recall is analogous to that used by Markus et al. (1985), and is intended to represent a continuum from incidents observable on the film to statements representing complete conjecture on the part of the rater. Interrater agreement for the classifications made by the research assistants was .85, and all discrepancies were resolved by the experimenter prior to statistical analyses, and blind to the condition.

Overall Rating Scale. Each subject was asked to make two overall performance ratings, one for each of the assigned players. These ratings were made on a scale of one to one-hundred. Because multiple performance dimensions were not used, the traditional Cronbach (1955) operationalizations of rating accuracy were not directly applicable. Only two measures of accuracy were computed.
Elevation was computed in the normal manner. The second component of accuracy was differential accuracy, but the adjustment for stereotype accuracy present in the usual computation of differential accuracy is not applicable because of the lack of multiple dimensions. Therefore the calculations of this second component will closely reflect the calculation for differential elevation except there is no average across dimensions for each player, but rather a single rating (See Appendix G for accuracy formulae). These measures are further described in the section below.

Behavior Frequency Ratings. For the purposes of the study, it was necessary to create a set of behavior frequency rating scales (See Appendix E). This was accomplished in three steps. First, separate behavior categories for each position were generated. These categories were designed to be dichotomous (yes/no) for each play (e.g., How many plays did the tackle push his opponent backwards?; How many plays did the running back block?) Second, for each position, the film was viewed by the researcher and an assistant to determine whether an example of each behavior category occurred. All behavioral categories occurring on the film were included in the rating questionnaire (10 items for each running back, 9 items for each tackle). Also, behavioral categories not occurring on the film for at least one of the players were also included.
for both running backs and tackles.

Since the behavioral frequencies were rated on multiple categories with multiple ratees, all four of Cronbach's (1955) components of accuracy were used to score the behavior frequency ratings (See Appendix G). Elevation refers to the accuracy of the overall average rating aggregated across both ratees and dimensions. Differential elevation indicates a subject's ability to discriminate among players by comparing observed and true averages of each ratee across all dimensions (i.e., behaviors). A subject's stereotype accuracy refers to the ability to discriminate among dimensions by aggregating observed and true ratings across players. Finally, differential accuracy refers to the subject's ability to evaluate ratee differences within behavior dimensions.

**Target Scores.**

**Overall Ratings.** The current study used the same target scores as those used in Hauenstein and Walker (1982). Four players were chosen to be the target ratees for the study. Target scores for the performances of the players were provided by the players' coaches. The coaching staff reviewed each play on film several times before rendering consensus judgments of the players' performances. Consistent with their usual practice, the coaches only rated each player in terms of overall performance (i.e., there
were no specific dimensions of performance). Also, their overall rating scale ranged from zero (totally unsuccessful performance) to 100 (perfect performance).

**Behavior Frequencies.** The target behavior frequencies were generated as part of the developmental process for the behavior frequency rating scales. In order to determine the target frequency of behavior categories, the film was viewed by the researcher and a research assistant, both of whom were very familiar with football, and the positions viewed. Each play was reviewed repetitively until the viewers reached consensus concerning the presence or absence of each behavior category for each player. The target behavior frequencies represent the number of plays in which the behavior category was present. The true behavior frequencies for both positions are provided in Appendix H.

**Manipulation Checks.**

The manipulation check questionnaire included in Appendix F was used to assess the effectiveness of the operationalization of observational purpose and task complexity. Three items concerning observational purpose were evaluated by the subjects on five point scales. The specific items will assess the extent to which subjects: 1) "found the observational purpose clear" (1 - Totally unclear; 7 - Very Clear), 2) "tried to form an impression" (1 - None of the time; 7 - All of the time), 3) "tried to
memorize details" (same scale as #2). The success of the


task complexity manipulation was assessed by the same scale


as described in the pilot study above.
Results

Manipulation Checks

Subjects' responses to the manipulation check items are summarized in Table 1. A composite of each subject's responses to three different items (seven point scales) assessing the complexity of critically observing, evaluating, and forming an overall impression of the target players were assessed using a hierarchical regression approach in which three main effects: job knowledge, task complexity, and observational purpose were entered in the first three steps followed by each two-way interaction term, and finally the three-way interaction term. The results of this analysis are presented in Table 2. As indicated, results yielded three main effects influencing complexity perceptions.

Viewing running backs was reported to be less complex than viewing offensive tackles, $\sqrt{R^2} = .06$, $F(1,120) = 8.88$, $p < .01$. Observation of offensive tackles led to significantly higher perceptions of complexity ($M = 11.87$) compared to observation of running backs ($M = 9.83$). However, perceptions of complexity were also influenced by job knowledge and observational purpose. Although no relationship between job knowledge and perceptions of complexity was demonstrated in the pilot study previously discussed, it was found in the present study. Subjects with
less knowledge tended to report greater complexity. Knowledge and perceived complexity were significantly negatively related, $R^2 = .09$, $F(1, 121) = 11.32, p < .01$. In addition, the observational purpose manipulation also significantly influenced perceptions of complexity, $R^2 = .05$, $F(1, 119) = 7.42, p < .01$. Significantly greater perceptions of complexity were reported in the memory-set condition ($M = 11.60$) compared to the impression-set condition ($M = 10.20$). Results of the manipulation check items for task complexity present some concerns for interpreting analyses of process and outcome measures since complexity levels seem to be a function of other variables beyond the position being observed as intended. Specifically, complexity increased with decreasing levels of job knowledge and requirements to remember detail.

Table 1 also summarizes the results for three items assessing the observational purpose manipulation, and hierarchical regression results for each are provided in Table 3. There was no significant difference between the groups in their response to the extent to which subjects felt the observational purpose was clear (item #4 in Appendix F). The null finding between the impression formation and memory for detail group was expected since detailed instruction, and standardized procedures should not have yielded any more or less clarity of purpose for either
Item #5 asked subjects: "To what extent did you try to
form an impression of the target player's you observed?". As expected, subjects receiving the impression-set did indicate that they were more frequently involved in impression-formation ($M=5.23$), than those subjects in the memory-set condition ($M=4.40$), $\sqrt{R^2} = .11$, $F = 8.72$, $p < .01$. Unfortunately, the results for the final item were not as expected. When asked: "To what extent did you try to
c memorize details while observing the target players?", subjects in the impression-set ($M=4.13$) gave similar responses as those in the actual memory-set ($M=4.08$). This result may be due to the difficulty in attempting to
remember ongoing detail as reported above in both task
complexity conditions, or it may reflect the difficulty in
manipulating an information processing strategy under experimental conditions that does not come as naturally to the subjects as impression-formation. In either case, as with task complexity, results of manipulation checks for observational purpose also indicate some concern and problems in testing hypotheses.

Finally, as shown in Table 1, levels of job knowledge were similar in all four experimental conditions ($M = 32.04$, $33.72$, $32.99$, and $31.45$). A 2 (Task Complexity) X 2 (Observational Purpose) ANOVA yielded no significant main
effects or interaction.

**Hypotheses**

**Process Measures.**

The first two hypotheses concern the type of information being processed by subjects in the various experimental conditions, and at different levels of job knowledge. Therefore the results pertaining to the free recall data (which reflects the content of what subjects remembered from the performance stimulus) will be discussed in terms of the total amount of information recalled, and then broken down into the two classifications in which the statements were classified: the number of judgments recalled, and the number of behaviors recalled. Each of these three variables were subject to a hierarchical regression analysis. Task complexity and observational purpose, were effect coded: -1 representing the complex task (offensive tackles) and impression-set condition respectively; and 1 representing the simple task (running backs) and memory-set condition respectively. Job knowledge preceded task complexity which was entered just prior to observational purpose in the first three steps of the hierarchical regression analyses. This was followed up by entering the three two way interaction terms, and finally the three way interaction term. The descriptive statistics for each of these variables in the four experimental
conditions can be found in Table 4, and the results of the hierarchical regression analyses are summarized in Table 5.

The first hypothesis predicted that there would be significantly more information recalled in the simple task condition compared to the complex task condition due to the increased difficulty in processing more complex information (Bodenhausen & Lichtenstein, 1987). There was a significant positive relationship between the total amount of information recalled and job knowledge, $R^2 = .08$, $F(1,121) = 10.25$, $p < .01$. However, when the task complexity variable was entered it did not significantly increase the $R^2$ of the model; thus failing to support the first hypothesis. Subjects in the simple task condition ($M=10.88$) did not recall a significantly greater amount than subjects in the complex task condition ($M=10.16$). Neither the observational purpose main effect, or any of the interaction terms yielded significant changes in $R^2$.

Interestingly, when judgments and behaviors were analyzed separately, not only did the task complexity variable fail to yield a significant increase in $R^2$ for either type of recalled statement, but also job knowledge was not significantly related to either type of recalled statement. This was very unusual since total recall did yield a main effect. Upon closer examination, it was revealed that the explanation lay in the correlation between
behaviors recalled and judgments recalled. The overall correlations between job knowledge and the recall measures are provided in the first section of Table 6. The large negative correlation between behaviors and judgments, \( r = -0.62 \), suggests that subjects tend to be idiosyncratic in their style of free recall by using one type of recall or the other almost exclusively. It seems that this idiosyncratic tendency cannot be predicted by job knowledge or the nature of the task. In general, the total amount of information recalled is positively related to job knowledge.

Hypothesis two predicted interaction effects between the three variables: job knowledge, task complexity, and observational purpose on memory of information. Specifically, hypothesis 2a predicted a two way interaction between job knowledge and task complexity for number of judgments and number of behaviors. However, job knowledge would be positively related to judgments recalled and inversely related to behaviors recalled, under the complex task condition only. As stated above, neither job knowledge or task complexity were significantly related to behavior or judgments recalled. The interaction between job knowledge and task complexity was not significant for either type of recall statement. The relationships between job knowledge and either type of recall within both simple and complex task conditions are displayed in the bottom of Table 6. In
all cases, the relationship is weak and positive.

This suggests that the type of information processing strategy adopted by a rater (i.e. behavioral or evaluative traits) could not be determined by the level of the rater's job knowledge or the complexity of the task. Therefore, the present study failed to support the notion that knowledgeable rater's tend to rely on recalled judgments when evaluating ratees relative to less knowledgeable raters, especially under complex task conditions.

Hypotheses 2b and 2c predicted that the variable observational purpose, would moderate the relationship between job knowledge and the recall of behaviors and judgments, again under the condition of complex task only. Observational purpose did have a strong effect on behaviors and judgments recalled. Collapsing the means across task complexity conditions in Table 6, there were significantly more behaviors recalled under the memory-set condition (M=8.00) than in the impression-set condition (M=4.57) resulting in a R² = .17, F(1,119) = 25.89, p < .01, and significantly fewer judgments recalled under the memory-set condition (M=2.56) than in the impression-set condition (M=5.91) resulting in a R² = .22, F(1,119) = 34.92, p < .01. However, once again, interaction terms were not significant. The expected effect of observational purpose in causing a shift in the the processing styles, as
reflected in the type of information recalled, by knowledgeable raters exclusively in complex tasks was not supported as the significant main effect of observational purpose seemed to operate independent of both level of job knowledge and level of task complexity.

In summary, the first two hypotheses were not supported by the findings of the present study. Although job knowledge was shown to be positively related to the amount of information recalled, it was not significantly related to judgments or behaviors recalled when analyzed separately. In addition, the complexity of the task as operationalized by observing running backs (simple) or offensive tackles (complex) failed to show a difference in behaviors, judgments, or total information recalled, except when perceptions of complexity were controlled for. The observational purpose was very powerful in influencing the type of statements recalled. The three variables showed no significant interactions. Therefore, the expected ability of knowledgeable raters to shift modes of information processing strategies depending on the purpose assigned to them, as suggested by the person memory model proposed by Srull and Wyer (1989) was not supported.

**Outcome Measures.**

The outcome measures collected involved single overall
ratings of the two target players, and behavior frequency ratings on 9 (offensive tackles) or 10 (running backs) applicable behaviors for each player. It should be noted that accuracy measures are counter-intuitive in that larger values represent less accuracy. Similar to hypothesis one, the third hypothesis predicted greater accuracy under the simple task condition compared to the complex task condition. Hypothesis 4 predicted that job knowledge and accuracy would be related only under the complex task condition. More specifically, hypothesis 4a basically predicted a positive relationship between job knowledge and rating accuracy when subjects were given impression-set instructions in the complex task condition only. Hypothesis 4b went on to specify that within the complex task condition the relationship between job knowledge and behavior frequency accuracy would be moderated by the observational purpose such that job knowledge and accuracy for behavior frequencies would be positively related in the memory-set condition, and inversely related in the impression-set condition. Table 7 contains the descriptive statistics for each accuracy measure, and Tables 8 - 10 provide the results of the regression analyses. In addition, Table 11 presents correlations of job knowledge with each accuracy measure overall, within each level of the task complexity and observational purpose variables, as well as within each
experimental group

Examining rating accuracy first, results will be presented in the context of how they support or fail to support hypotheses 3, 4, and 4a. Due to the use of only one overall dimension, analyses are confined to two of the Cronbach (1955) accuracy measures, elevation and differential accuracy. Regressing elevation onto all three main effects, the two way interaction terms, and the three way interaction term yielded null results. For differential accuracy the only significance revealed was the main effect for observational purpose as subjects were more accurate under the impression-set condition (M=8.12) than under the memory-set condition (M=10.17). This suggests that memorizing details does hinder the ability to evaluate and form accurate impressions of each of the individual target players' performances.

In general, hypothesis 4 was not supported by the overall rating measures, nor was hypothesis 4a. For both elevation and differential accuracy job knowledge was not significantly related to accuracy. In addition, the complexity of the task did not lead to any differences in accuracy either. Subjects in the simple task had almost the exact same accuracy means as subjects in the complex task for both elevation ((simple M=7.82; complex M=7.57)), and differential accuracy ((simple M=9.20; complex M=8.10)).
The results for behavior frequency ratings will be discussed in terms of how they support, or fail to support hypotheses 3, 4, and 4b. The elevation measure for the behavior frequency ratings supported hypothesis 3 in that the task complexity variable provided /\ \ R² = .08, F(1,120) = 12.51, p < .01. Subjects in the simple task condition were significantly more accurate (M=0.67) than subjects in the complex task (M=1.02). Elevation for behavior frequency ratings also yielded a significant three way interaction, /\ \ R² = .05, F(1,115) = 7.61, p < .01, which is illustrated in Figure 2. This interaction supports hypothesis 4b. The strongest relationship between job knowledge and elevation is in the memory-set / complex task condition. The more knowledgeable raters tend to have more accurate behavior frequency ratings as predicted. The inverse relationship expected in the complex task / impression-set condition exists, but is weak, r = .07. Within the simple task condition, as expected the impression-set (r = -.08) and memory-set (r = .18) conditions did not demonstrate a significant relationship between job knowledge and accuracy (See Table 11). This reflects a tendency on the part of knowledgeable raters in the complex task condition to more accurately remember overall behavior frequencies after observing performance under the memory-set condition.

For the differential elevation measure for behavior
frequency ratings, a task complexity main effect, $\sqrt{R^2} = .14$, $E(1,120) = 21.81$, $p < .01$, again supported the third hypothesis in that raters in the simple task condition ($M=0.41$) tended to be more accurate than raters in the complex task condition ($M=0.71$). This effect was moderated by the observational purpose manipulation, $\sqrt{R^2} = .03$, $E(1,117) = 4.06$, $p < .05$, as illustrated in Figure 3. According to separate t-tests, within the simple task condition, a significant increase in accuracy occurred in the memory-set condition compared to the impression-set condition, $t = 5.47$, $p < .01$. This would be expected since memory for detail would facilitate the recall of behaviors. This did not occur with the complex task condition, suggesting that the memory-set directions given to subjects observing offensive tackles did not influence the subjects’ ability to store and recall specific behaviors, particularly discriminating between players in terms of the overall mean of behavior frequencies representing each player’s involvement.

Job knowledge was significantly related to differential elevation, $\sqrt{R^2} = .07$, $E(1,121) = 9.01$, $p < .01$. The positive correlation meant that more knowledgable raters were less accurate. As shown in Figure 4, this effect was much stronger for the complex task condition than for the simple task condition. Although observational purpose did
not significantly moderate this effect within the complex task, the job knowledge by task complexity interaction was significant, \( R^2 = .03, E(1,118) = 4.73, p < .05 \). A significant relationship between job knowledge and accuracy was displayed within the complex task condition, \( r = .44, p < .01 \), unlike the simple task condition, \( r = .08 \) (See Table 11). The stronger correlation was within the impression-set condition (\( r = .50 \) see Table 11) as hypothesized, but the positive correlation for the memory-set condition (\( r = .35 \)) was unexpected. Knowledgable raters were expected to lead to greater accuracy in the memory-set condition. Therefore, the observed correlation for that particular cell was in the opposite direction of hypothesis 4b, and the results for elevation discussed above. Unfortunately, experiencing memory-set instructions apparently did not evoke the shift in information processing predicted for knowledgable raters.

The results of the hierarchical regression analyses for both stereotypical accuracy and differential accuracy for behavior frequency ratings provided almost identical results for job knowledge. A significant negative relationship between job knowledge and each of the accuracy measures suggested that more knowledgable raters are more accurate in terms of stereotypical accuracy, \( R^2 = .04, E(1,121) = 4.73, p < .05 \), and differential accuracy, \( R^2 = .05, E(1,121) = \)
6.58, p < .05. The significance of the latter variable suggests that, in general, knowledgeable raters were more adept at accurately recalling behavior frequencies within each dimension, perhaps due to greater familiarity with the dimensions and a greater ability at distinguishing between some of the more similar dimensions for each task's target position (e.g. blocked a linebacker, blocked a lineman; tackled by a linebacker, tackled by a lineman).

In addition, for both stereotypical accuracy and differential accuracy, significant F values were obtained for task complexity. For stereotypical accuracy, \( \sqrt{R^2} = .28, F(1,120) = 48.08, p < .001 \), raters in the simple task (\( \bar{M}=1.33 \)) were significantly more accurate than raters in the complex task (\( \bar{M}=2.22 \)). The results for the differential accuracy measure were unexpectedly in the opposite direction, \( \sqrt{R^2} = .27, F(1,120) = 47.15, p < .001 \). Raters in the simple task (\( \bar{M}=2.22 \)) were significantly less accurate than the raters who participated in the complex task (\( \bar{M}=1.54 \)). Basically, the stereotypical measure findings support hypothesis 3; while differential accuracy measures support the opposite effect.

For differential accuracy the job knowledge X task complexity interaction was significant, \( \sqrt{R^2} = .02, F(1,118) = 3.93, p < .05 \). This interaction is embedded within a significant three way interaction, \( \sqrt{R^2} = .04, F(1,115) = \)
7.49, \( p < .01 \) which is illustrated in Figure 5. As indicated in the within cell correlations presented in Table 11, the overwhelming effect is the strong negative correlation between job knowledge and differential accuracy within the simple task / impression formation condition, \( r = -.75, p < .01 \). This result does not support hypothesis 4b, in that knowledge and accuracy were supposed to be related within the complex task condition, not the simple task condition. In addition, the positive relationship between knowledge and accuracy was supposed to occur under memory-set conditions, and in fact, the impression-set condition was expected to be inversely related to accuracy of behavior frequency ratings since knowledgeable raters were expected to encode the specific behavioral information into broader qualitative traits.

In summary, accuracy measures pertaining to overall ratings did not support the third hypothesis, nor were there significant interactions to demonstrate the moderating effects of observational purpose on the relationship between job knowledge and accuracy, as predicted by hypothesis 4a. For behavioral frequency ratings, support was obtained for the last two hypotheses. Except for differential accuracy, results from the behavior frequency ratings supported hypothesis 3 by demonstrating that task complexity does lead to a decrease in rating accuracy for recall frequencies.
within behavioral dimensions. Raters tended to be significantly more accurate in the simple task condition compared to the complex task condition for all accuracy measures pertaining to the behavior frequency ratings, except for differential accuracy. Unfortunately, similar findings for the task complexity variable were not found for the overall evaluative ratings since the evaluative rating data yielded null results, and relied on the use of a single dimension. This prevents any conclusions to be drawn concerning the relationship between behavior frequency accuracy and rating accuracy.

Hypothesis 4 and 4b received support from the elevation accuracy measure only. The three way interaction found for elevation supported the moderating effect of observation purpose on the relationship between job knowledge and accuracy in the complex condition. The only shortcoming was that the inverse relationship expected for the impression-set condition was not significant. As predicted in hypothesis 4, job knowledge did interact with task complexity for differential elevation accuracy. However, the significant job knowledge - accuracy relationship was in the simple task condition instead of the complex task condition. This was in contrast to the elevation accuracy findings and in the opposite direction of the prediction made in hypothesis 4b.
Stereotypical accuracy supported hypothesis 3, but no significant interactions were found to support hypothesis 4 or 4b. Finally, results for differential accuracy of behavior frequency ratings were in the opposite direction as the predictions offered in hypothesis 4b. Although unexpected, the strong correlation between knowledge and differential accuracy in the simple task/impression formation condition may be explained by categorization processes as suggested by some of the data pertaining to individual behavior items.

Job Knowledge and Perceptions of Complexity

Due to the three main effects obtained for the manipulation check for task complexity, one final set of analysis were necessary. Table 12 presents the results of a series of three step hierarchical regression analyses in which each of the dependent variables were reanalyzed. The order of the variables submitted were the subjects' job knowledge measure, perceived complexity composite, and the interaction of the two variables.

The R²'s generated when entering job knowledge would not have changed from the previous analyses, so there is no new findings to report for the first variable. The primary focus of this analysis is to investigate significant effects for the perceived complexity measure and its interaction
with job knowledge to shed further light on the role of task complexity in the performance appraisal context. Complexity perception main effects may show further support for hypotheses one and three where the actual task complexity manipulation was not supportive. In fact, judgments,\( r = -0.37, \sqrt{R^2} = 0.13, F(1, 120) = 17.66, p < 0.01 \), and total recall, \( r = -0.42, \sqrt{R^2} = 0.13, F(1, 120) = 19.41, p < 0.01 \), showed significant inverse relationships with complexity perceptions. These findings support hypothesis one in that complexity does detract from subjects' ability to recall information in general, and judgments specifically.

The number of behaviors was the only variable in the reanalysis to yield a significant interaction between the two variables, \( \sqrt{R^2} = 0.03, F(1, 119) = 3.86, p < 0.05 \). This interaction is illustrated in Figure 5. Job knowledge is positively related to the recall of behaviors when complexity perceptions are high, but the relationship between the two variables disappears when perceived complexity is low. Although the original hierarchical regression analysis for behaviors did not reveal any significant interaction, when perceptions of complexity are used hypothesis two is supported for recall of behaviors. This suggests that the complexity experienced by subjects does seem to moderate the influence of job knowledge on effectiveness of behavior recall.
Several outcome measures also displayed significant complexity main effects. Unlike the original analysis, hypothesis three was supported for differential accuracy of overall ratings, $\sqrt{R^2} = .05$, $F(1,120) = 6.58$, $p < .05$. The significant positive relationship with complexity perceptions, $r = .22$ suggests that differential accuracy increased as perceived complexity decreased (accuracy components are reverse coded). For behavior frequency ratings, significant complexity effects for stereotypical accuracy and differential accuracy repeated the task complexity main effects found in the original analysis. Stereotypical accuracy was related positively to complexity perceptions, $r = .23$, $\sqrt{R^2} = .03$, $F(1,120) = 4.23$, $p < .05$, and differential accuracy was related negatively to complexity perceptions, $r = -.16$, $\sqrt{R^2} = .05$, $F(1,120) = 6.95$, $p < .01$. The significant task complexity main effects found for elevation and differential elevation were not repeated for complexity perceptions.

As suggested by the manipulation check results, the source of the complexity perceptions cannot be determined with absolute confidence. In general, the analysis of the data using complexity perceptions adds further support for the importance of understanding the effect of different levels of complexity in the information processing and rating accuracy of performance evaluators.
Discussion

Walker (1989) explored processing and outcome differences as a function of job familiarity by crossing rater populations and performance stimulus (e.g., students and carpenters rating both teacher and carpenter performance). His primary focus was to test the constant familiarity assumption (i.e., raters need to be sufficiently familiar with job domain) against the constant category assumption (i.e., the target performance may be tested in the lab and results will transfer to actual supervisors familiar with the job) in order to critique the generalizability of performance appraisal lab studies. He asserted that many of the null findings revealed in his study may be attributed to the overly simplistic nature of the teacher films. Using staged lecture films to manipulate poor, average and good levels of performance, Walker asserted that the critical behaviors incorporated into the film may have been overly salient resulting in easy discrimination by any rater population. This critical difference between the lecture and carpentry tapes, therefore, may have led to an inappropriate empirical test of the two competing assumptions.

The present study explored the potential moderating effects of task complexity on the relationship of job knowledge to rater accuracy and information processing
strategies. This was facilitated by manipulating the nature of the task, simple versus complex. Using real performance within the same general work domain, subjects viewed either running backs (simple) or offensive tackles (complex). Differences in levels of complexity were verified by subject matter experts and pilot data. It was expected that job knowledge would be related to greater rating accuracy and differences in information processing within the complex task conditions only, and that knowledgeable raters would be able to demonstrate shifts in processing strategies as a function of the observational purpose given. Although hypotheses were generally not supported, the following section will provide some post hoc explanations that may shed some light on explaining the results, what limitations of the study need to be considered, and what implications can be made from the results.

**Manipulation Effectiveness**

The results of the pilot work discussed earlier demonstrating independence between the variables job knowledge and task complexity were not repeated in the present study. Manipulation of the position observed by subjects did influence perceptions of complexity as expected. However, the perception of complexity manipulation check was also significantly influenced by job knowledge and the observational purpose manipulation. The
significant negative relationship between job knowledge and perceived complexity which contradicted pilot study results is very unfortunate since it limits the interpretation of results. This effect suggests that not only do researchers need to manipulate aspects of the task in the operationalization of task complexity, but the subjects' knowledge of the task domain may confound the manipulation. The only explanation to account for the contradicting results of the pilot study and present data seems to be sampling error. The present data suggests that having different levels of background knowledge of a domain may influence perceptions of complexity for the same aspect within that particular domain.

In relation to the task complexity research previously discussed, it may be that varying the positions observed by the raters (running backs versus offensive tackles) may have been a manipulation of one aspect of task complexity, while manipulating observational purpose may have varied another aspect. The evaluation of running backs may not be as simple as the operationalization intended. Indeed, one of the aspects of complexity mentioned earlier was multiplicity of tasks which can be more readily attributed to running backs than to offensive tackles; although other aspects of complexity are more applicable to tackles, such as availability of salient, diagnostic information and lack of
direct link to overall success. Also, remembering detail probably is most likely more demanding in terms of cognitive load than impression formation, the more natural cognitive set.

Dimensions of task complexity described by Wood (1986) are component, coordinative, and dynamic complexity. These pertain to the number of acts or processed information cues required within the task, the interrelatedness of those acts/cues, and the need for flexibility within the task, respectively. Observation of running backs and offensive tackles can certainly be differentiated on the coordinative and dynamic complexity dimensions since evaluating offensive tackles requires the rater to integrate more multi-faceted, and subtle behavioral cues that are less directly related to the overall success of each football play compared to running backs, in general. This is based on information provided by subject matter experts (Hauenstein & Walker, 1989). These same aspects of task complexity are referred to by Campbell (1988) when he discusses levels of information diversity and rate of information change. Similar to Wood's component complexity, Campbell suggests a third, and perhaps the most basic dimension, information load. This may be the dimension of task complexity influenced by the observational purpose manipulation. The memory-set instructions given to subjects certainly seems to
demand much more than the impression-set instructions. Asked to remember as much detail as possible, subjects may have felt that the directions contributed to the complexity of viewing a running back, especially since they were also asked to recall that information twenty-four hours later. This would parallel some of the information overload research which has been shown to influence processing and outcome measures (Bargh & Thein, 1985; White & Carlston, 1983).

Bodenhausen & Lichtenstein (1987) manipulated task difficulty by varying levels of "expected" information load using bogus pages of information that the subjects believed they would use, but never did. Manipulation checks indicated that perceptions of task difficulty were not affected. Unlike that study, subjects in the present study actually viewed fourteen plays under the experimental directions given. In addition, since the manipulation check was administered at the end of the study, memory-set subjects may not have differentiated the recall task from the viewing task of the previous day when considering their response to the manipulation check items. The recall task was a surprise to the impression-set people, and therefore may have remained unassociated with the observation task.

The job knowledge effect on perceptions of task complexity may be a function of familiarity with, and
exposure to football. Both factors, familiarity and exposure, should be closely associated with knowledge level. In addition, these same factors may be confounding the complexity differences between the two positions. For instance, the use of offensive tackles may have been an effective manipulation of task complexity due to subjects' lack of familiarity with the offensive tackle position compared to running backs. Familiarity differences between the positions may occur regardless of subjects' knowledge level. Rosch (1978) suggests that category prototypes, which function in the processing of information, may develop as a function of an individual's experience with the categories associated with a particular entity, in this case the target position. Media attention and television focus are typically geared more toward running backs than toward offensive tackles. Thus, perceptions of lower levels of complexity for the running back position may be attributed to its higher profile in sports media compared to offensive tackles. The intended sources of task complexity are confounded, in this case by systematic rater differences in levels of familiarity with the two different targets (Foti & Luch, 1992).

For observational purpose, two of the three manipulation check items supported successful manipulation of this variable through the directions given to subjects.
Although the same manipulation check items were found to be significant in other research contexts (Foti & Lord, 1987), the lack of a significant difference between groups for the single manipulation check item referring to time spent remembering detail did not seem to be too problematic. The effectiveness of the manipulation did not seem to be diminished, and in fact, was substantiated by the results of the data. The expected main effects associated with the manipulation were indeed significant. In addition, the observational purpose was reportedly clear to both groups, and the counter item for time spent forming an impression also supported the effectiveness of the manipulation.

**Processing Differences**

The expected shift in information processing strategies as indicated by the moderating effects of the observational purpose manipulation within the complex task condition may not have been obtained for two reasons: the time delay between observation and recall, or the lack of experience of college students in developing memory-based evaluations.

First, the twenty-four hour delay was implemented as an improvement of other studies using the observational purpose manipulation with immediate recall (cf. Hauenstein & Walker, 1989) since delays ensure that evaluations and free recall measures were based on memory, and therefore a valid test of
the cognitive process models (Ilgen & Feldman, 1983; DeNisi, Cafferty, & Meglino, 1984). The null results associated with process measures in the present study do not preclude the possibility that performance information is being encoded at the level of evaluative traits at time of observation by knowledgeable raters. Processing differences were demonstrated in other research using unitizing tasks (Markus, et al., 1985), in which information is grouped during observation, as well as in studies using immediate recall (Hauenstein & Walker, 1982). Performance information may have been initially encoded differently as a function of job knowledge and observational purpose, but the time lapse may have diminished the processing effect differences.

Apparently, if performance information is processed differently by knowledgeable raters, as suggested in previous research, requesting subjects to retrieve information using the free recall task after a time delay may be insufficient for tapping into the different processing strategies. On the other hand, it may be that the time delay is appropriate, but was too short in this instance. Murphy, Philbin, and Adams (1989) found large differences in rating accuracy and recognition tasks when they varied the time delay, particularly a large decrease in accuracy occurred when measured after a three day interval compared to immediate ratings or one day interval
conditions. Nathan and Lord (1983) found that a two day delay influenced the number of behaviors recalled without influencing the accuracy of the subject's ratings compared to an immediate condition. They concluded that the rater's overall impression may guide the recall of behaviors, rather than the use of behavioral information to form evaluations. Finally, Murphy and Balzer (1986) also compared behavior and performance ratings taken immediately after observation and with a delay. Subjects viewed two sets of performance tapes. One they rated immediately, and the other set was evaluated after they returned the following day. They found that the memory-based ratings (delayed condition) yielded greater halo and accuracy compared to immediate ratings. Again this suggests a general trend toward a loss of behaviorally specific information and a greater reliance on general impressions when a time lag is implemented.

Since the Hauenstein and Walker study used the same performance stimulus tapes under impression-formation conditions only, a comparison of results may offer some understanding of the effects caused by the time lag. Hauenstein and Walker (1982), using immediate recall, found larger mean differences (although not significant) within the simple task condition compared to the complex task condition for both behaviors and judgments compared to the present study. These differences were seemed to be
eliminated as a result of two factors within the present study: 1) an increase in judgments within the complex task condition (M=5.8 versus M=4.6), and 2) a decrease in behaviors in the simple task condition (M=5.4 versus M=8.5); while behaviors in the complex task and judgments in the simple task condition were similar. As a result there were no differences between the two studies for the total amount recalled. This rudimentary analysis suggests that the effects of the time lag on memory and the type of information recalled may interact with the complexity of the task. In the simple task behaviors seem to be more difficult to recall due to the time lag; yet recall of judgments in the complex task may be facilitated by the delay. It may be that the integration of performance information into broader evaluative traits is easier or more natural with the simple task, and as a result specific behavioral information is more readily disregarded in terms of deeper processing. For more complex tasks the integration of information into broader evaluative traits may require longer processing or may be a function of making inferences from the general impression. These results suggest that the nature of the task and length of delay may interact to determine how the information is processed. These findings support Funder's (1987) ideas that processing phenomena may be stimulus-bound, dependent upon the content
of the performance domain observed.

In terms of job knowledge, Hauenstein and Walker's (1982) data suggested that knowledgeable raters may tend to recall more judgments and fewer behaviors compared to less knowledgeable subjects. Again, this trend was wiped out as job knowledge demonstrated only a weak positive correlation with both judgments and behaviors. The relationship between job knowledge and total recall was not significant in the Hauenstein and Walker study, and therefore the superiority of recall by knowledgeable raters may depend on whether or not recall is immediate or delayed.

The second issue concerns the fact that the subjects (i.e., college students) do not have actual supervisory experience, and are not used to the process of observation for evaluation purposes, or rating tasks in general. Also, it may be a more basic difference of motivation. Subjects whose incentive, course credit, is dependent solely on participation, may fail to exert the effort that such cognitive processes require. Research has shown that evaluation often requires cognitive processing that requires a great deal of mental effort (Murphy & Cleveland, 1981). This issue reflects an ongoing debate concerning the external validity of performance appraisal research employing college students. In reaction to the question of whether or not performance appraisal studies using college
students are externally valid, a review of studies with student and non-student subject populations found that the majority of these contained statistically significant differences between the two populations on process and outcome variables (Gordon, Slade, and Schmitt, 1986). This suggests that the generalizability issue is of justifiable concern, particularly for business-related domains. The use of football as the target performance domain may have improved upon some, but perhaps not all, of the problems suggested by this debate.

The attempt to connect the information processed, as determined by the free recall task, to the rating behavior or evaluation task may have been futile since subjects responses in each were not related. One way to conceptualize this notion is in terms of Hastie and Park’s (1986) distinction between on-line and memory-based judgments. On-line judgments, in contrast to memory-based judgments, are made at the time the information is encoded, and therefore memory retrieval and accuracy would not be correlated. It may be that subjects in all experimental conditions relied on their on-line evaluations when asked for their evaluation. Free recall responses, however, came from the detail they were able to remember. This detail may have had very little influence or relationship with the rating provided. This assertion is similar to that posited
by Murphy and colleagues (1988) who also attributed null findings to differences between the evaluation and behavior recognition tasks. The former involved on-line judgments while the latter was memory-based.

Two final notes relevant to the process measures concern a potential heuristic bias for the memory-set subjects, and the modality by which responses were collected. When asked to remember as much detail as possible, it may have been a tendency for some subjects in this experimental condition to use the two players as points of reference, especially since both players were being observed simultaneously. In other words, subjects remembered information by comparing the two target players and therefore their recall information was recorded as judgments since that was the heuristic they chose to use. This is one example of what is meant by idiosyncratic processing styles. Secondly, subjects' responses may have been contaminated by the translation of thought into words, and therefore judgments were recorded as pure behaviors or vice versa. It would be interesting to test differences between verbal and written recall data collection methods.

Task Complexity

Except for differential accuracy, the improved accuracy in the simple task condition was supported within the
behavior frequency ratings, but not at all in evaluation ratings. In addition, the role of task complexity within the hypothesized interactions were, for the most part, unsupported. One critical reason for this may be that the evaluative task involved in the simple task condition was not simple for the subjects. Task complexity was investigated in the present study in reaction to Walker's (1989) assertion that the teacher tapes were too simple in comparison to the carpentry tapes, and this systematic difference accounted for some of the null findings in his study. In attempting to demonstrate that task complexity is a pre-condition to revealing differences in processing and rating accuracy, it may be the case that two complex tasks were chosen instead of a simple and a complex task. With respect to the pilot data, it may be the case that the offensive tackles did lead to a more complex task compared to running backs, but in comparison to the performance stimuli presented on Walker's teacher tapes, running backs were also relatively complex. The implication is that a more thorough representation of the complexity continuum should have been manipulated, one that would have more effectively tapped the simpler end of the continuum. An alternative explanation to the pilot study results is that perceptions of complexity were confounded by familiarity as influenced by the media and typical television exposure
since running backs are much more frequently the focus or center of attention. This confounding may have led to more accessible, and therefore more useful, implicit theories and/or developed categories of running back performance.

It is also plausible that different types of task complexity are operating when the running backs / offensive tackles are used to operationalize the construct. Therefore any further attempts to manipulate it should be preceded by a thorough understanding of the construct and its various dimensions. For instance, running backs may perform a more diverse set of behaviors than tackles, but the offensive tackle behaviors are much more difficult to observe and evaluate since they are less salient, less connected to the overall success of the play, and involve more subtle differentiation between successful and failed execution. In addition, the task of evaluating the running backs seems to involve dealing with two distinct positions, fullback and halfback. The different functions and behaviors between the two running backs seem to be greater than between the two tackles. This may have contributed to the complexity level of evaluating running backs. As stated, this may explain why the simple task led to significant decreases in differential accuracy for behavior frequencies.

The hierarchical regression analyses lend further support to the potentially important role of the task
complexity variable in the information processing associated with performance appraisal ratings. The initial problems with the manipulation check provide a hint of the challenge facing researchers to better understand how the task, the purpose, and rater knowledge level influence the construct of complexity.

Job Knowledge

Before discussing the results of the data concerning level of job knowledge, some general concerns should be expressed relating to the measure of job knowledge and the evaluative rating measures. The construct validity of the job knowledge instrument may be of concern since it taps knowledge of football and not directly knowledge of the offensive tackle or running back job. This may account for why more and greater effects of job knowledge are not reported, especially for evaluative ratings. Also, the use of a single overall rating on a one hundred point scale may have been very burdensome for all subjects regardless of knowledge level. Although passing grades were provided, standards were not explicitly stated. As a result, criteria may have been extremely vague, and therefore varied a great deal among raters.

Similar to task complexity, the effects of job knowledge on accuracy were demonstrated among the behavior frequency ratings, except for differential elevation; but
not at all for evaluative rating accuracy. A large focus of the study was on the moderating effects of task complexity and observational purpose on the relationship between job knowledge and accuracy for both types of measures. Significant interactions were not found for the two evaluative accuracy measures. The hypothesized three-way interaction was demonstrated for elevation accuracy of behavior frequency ratings only. The three-way interaction for differential accuracy did not involve the relationships that were hypothesized. Significant interactions for differential elevation of behavior frequency ratings were in the opposite direction of the prediction made in the hypothesis, and no significant interactions were associated with the stereotypical accuracy measure. These divergent results across accuracy measures for the behavior frequency ratings may shed some light on the use of such measures to explore both rating styles and underlying cognitive processes.

The findings pertaining to stereotypical accuracy and differential accuracy may indicate different types of categorization processes for different tasks. For stereotypical accuracy, the two main effects taken together, suggest that knowledgable raters had an advantage as did subjects in the simple task condition. The lack of significant interactions suggest that the results for
stereotypical accuracy may be reflecting different cognitive processes than the results for differential accuracy discussed below. The increase in stereotypical accuracy associated with greater levels of job knowledge suggests that knowledgeable raters may have been more sensitive (Jackson, 1972) to the implicit theory of performance by offensive tackles and running backs, and that this implicit theory was generally more stable or clearly understood for running backs than for offensive tackles regardless of knowledge level. In other words, since most of the behavior frequency ratings involved some degree of inference, the accuracy of these inferences may be dependent upon the implicit theories subjects have of each position. The effectiveness of drawing inferences from such implicit theories seems to be a function of both job knowledge and the particular position observed. Both factors seem to function independently.

Post hoc assessment of individual behavioral frequency items suggest that differential accuracy effects may reflect specific cognitive categorization strategies (Rosch, 1978). Without the knowledge base, or the pre-existing framework that less knowledgeable raters may have been lacking, it was difficult to differentiate between the two running backs within each dimension, especially in the impression formation condition. Categorization theory (Rosch, 1978)
suggests that what less knowledgable raters may have been lacking in this condition are efficient categories by which to organize the information they observed. In other words, individuals may have certain categories available to them by which they compare and classify ratees. In the context of the present study, relevant categories may be "fullback", "halfback", or even "good player" and "poor player" since once again evaluative true scores indicate that one running back clearly performed superior to the other. It may be that less knowledgable raters were not aware of the differentiation of responsibilities discussed above, and therefore did not possess the position categories. Without such categories less knowledgable subjects failed to notice the variation in behavior frequency that the specialization differences manifested unless they were directed to remember detail. Under the memory-set condition the correlation dropped to $r = -.08$ (See Table 11), suggesting that observation under the memory-set condition countered this disadvantage for less knowledgable raters.

Closer examination of some of the individual behavior frequency items explore the possible influence of categorization. The results suggest that the advantage that the knowledgable raters had may have reflected the two types of categorization mentioned. To determine the extent to which knowledge of running back specialization, or the
"fullback" and "halfback" category, may have influenced the differential accuracy measure, accuracy of responses for number of times blocking and carrying the ball for each running back was correlated with job knowledge. For three out of the four items greater knowledge did lead to more accuracy (i.e. negative correlation with the accuracy component). The correlations were -.67 which was significant, \( p < .01 \), and .05, -.17, -.12. Similar analyses were conducted for behavioral items that suggested some indication of performance quality (e.g. times missed a block, gained yardage, broke a tackle). These items were classified into four categories developed on the basis of two dimensions, player performance (good/poor) and type of behavior evaluation (positive/negative). If knowledgable raters were making more accurate behavior frequency ratings due to effectively making use of "good player" and "bad player" categories this analysis should yield strong negative correlations. This was the case for rating negative behavior frequencies for the good player, \( r = -.51, \ p < .01 \), and rating positive behavior frequencies for the poor player, \( r = -.65, \ p < .01 \). The other two categories, positive behaviors of good players and negative behaviors of poor players yielded weak, but positive correlations, .11 and .12 respectively. These are examples where the categorization process can cause bias. In both cases
knowledgable raters tended to overestimate the occurrence of these types of behaviors. Knowledge was weakly correlated with attributing more negative behaviors to the poor player, $r = .14$, $p > .05$, but strongly correlated with frequency ratings of positive behaviors for the good player, $r = .50$, $p < .01$.

The simple task / impression formation condition was perceived by subjects as significantly less complex than the other experimental conditions as determined by responses to the manipulation check; yet it demonstrated the strongest knowledge - accuracy relationship. Forming impressions is much easier than remembering detail, and viewing running backs may be more simplistic than viewing offensive tackles due to more highly diagnostic information and greater familiarity for the running back position. Subsequently this condition may have been much less demanding on subjects cognitive abilities at time of observation compared to the other conditions. However, this may have increased the challenge and difficulty at time of retrieval when responding to the behavior frequency items. In order to respond to the behavior frequency measure, subjects in this condition may have needed to rely on the pre-existing cognitive frameworks, particularly their use of categorization at time of encoding which was readily available to knowledgable raters only. It may have been
necessary for knowledgable raters to draw on this pre-existing framework to provide behavior frequency estimations, and this provided an advantage in the impression-set condition. It was expected that when asked to form an impression, knowledgable raters would lose specific behavioral information. This still may have occurred, but at time of retrieval of that information, when the behavior categories were provided to the subjects, the knowledgable raters were able to infer better approximations of the behavior frequencies relevant to each of the running backs. Post hoc analyses suggest that this increased accuracy may be due to the pre-existing knowledge base which provided an understanding of the task specialization and quality of performance differences between the two running backs and facilitated the inferential process. The knowledge-accuracy relationship may have been influenced more by this inferential process than actual recall of detailed behaviors. This would account for the unexpected positive relationship in this particular condition.

Apparently, the simple task led to greater accuracy in all aspects, except in the ability to accurately discriminate between the frequency of ratee behaviors within each dimension (differential accuracy). It is interesting that compared to raters in the complex task condition, raters in the simple task condition were able to better
discriminate between ratees and dimensions in general when evaluating running backs, but were unable to distinguish between the running backs within each particular behavior dimension. This is especially surprising when one considers the differences between the two pairs of players. The two running backs tend to have more specialization differences between them compared to offensive tackles who have similar responsibilities. One running back, the halfback, is primarily responsible for running and secondarily responsible for blocking; which is reverse of that for the other running back, the fullback. These differences are reflected in the behavior frequency true scores for running backs, and in a sense, within the behavior items themselves (e.g. carry the ball, blocking). The behavior items for the complex task condition seem to be more closely related (e.g. missed a block, pushed defender backward, knock defender down, etc.). In terms of true scores, the player who was a halfback "carried the ball" a greater amount of times, and "blocked" fewer times relative to the fullback. Such specialization should facilitate easier rank ordering and relative frequencies of such behaviors between the two running backs compared to offensive tackles.

Significant interactions occurred for the differential elevation measure associated with the behavioral frequency measures, but they did not support the effects predicted in
the hypotheses. Close examination of subjects' overall means for each target player produced some interesting speculation concerning the differential elevation component. The discussion below explains how the differential elevation effects may have been an artifact of differences in elevation bias among subjects observing each position. It is concluded that the distance formula for differential elevation may need to be more closely analyzed, especially in the context of significant elevation effects.

The interaction displayed in Figure 4 suggests that more knowledgable raters were less effective at ranking or discriminating between target players in terms of overall behavior frequencies within the complex task condition. Initially it was believed that the effect was due to categorization bias. Knowledgable raters overestimating good behaviors for the better player (based on true score evaluations), and underestimating good behaviors for the poorer player. This was not confirmed by the data for the relationship between job knowledge and either frequencies for good behaviors for the better player ($r = -0.12$), or frequencies for good behaviors for the poorer player ($r = 0.00$).

Nathan & Alexander (1965) suggest that elevation and differential elevation reflect rating styles or bias; while stereotypical and differential accuracy are more closely
related to inferential and cognitive processes of performance ratings. With this in mind, a closer examination of the data suggests that the job knowledge X task complexity interaction effect may be an artifact resulting from a general elevation bias by all raters for the better offensive tackle, and a strong knowledge - accuracy correlation for the poorer player. The distance formula for calculating differential elevation was used (See Appendix G) since the correlation component was not arithmetically viable due to the use of only two rateses. The general elevation bias refers to the tendency of the raters to overestimate behavior frequency of offensive tackles in general (M=3.73 compared to a true score M=2.76), and for each of the tackles individually (observed M = 3.26 and 4.19 for respective true score M = 3.00 and 2.55). It is important to note that the offensive tackle with fewer true relevant behavior frequencies received the higher observed frequency rating average. True behavior frequencies are provided in Appendix H. In fact, only one subject (of 63) underestimated the behavior frequency average for this player. In addition, only 12 subjects (19%) even had the direction of the two means correct.

The raters seemed to have been biased toward attributing more behaviors to the better player (based on overall true score ratings) even though this was in contrast
to reality as indicated by the true scores. More behaviors should not automatically lead to better overall performance for two reasons. First, many of the behavior dimensions reflect poor (e.g. times missed a block) or neutral (e.g. times blocked a linebacker) performance. Second, the overall rating should entail a qualitative evaluation of the player's behavior, and not just a quantitative assessment of the behaviors performed. Since this bias seemed to focus on the better player, it would seem that negative items (e.g. missed a block) should not be included in this elevation bias. This tended to be truer for more knowledgable subjects. In fact, knowledgable raters tended to be more conservative (r = -.28, p < .05) in providing frequencies for negative items for the better player, and for the poorer player as well (r = -.25, p < .05).

Upon examining the rater's accuracy for each player's overall behavior frequency average (elevation component applied to each individual ratee), the degree of elevation bias for the better offensive tackle was independent of level of job knowledge, r = -.02. The same relationship for the poorer offensive tackle indicated that knowledgable raters were significantly more accurate at determining the behavior frequency average, r = -.27, p < .05. Combining both the elevation and differential elevation results suggests that the two results may be related and provide an
understanding of differences in rating styles between knowledgable and less knowledgable subjects. The latter tended to consistently overestimate behavior frequencies for both players, while knowledgable raters were subject to this lenient tendency only for the better player, perhaps due to generalization of better performance to behavior involvement which was in contrast to reality as suggested by the true scores. This suggests that the frequency ratings for each player were obtained in an independent fashion. The two ratee performances were distinctly different in that one evoked a general elevation bias toward behavior frequency, while the other was more accurately portrayed by more knowledgable raters. In other words, at time of encoding, behavioral information was processed differently as a function of qualitative differences within each target performance. A type of halo effect occurred as good performance was inaccurately generalized to other behaviors. In addition, poor performance may have been perceived as being more salient behavioral information for knowledgable raters and subsequently enabled them to be more accurate.

Perhaps examining the pattern of means among ratees, especially within each experimental group, could reveal processing phenomena which may be unique to each target ratee. In the present study, the two offensive tackles may have been treated differently. Mount and Thompson (1987)
discovered that managers were rated more accurately and with greater halo and leniency when they fell into a subordinate's congruent category compared to incongruent category. Congruent/incongruent referred to the extent to which subordinates perceived each manager's behaviors represented the expectations of those subordinates. In the present study a similar phenomenon may have occurred for the good performing offensive tackle as behaviors were consistently overestimated, and often incorrectly elevated above the poor players involvement.

To summarize the effects found for behavioral frequency ratings, knowledgable raters seem to be more accurate at providing behavior frequencies for all measures, except differential elevation. This may be due to differences in the dimensions representing each position, and the ability of knowledgable raters to handle those differences. As suggested in the discussion of task complexity, the dimensions for the tackles may be more difficult to distinguish and this contributes to a general lack of clear and useful implicit theories (Nathan & Alexander, 1965) for offensive tackle performance. The dimensions may also be less salient making it difficult to isolate specific behaviors since the offensive tackles spend much more time "in the thick of things". This would cause the task complexity main effect that was hypothesized and supported
in three of the four accuracy measures pertaining to behavior frequency ratings.

The advantage of knowledgable raters seems to be that they are sufficiently familiar with the dimensions due to the pre-existing framework from which, hypothetically, they operate. This would seem especially true for dimensions that may involve other defensive positions (e.g. times blocked a linemen, times blocked a linebacker, etc.) since the various defensive positions have different probabilities with which they would interact with the target player. Behavior frequency rating accuracy is facilitated by the pre-existing categories (Rosch, 1978) used by knowledgable raters in that it provides a relative guide for them at time of observation to more efficiently store specific behavioral information, and perhaps to retrieve that information at time of recall. If knowledgable raters cannot remember specific frequencies their estimations can be more accurately guided by the "typical" ratio pertaining to the behaviors of offensive tackles or running backs provided by the pre-existing categories. Basically, the categories assist the knowledgable raters in differentiating between the dimensions, as well as differentiating the players within each dimension which, as indicated by the data, is much more difficult when evaluating the two running backs. Srull and Wyer (1988) suggest that categories are more
accessible as a function of the degree to which they are used. In the present context, it seems probable that job knowledge serves as a good indicator of the degree to which subjects have previously employed or experienced using relevant categories in football.

Implications

Although hypotheses were unsupported for the most part, implications for research can be derived from the study.

The first implication focuses on the task complexity variable. Probably the most salient implication from this study is the need to further the efforts of Campbell (1988) and Wood (1988) in order to understand and appropriately classify the various dimensions of the task complexity construct. This is especially critical in the context of performance appraisal where the complexity of the task being observed may influence, but also be quite distinct from the task of the observer-evaluator. The focus of the latter may be on the number and interrelatedness of the performance behaviors. The former deals primarily with handling and integrating the information cues in a particular context. For instance, research on competing tasks (Balzer, 1986) may make the rating task more complex regardless of the level of complexity involved in the actual job being observed. Although the two sources of complexity may be discussed
separately, understanding their relationship will be a critical step in furthering the social cognitive research in performance appraisal. In addition, the contrast between the results for the four accuracy measures for behavior frequency ratings suggest that each of the components are quite distinct, and the notion of independence between them seems to be supported (Murphy & Balzer, 1981). More importantly, the influence of job knowledge and task complexity seem to affect each component differently. Perhaps more systematic investigation of what each measure represents in terms of cognitive processing should be considered and explored in the future. Such attempts should emulate Murphy and Balzer's (1986) empirical demonstration of how halo error improves rating accuracy. It may be that within each research endeavor, separate hypotheses regarding each measure should be encouraged.

Secondly, the concern with the time lapse in the present study should not be misconstrued as evidence for denouncing processing differences since the data suggests that they disappear soon after observation. Stable cognitive differences among raters with different levels of knowledge or expertise has been well documented, whether it involve prototype attributes (Foti & Luch, 1982), category familiarity (Hauenstein & Kovach, 1986), or conceptual similarity schemas (Smither & Reilly, 1987). Also, it has
been demonstrated that time delays affect rating accuracy (Murphy & Balzer, 1986), and behavior recall (Nathan & Lord, 1983). Therefore, it would make sense to explore how variables may moderate the relationship between memory and accuracy. One specific factor to be explored is time, how cognitive processes are moderated by various time intervals spanning from the point of observation to longitudinal-types of exploration, especially since appraisals are often done on a semi-annual to annual basis. For example, drastic changes in recognition accuracy were observed by Murphy, Philbin, and Adams (1989) from immediate recall to one day delay to three days delay when students were evaluating teacher performance. The question may be raised that if changes in the process are apparent between one and three days, what type of effect does the more realistic delay of a month or more have on the process? Research should continue to focus on how the dynamics of different processing strategies change over time within different appraisal contexts as defined by the purpose of the appraisal, availability of behavioral information, and competition with other tasks besides appraisal, just to mention a few of the relevant contextual factors.

Third, despite the suggestion that the results tended to display idiosyncratic styles of information processing (behavior- or judgment-oriented) independent of job
knowledge, such individual difference variables should continue to be investigated. The job knowledge variable is just one of several other distinct relevant constructs pertaining to individual differences among raters, such as levels of expertise in actually performing the job being evaluated, and familiarity which may be determined by direct or indirect experience, as well as simple exposure as exemplified by the discussion concerning differences in sports media coverage for running backs and offensive tackles. The relationship between these constructs needs to be better understood, and so research should draw on the example of others (Foti & Luch, 1982; Kozlowski et al., 1987) and continue to look at such individual difference variables simultaneously, and with consideration of how they may influence the perceptions of task complexity discussed above.

Finally, a widely-held belief that supervisor ratings may be necessary, but are rarely suitably accurate is an insufficient generalization. The current trend in exploring social cognitive issues in the realm of performance appraisal (Landy & Farr, 1980; Feldman, 1981) will continue to provide further insight into understanding the information processing phenomena that leads to those ratings, and in turn better applications for ensuring more accurate and reliable ratings within the workplace. The
implementation of more cognitive measures (e.g. free recall, frequency estimates, etc.) is encouraged in order to get more information concerning raters' schemas and processing strategies. However, two improvement suggestions include the possible use of verbal free recall, to avoid possible contamination in transferring thoughts to words, as well as a more refined classification of the qualitative data gained from free recall tasks should be explored. Simply because the internal functions of the rater cannot be directly observed, should not prevent attempts of empirical investigation of that subject matter.

Limitations

This study has several limitations that may also account for the lack of support for the hypotheses tested. First, the interaction between variables for the complexity manipulation check suggests that the two levels of both independent variables manipulated may not have operated independently. Second, only one overall rating was obtained on a 100 point scale for the evaluative ratings. This limited the assessment of rating accuracy to only elevation and differential elevation. This was most unfortunate since stereotypical accuracy and differential accuracy yielded a significant relationship with job knowledge in the behavior frequency ratings suggesting that for the performances being studied (running backs and offensive tackles) the ability to
differentiate among dimensions, and differentiate rates within dimensions may be dependent on levels of job knowledge. Third, the use of students as raters may limit the generalizability of the results. The generalizability of lab studies in performance appraisal have long been debated (Banks & Murphy, 1985; Ilgen & Favero, 1985). Although the present study attempted to capture a typical, more realistic rating experience by using football, a performance domain which students should be sufficiently familiar, paralleling this phenomena to that of evaluation in the workplace by supervisors should be done with caution. The use of real performance stimulus and the degree of control implemented in the present study was were strong points necessary to scientifically test the hypotheses. Students evaluating football players may even meet the criteria of the constant familiarity assumption advocated by many researchers (Athey & McIntyre, 1987; Smither and Reilly, 1987) in the sense that football is an occupation to which students have been exposed. However, this does not guarantee that the information processing done by college students in a laboratory setting is equivalent to that done by actual supervisors in a "real world" setting. It may well be that the generalizability concern of such performance appraisal studies is a valid point. A related limitation is the laboratory setting which lacks many of the
distractions typical in organizations. It would be extremely advantageous if research of this nature could somehow be investigated in the field. Such innovative methodologies should be considered for future research endeavors. Finally, the operationalization of job knowledge may be considered a limitation in that the test used was made up of questions pertaining to the general game of football: rules, terms and collegiate trivia for the most part. The construct validity of this measure may be criticized for not truly tapping knowledge levels for each of the actual jobs studied. In other words, the instrument did not directly measure relevant knowledge for being a running back or an offensive tackle per se, but rather a more general construct.

It is highly encouraged that similar research conducted in this area in the future, particularly those that explore the same variables, take these limitations into consideration, and correct for them as effectively as possible.
References


Associative storage and retrieval processes in person 

processing invariance in laboratory studies of performance 
Polytechnic Institute & State University, Blacksburg, VA.

attention, impressions, & recall complex social interactions 

Williams, K. J., DeNisi, A. S., Meglino, B. M., & Cafferty, T. P. 
(1985). The role of appraisal purpose: Effects of purpose on 
information acquisition and utilization. Organizational 
Behavior and Human Decision Processes, 36, 314-338.

Organizational Behavior and Human Decision Processes, 37, 
60-82.

cognitive representation of persons and groups and its effect 
on recall and recognition memory. Journal of Experimental 

Wyer, R. S., & Gordon, S. E. (1982). The recall of information 
about persons and groups. Journal of Experimental Social 
Psychology, 18, 128-154.

as a function of rater training and purpose of the appraisal. 
Table 1

Means and Standard Deviations for Manipulation Checks of Task Complexity and Observational Purpose by Task Complexity Condition and Observational Purpose Condition

<table>
<thead>
<tr>
<th>Memory-set</th>
<th>Observational Purpose</th>
<th>Impression-set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple (n=30)</td>
<td>Complex (n=32)</td>
</tr>
<tr>
<td>DV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q#1</td>
<td>3.60 (1.43)</td>
<td>3.84 (1.43)</td>
</tr>
<tr>
<td>Q#2</td>
<td>3.80 (1.27)</td>
<td>4.06 (1.37)</td>
</tr>
<tr>
<td>Q#3</td>
<td>4.07 (1.46)</td>
<td>4.09 (1.42)</td>
</tr>
<tr>
<td>Q#4</td>
<td>4.10 (1.86)</td>
<td>3.75 (1.44)</td>
</tr>
<tr>
<td>Q#5</td>
<td>4.60 (1.40)</td>
<td>4.22 (1.39)</td>
</tr>
<tr>
<td>Q#6</td>
<td>4.23 (1.50)</td>
<td>3.94 (1.37)</td>
</tr>
<tr>
<td>JK</td>
<td>32.04 (8.38)</td>
<td>33.72 (8.59)</td>
</tr>
</tbody>
</table>

Note: Q#1=degree of difficulty critically observing targets
Q#2=degree of difficulty evaluating targets’ performance
Q#3=degree of difficulty in forming overall impression
Q#4=extent to which the observational purpose was clear
Q#5=extent to which subjects tried to form an impression
Q#6=extent to which subjects tried to memorize details
JK =results of job knowledge (Football) test

Note: Number in parentheses denotes standard deviation.
Table 2

Hierarchical Regression Analyses Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Perceptions of Complexity Manipulation Check Items.

<table>
<thead>
<tr>
<th>Source</th>
<th>Complexity Perceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>JK</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.09*</td>
</tr>
<tr>
<td>$F$</td>
<td>11.32</td>
</tr>
<tr>
<td>TC</td>
<td></td>
</tr>
<tr>
<td>$\setminus R^2$</td>
<td>.06*</td>
</tr>
<tr>
<td>$F$</td>
<td>8.68</td>
</tr>
<tr>
<td>OP</td>
<td></td>
</tr>
<tr>
<td>$\setminus R^2$</td>
<td>.05*</td>
</tr>
<tr>
<td>$F$</td>
<td>7.42</td>
</tr>
<tr>
<td>JK X TC</td>
<td></td>
</tr>
<tr>
<td>$\setminus R^2$</td>
<td>.01</td>
</tr>
<tr>
<td>$F$</td>
<td>1.08</td>
</tr>
<tr>
<td>JK X OP</td>
<td></td>
</tr>
<tr>
<td>$\setminus R^2$</td>
<td>.01</td>
</tr>
<tr>
<td>$F$</td>
<td>1.21</td>
</tr>
<tr>
<td>TC X OP</td>
<td></td>
</tr>
<tr>
<td>$\setminus R^2$</td>
<td>.02</td>
</tr>
<tr>
<td>$F$</td>
<td>3.64</td>
</tr>
<tr>
<td>TC X TC X OP</td>
<td></td>
</tr>
<tr>
<td>$\setminus R^2$</td>
<td>.00</td>
</tr>
<tr>
<td>$F$</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: JK: Job Knowledge
TC: Task Complexity
OP: Observational Purpose

Note: * p<.05
### Table 3

**Hierarchical Regression Analyses Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Observational Purpose Manipulation Checks**

<table>
<thead>
<tr>
<th>Source</th>
<th>MC4</th>
<th>MC5</th>
<th>MC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>JK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>F</td>
<td>1.02</td>
<td>1.04</td>
<td>0.00</td>
</tr>
<tr>
<td>TC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.00</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>F</td>
<td>0.00</td>
<td>2.58</td>
<td>1.11</td>
</tr>
<tr>
<td>OP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.01</td>
<td>.11</td>
<td>.01</td>
</tr>
<tr>
<td>F</td>
<td>1.82</td>
<td>8.92**</td>
<td>1.50</td>
</tr>
<tr>
<td>JK X TC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>F</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>JK X OP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>F</td>
<td>1.22</td>
<td>0.00</td>
<td>1.08</td>
</tr>
<tr>
<td>TC X OP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>F</td>
<td>0.73</td>
<td>1.27</td>
<td>2.68</td>
</tr>
<tr>
<td>JC X TC X OP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>F</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Note:**
- MC4 = extent to which the observational purpose was clear
- MC5 = extent to which subjects tried to form an impression
- MC6 = extent to which subjects tried to memorize details

**Note:**
- **p < .01**
Table 4

Means and Standard Deviations for Free Recall Measures by Task Complexity Condition and Observational Purpose Condition

<table>
<thead>
<tr>
<th>Observational Purpose</th>
<th>Memory-set</th>
<th>Impression-set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td></td>
<td>(n=30)</td>
<td>(n=32)</td>
</tr>
<tr>
<td>DV</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Beh.</td>
<td>8.10</td>
<td>7.91</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(4.51)</td>
</tr>
<tr>
<td>Jdg.</td>
<td>2.30</td>
<td>2.81</td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td>(2.53)</td>
</tr>
<tr>
<td>Tot.</td>
<td>10.49</td>
<td>10.72</td>
</tr>
<tr>
<td></td>
<td>(2.66)</td>
<td>(4.36)</td>
</tr>
</tbody>
</table>

Note: Beh.= number of behaviors recalled
Jdg.= number of judgments recalled
Tot.= total number of items recalled

Note: Number in parentheses denotes standard deviation.
Table 5

Hierarchical Regression Analyses Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Free Recall Measures

<table>
<thead>
<tr>
<th>Source</th>
<th>DV</th>
<th>Judgments</th>
<th>Behaviors</th>
<th>Total Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>JK</td>
<td>$R^2$</td>
<td>.01</td>
<td>.02</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>1.52</td>
<td>2.18</td>
<td>10.25**</td>
</tr>
<tr>
<td>TC</td>
<td>$R^2$</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>0.05</td>
<td>1.49</td>
<td>1.49</td>
</tr>
<tr>
<td>OP</td>
<td>$R^2$</td>
<td>.22</td>
<td>.17</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>34.92**</td>
<td>25.99**</td>
<td>0.00</td>
</tr>
<tr>
<td>JK X TC</td>
<td>$R^2$</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>JK X OP</td>
<td>$R^2$</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TC X OP</td>
<td>$R^2$</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>0.73</td>
<td>1.27</td>
<td>2.69</td>
</tr>
<tr>
<td>JC X TC X OP</td>
<td>$R^2$</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>$E$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: JK: Job Knowledge  
TC: Task Complexity  
OP: Observational Purpose

Note: ** p<.01
Table 9
Correlations between Job Knowledge and the Free Recall Measures Overall, and within each Task Condition

<table>
<thead>
<tr>
<th></th>
<th>JOBK</th>
<th>BEH</th>
<th>JUDG</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEH</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JUDG</td>
<td>.13</td>
<td>-.62**</td>
<td></td>
</tr>
<tr>
<td>TOTR</td>
<td>.28**</td>
<td>.31**</td>
<td>.56**</td>
</tr>
</tbody>
</table>

SIMPLE TASK (top triangle - above diagonal)
COMPLEX TASK (bottom triangle - below diagonal)

<table>
<thead>
<tr>
<th>JOBK</th>
<th>BEH</th>
<th>JUDG</th>
<th>TOTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOBK</td>
<td>1.00</td>
<td>.16</td>
<td>.12</td>
</tr>
<tr>
<td>BEH</td>
<td>.15</td>
<td>1.00</td>
<td>-.70**</td>
</tr>
<tr>
<td>JUDG</td>
<td>.07</td>
<td>-.55**</td>
<td>1.00</td>
</tr>
<tr>
<td>TOTR</td>
<td>.23</td>
<td>.28*</td>
<td>.67**</td>
</tr>
</tbody>
</table>

Note: JOBK = Job Knowledge
      BEH = Number of Behaviors Recalled
      JUDG = Number of Judgments Recalled
      TOTR = Total Number of Items Recalled

Note: * p < .05
     ** p < .01
Table 7

Means and Standard Deviations for Accuracy Measures by Task Complexity Condition and Observational Purpose Condition

<table>
<thead>
<tr>
<th>Memory-set</th>
<th>Observational Purpose</th>
<th>Impression-set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple (n=30)</td>
<td>Complex (n=32)</td>
</tr>
<tr>
<td></td>
<td>Simple (n=30)</td>
<td>Complex (n=31)</td>
</tr>
<tr>
<td>EVL</td>
<td>Evaluative Ratings</td>
<td></td>
</tr>
<tr>
<td>ELR 6.60</td>
<td>6.88</td>
<td>9.03</td>
</tr>
<tr>
<td>(4.74)</td>
<td>(4.88)</td>
<td>(4.31)</td>
</tr>
<tr>
<td>DAR 9.53</td>
<td>10.78</td>
<td>8.87</td>
</tr>
<tr>
<td>(4.90)</td>
<td>(8.26)</td>
<td>(4.94)</td>
</tr>
<tr>
<td>BEV</td>
<td>Behavior Frequency</td>
<td></td>
</tr>
<tr>
<td>ELF 0.66</td>
<td>1.06</td>
<td>0.69</td>
</tr>
<tr>
<td>(0.55)</td>
<td>(0.65)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>DE 0.31</td>
<td>0.75</td>
<td>0.51</td>
</tr>
<tr>
<td>(0.27)</td>
<td>(0.37)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>SA 1.32</td>
<td>2.15</td>
<td>1.34</td>
</tr>
<tr>
<td>(0.67)</td>
<td>(0.86)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>DA 2.23</td>
<td>1.60</td>
<td>2.20</td>
</tr>
<tr>
<td>(0.46)</td>
<td>(0.57)</td>
<td>(0.54)</td>
</tr>
</tbody>
</table>

Note: ELR = Elevation for Evaluative Ratings
DAR = Differential Accuracy for Evaluative Ratings
ELF = Elevation for Behavior Frequency Ratings
DE = Differential Elevation for Behavior Freq. Ratings
SA = Stereotypical Accuracy for Behavior Freq. Ratings
DA = Differential Accuracy for Behavior Freq. Ratings

Note: Number in parentheses denotes standard deviation.
Table 8
Hierarchical Regression Analyses Testing the Relationship
Between Job Knowledge, Task Complexity, and Observational
Purpose on Accuracy Measures for Evaluative Ratings

<table>
<thead>
<tr>
<th>Source</th>
<th>Elevation</th>
<th>Differential Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>JK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>E</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>E</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>E</td>
<td>2.78</td>
<td>4.63*</td>
</tr>
<tr>
<td>JK X TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>E</td>
<td>3.78</td>
<td>0.00</td>
</tr>
<tr>
<td>JK X OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>E</td>
<td>1.23</td>
<td>0.00</td>
</tr>
<tr>
<td>TC X OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>E</td>
<td>0.87</td>
<td>2.08</td>
</tr>
<tr>
<td>JC X TC X OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>E</td>
<td>0.62</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Note: JK: Job Knowledge
TC: Task Complexity
OP: Observational Purpose

Note: * P < .05
Table 8

Hierarchical Regression Analyses Testing the Relationship
Between Job Knowledge, Task Complexity, and Observational Purpose on Accuracy Measures for Behavior Frequency Ratings
(Elevation and Differential Elevation)

<table>
<thead>
<tr>
<th>Source</th>
<th>Elevation R²</th>
<th>Elevation</th>
<th>Differential R²</th>
<th>Differential</th>
</tr>
</thead>
<tbody>
<tr>
<td>JK</td>
<td>.01</td>
<td>1.42</td>
<td>.07</td>
<td>9.01**</td>
</tr>
<tr>
<td>TC</td>
<td>.09</td>
<td>12.51**</td>
<td>.14</td>
<td>21.81**</td>
</tr>
<tr>
<td>OP</td>
<td>.00</td>
<td>0.00</td>
<td>.01</td>
<td>0.85</td>
</tr>
<tr>
<td>JK X TC</td>
<td>.02</td>
<td>3.31</td>
<td>.03</td>
<td>4.33*</td>
</tr>
<tr>
<td>JK X OP</td>
<td>.02</td>
<td>2.75</td>
<td>.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TC X OP</td>
<td>.00</td>
<td>0.00</td>
<td>.03</td>
<td>4.37**</td>
</tr>
<tr>
<td>JC X TC X OP</td>
<td>.05</td>
<td>7.61**</td>
<td>.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: JK: Job Knowledge
TC: Task Complexity
OP: Observational Purpose

Note: * p<.05
** p<.01
Table 10

Hierarchical Regression Analyses Testing the Relationship Between Job Knowledge, Task Complexity, and Observational Purpose on Accuracy Measures for Behavior Frequency Ratings (Stereotypical Accuracy and Differential Accuracy)

<table>
<thead>
<tr>
<th>Source</th>
<th>Stereotypical Accuracy</th>
<th>Differential Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>JK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>$F$</td>
<td>4.73*</td>
<td>6.58*</td>
</tr>
<tr>
<td>TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\backslash R^2$</td>
<td>0.28</td>
<td>0.32</td>
</tr>
<tr>
<td>$F$</td>
<td>49.09**</td>
<td>47.15**</td>
</tr>
<tr>
<td>OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\backslash R^2$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$F$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>JK X TC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\backslash R^2$</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>$F$</td>
<td>0.00</td>
<td>3.93</td>
</tr>
<tr>
<td>JK X OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\backslash R^2$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$F$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TC X OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\backslash R^2$</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>$F$</td>
<td>0.00</td>
<td>2.44</td>
</tr>
<tr>
<td>JC X TC X OP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\backslash R^2$</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>$F$</td>
<td>0.00</td>
<td>7.49**</td>
</tr>
</tbody>
</table>

**Note:** JK: Job Knowledge
TC: Task Complexity
OP: Observational Purpose

**Note:** * $p<.05$
** $p<.01$
Table 11
Correlations Between Job Knowledge and Accuracy Measures
Overall, and Within Each Experimental Condition

<table>
<thead>
<tr>
<th></th>
<th>ER</th>
<th>DAR</th>
<th>EF</th>
<th>DEF</th>
<th>SA</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOBK</td>
<td>-.06</td>
<td>.01</td>
<td>-.11</td>
<td>.26*</td>
<td>-.19*</td>
<td>-.22*</td>
</tr>
</tbody>
</table>

Complex Task / Memory-Set

<table>
<thead>
<tr>
<th></th>
<th>ER</th>
<th>DAR</th>
<th>EF</th>
<th>DEF</th>
<th>SA</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOBK</td>
<td>-.08</td>
<td>.07</td>
<td>-.54*</td>
<td>.35</td>
<td>-.17</td>
<td>-.17</td>
</tr>
</tbody>
</table>

Complex Task / Impression-Set

<table>
<thead>
<tr>
<th></th>
<th>ER</th>
<th>DAR</th>
<th>EF</th>
<th>DEF</th>
<th>SA</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOBK</td>
<td>-.28</td>
<td>.03</td>
<td>.07</td>
<td>.50*</td>
<td>-.33</td>
<td>-.07</td>
</tr>
</tbody>
</table>

Simple Task / Memory-Set

<table>
<thead>
<tr>
<th></th>
<th>ER</th>
<th>DAR</th>
<th>EF</th>
<th>DEF</th>
<th>SA</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOBK</td>
<td>.12</td>
<td>.10</td>
<td>.18</td>
<td>.16</td>
<td>-.18</td>
<td>-.08</td>
</tr>
</tbody>
</table>

Simple Task / Impression-Set

<table>
<thead>
<tr>
<th></th>
<th>ER</th>
<th>DAR</th>
<th>EF</th>
<th>DEF</th>
<th>SA</th>
<th>DA</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOBK</td>
<td>.19</td>
<td>-.28</td>
<td>-.08</td>
<td>.01</td>
<td>-.24</td>
<td>-.75*</td>
</tr>
</tbody>
</table>

Note:

JOBK = Job Knowledge
ER = Elevation for Evaluative Ratings
DAR = Differential Accuracy for Evaluative Ratings
EF = Elevation for Behavior Frequency Ratings
DEF = Differential Elevation for Behavior Frequency Ratings
SA = Stereotypical Accuracy for Behavior Frequency Ratings
DA = Differential Accuracy for Behavior Frequency Ratings

Note: ** p < .01

Note: Negative correlations indicate increased accuracy with increased job knowledge.
Table 12

Hierarchical Regression Analyses Testing the Relationship
Between Job Knowledge and Complexity Perceptions on Process and
Accuracy Measures.

<table>
<thead>
<tr>
<th>Source</th>
<th>Job Knowledge (JK)</th>
<th>Complexity (C)</th>
<th>JK X C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>$F$</td>
<td>$\Delta R^2$</td>
</tr>
<tr>
<td><strong>PROCESS MEASURES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviors</td>
<td>.02</td>
<td>2.18</td>
<td>.00</td>
</tr>
<tr>
<td>Judgments</td>
<td>.01</td>
<td>1.52</td>
<td>.13</td>
</tr>
<tr>
<td>Total Recall</td>
<td>.08</td>
<td>10.25**</td>
<td>.13</td>
</tr>
<tr>
<td><strong>OVERALL RATINGS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>.00</td>
<td>0.00</td>
<td>.00</td>
</tr>
<tr>
<td>Differential Accuracy</td>
<td>.00</td>
<td>0.00</td>
<td>.05</td>
</tr>
<tr>
<td><strong>BEHAVIOR FREQUENCIES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>.01</td>
<td>1.42</td>
<td>.00</td>
</tr>
<tr>
<td>Differential Elevation</td>
<td>.07</td>
<td>9.17**</td>
<td>.01</td>
</tr>
<tr>
<td>Stereotypical Accuracy</td>
<td>.04</td>
<td>4.73*</td>
<td>.03</td>
</tr>
<tr>
<td>Differential Accuracy</td>
<td>.05</td>
<td>6.04*</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note: * $p < .05$
** $p < .01$
Figure 1: The moderating effect of observational purpose on the relationship between job knowledge and elevation of behavior frequency ratings for simple and complex task conditions.
FIGURE 2: TASK COMPLEXITY X OBSERVATIONAL PURPOSE INTERACTION FOR DIFFERENTIAL ELEVATION MEASURE OF BEHAVIOR FREQUENCY RATINGS
FIGURE 3: THE MODERATING EFFECT OF TASK COMPLEXITY ON THE RELATIONSHIP BETWEEN JOB KNOWLEDGE AND DIFFERENTIAL ELEVATION FOR BEHAVIOR FREQUENCY RATINGS
FIGURE 4: THE MODERATING EFFECT OF OBSERVATIONAL PURPOSE ON THE RELATIONSHIP BETWEEN JOB KNOWLEDGE AND DIFFERENTIAL ACCURACY OF BEHAVIOR FREQUENCY RATINGS FOR SIMPLE AND COMPLEX TASK CONDITIONS
FIGURE 5: THE MODERATING EFFECT OF PERCEPTIONS OF COMPLEXITY ON THE RELATIONSHIP BETWEEN JOB KNOWLEDGE AND BEHAVIORS RECALLED
Appendix A

Instructions to Subjects
This is the football study. We are interested in how people view the actions of individual football players. First, please read the consent form, and if you agree to participate, sign and date the form. Thank you.

For this experiment you will be asked to carefully view a series of football plays and then answer some questionnaires. There will be 14 plays in all, and you will be able to see each play 3 times, once forward, once backward, and once again forwards. Thus, it is very important that you pay careful attention to EACH PLAY and follow your instructions accordingly.

Specifically, we want you to strictly focus on the actions of TWO INDIVIDUAL PLAYERS in the MAROON uniforms. These players are two running backs. Your purpose will be to recall as many of the actions of the two players as possible.

The running backs ALWAYS stand behind the quarterback (the quarterback is the player who receives the ball from the center.) The running backs, however, may stand one behind the other, or side-by-side. In both cases, they are positioned BEHIND the quarterback. Below are two diagrams indicating where the running backs stand. The "O's" are the defensive players and the "X's" are the offensive players. The two X's in boxes are the running backs.

```
  0  0  0  0  0
0  0  0  0  0
X  X  X  X  X  X
   X
   X
```

There are two separate running backs for you to watch. One player is number 43 and the other number 27. However, since these players occasionally switch places, the experimenter will tell you where each individual player is standing (i.e., which one is on the right or left, or front or back) BEFORE EVERY PLAY. Thus, please pay very careful attention to both the experimenter and the film throughout this experiment. Are there any questions?
This is the football study. We are interested in how people view the actions of individual football players. First, please read the consent form, and if you agree to participate, sign and date the form. Thank you.

For this experiment you will be asked to carefully view a series of football plays and then answer some questionnaires. There will be 14 plays in all, and you will be able to see each play 3 times, once forward, once backward, and once again forwards. Thus, it is very important that you pay careful attention to EACH PLAY and follow your instructions accordingly.

Specifically, we want you to strictly focus on the actions of TWO INDIVIDUAL PLAYERS in the MAROON uniforms. These players are two running backs. Your purpose will be to form an impression of the two players.

The running backs ALWAYS stand behind the quarterback (the quarterback is the player who receives the ball from the center.) The running backs, however, may stand one behind the other, or side-by-side. In both cases, they are positioned BEHIND the quarterback. Below are two diagrams indicating where the running backs stand. The "O's" are the defensive players and the "X's" are the offensive players. The two X's in boxes are the running backs.

```
  0  0  0  0  0
  0  0  0  0  0
    0  0  0  0
  X  X  X  X  X  X
     X  X
        X
```

There are two separate running backs for you to watch. One player is number 43 and the other number 27. However, since these players occasionally switch places, the experimenter will tell you where each individual player is standing (i.e., which one is on the right or left, or front or back) BEFORE EVERY PLAY. Thus, please pay very careful attention to both the experimenter and the film throughout this experiment.

Are there any questions?
This is the football study. We are interested in how people view the actions of individual football players. First, please read the consent form, and if you agree to participate, sign and date the form. Thank you.

For this experiment you will be asked to carefully view a series of football plays and then answer some questionnaires. There will be 14 plays in all, and you will be able to see each play 3 times, once forward, once backward, and once again forwards. Thus, it is very important that you pay careful attention to EACH PLAY and follow your instructions accordingly.

Specifically, we want you to strictly focus on the actions of TWO INDIVIDUAL PLAYERS in the MAROON uniforms. These players are two offensive tackles. Your purpose will be to form an impression of the two players.

The offensive tackle ALWAYS stand two players from the right and left of the center, respectively (The center is the player who snaps the ball.) Below is a diagram indicating where the offensive tackles stand. The "O's" are defensive players and the "X's" are offensive players. The two X's in boxes are the offensive tackles.

```
   O   O
   0   0   0
   0   0   0   0
   X   X   X   X   X
   X   X
```

There are two separate offensive tackles for you to watch. One player is number 70 and the other number 68. However, since these players occasionally switch places, the experimenter will tell you where each individual player is standing (i.e., which one is on the right or left, or front or back) BEFORE EVERY PLAY. Thus, please pay very careful attention to both the experimenter and the film throughout this experiment. Are there any questions?
This is the football study. We are interested in how people view the actions of individual football players. First, please read the consent form, and if you agree to participate, sign and date the form. Thank you.

For this experiment you will be asked to carefully view a series of football plays and then answer some questionnaires. There will be 14 plays in all, and you will be able to see each play 3 times, once forward, once backward, and once again forwards. Thus, it is very important that you pay careful attention to EACH PLAY and follow your instructions accordingly.

Specifically, we want you to strictly focus on the actions of TWO INDIVIDUAL PLAYERS in the MAROON uniforms. These players are two offensive tackles. Your purpose will be to form an impression of the two players.

The offensive tackle ALWAYS stand two players from the right and left of the center, respectively (the center is the player who snaps the ball.) Below is a diagram indicating where the offensive tackles stand. The "O's" are defensive players and the "X's" are offensive players. The two X's in boxes are the offensive tackles.

```
  0 0 0 0 0
  0 0 0 0 0
 0 0 0 0 0
X  X X X X
 X
 X
```

There are two separate offensive tackles for you to watch. One player is number 70 and the other number 68. However, since these players occasionaly switch places, the experimenter will tell you where each individual player is standing (i.e., which one is on the right or left, or front or back) BEFORE EVERY PLAY. Thus, please pay very careful attention to both the experimenter and the film throughout this experiment.

Are there any questions?
Appendix B

Job Knowledge Test
Instructions

DO NOT TURN THE PAGE UNTIL YOU ARE INSTRUCTED TO DO SO.

You will need a NUMBER 2 PENCIL to complete this test. If you do not have one, please ask the experimenter for a pencil.

Put your Social Security Number on your answer sheet.

This test contains 45 multiple-choice items covering various rules and facts of collegiate football. Please answer each question to the best of your knowledge. If you are not sure that your answer is correct, skip the question and go on to the next item. Please keep in mind that the test is timed. You will have 30 minutes to work.

Each question has four (4) possible answers: a, b, c, or d. On your OPSCAN sheet, darken the circle corresponding to the letter of the correct answer. Please make sure you COMPLETELY fill in the circle corresponding to the correct answer.

EXAMPLE: What is the nickname of the Virginia Tech football team?

a. Wolverines  b. Crimson Tide
   c. Hokies      d. Wildcats

The correct answer is "c". You would darken the circle of the letter "c" on your OPSCAN sheet.

Your score on this test will be the total number of questions you answered correctly in both sections, minus a penalty for incorrect answers. If you are sure you know the answer to a question, circle the appropriate letter on the answer sheet. It will be to your advantage to answer questions in which you can eliminate one or more options. Random guessing, however, will not pay off because of the penalty for wrong answers.
1. How many downs does a team get after gaining possession of the ball?
   a. one  
   b. two  
   c. three  
   d. four

2. For what team does Warren Moon play?
   a. Minnesota Vikings  
   b. Detroit Lions  
   c. Houston Oilers  
   d. New England Patriots

3. How many yards is a team penalized for "offensive holding"?
   a. 5  
   b. 10  
   c. 15  
   d. 20

4. What conference do the Ohio State Buckeyes belong to?
   a. Big Ten  
   b. Atlantic Coast (ACC)  
   c. Southeastern (SEC)  
   d. Independent

5. Who is the present coach of the Dallas Cowboys?
   a. Be Schembeckler  
   b. Jimmy Johnson  
   c. Tom Landry  
   d. Vince Dooley

6. How many points are awarded for a "safety"?
   a. 1  
   b. 2  
   c. 3  
   d. 6

7. Which of the following team is not in the NFC?
   a. Dallas Cowboys  
   b. Denver Broncos  
   c. Minnesota Vikings  
   d. Chicago Bears

8. What college did Eric Dickerson play for?
   a. University of Texas  
   b. Arizona State University  
   c. Southern Methodist University  
   d. University of North Carolina

9. Which conference is represented in the Rose Bowl each year?
   a. Big Ten  
   b. Atlantic Coast (ACC)  
   c. Southeastern (SEC)  
   d. Western Athletic (WAC)

10. What offensive player typically blocks AND catches passes?
    a. Quarterback  
    b. Tackle  
    c. Center  
    d. Tight End

11. What college did John Elway play for?
    a. Stanford University  
    b. Michigan State University  
    c. Iowa University  
    d. University of Miami

12. How many minutes are played per quarter?
    a. 10  
    b. 15  
    c. 20  
    d. 25
132

13. For one half, how many timeouts does each team receive?
   a. 1  b. 2  c. 3  d. 4

14. What team did Vince Lombardi coach?

15. Where is the ball spotted after a touchback?
   a. On the 1-yard line  b. On the 10-yard line  c. On the 20-yard line  d. On the 30-yard line

16. Which of the following is a defensive play?
   a. A blitz  b. A double-reverse  c. A roll out  d. delayed draw

17. What is the maximum number of points that can be scored directly after a touchdown in college football?
   a. 0  b. 1  c. 2  d. 3

18. In football, what does a "screen" mean?

19. Which of the following is an example of an offensive running play?
   a. A flare  b. A trap  c. A flea-flicker  d. A zone

20. What position did Charles White play?
   a. Quarterback  b. Tight-end  c. Halfback  d. Wide receiver

21. Which of the following is a pass play?
   a. A reverse  b. A button hook  c. An end-around  d. A quarterback draw

22. What school did Woody Hayes coach for?

23. What is meant by the "Wishbone"?
   a. It refers to a blitz  b. It refers to a field goal  c. It refers to a long pass  d. It refers to an offensive backfield

24. What is the EXACT rule for determining a touchdown?
   a. Whether the player crosses the end-line  b. Whether the ball is completely over the end-line  c. Whether the player’s knee is over the end-line  d. Whether the ball breaks the plane of the end-line
25. What is the penalty for offensive "clipping"?
   a. 5 yard loss  
   b. 10 yard loss  
   c. 15 yard loss  
   d. 15 yards & a loss of down

26. What is meant by the term "nickel defense"?
   a. the defense is using five men in the secondary
   b. the defense is willing to give up the long pass
   c. the defense is geared towards the run
   d. the defense is preparing to make a goal-line stand

27. Which of the following most closely resembles a "swing pass"?
   a. a bomb  
   b. a post pattern  
   c. a fly pattern  
   d. a flare pass

28. Which of the following is NOT an offensive position?
   a. cornerback  
   b. halfback  
   c. fullback  
   d. tailback

29. What college did Bo Jackson play for?
   a. University of Alabama  
   b. Ohio State University  
   c. University of Georgia  
   d. Auburn University

30. What team does Joe Paterno coach for?
   a. Temple University  
   b. Notre Dame University  
   c. University of Florida  
   d. Penn State University

31. Which of the following is NOT a post-season collegiate all-star game?
   a. The Blue-Grey Classic  
   b. The Pro Bowl  
   c. The Hula Bowl  
   d. The East-West Shrine Game

32. Who gets possession of the ball when it is fumbled out-of-bounds?
   a. The team that snapped the ball  
   b. The team whose player last touched the ball  
   c. The team whose player last controlled the ball  
   d. The defensive team when the ball was snapped

33. What is the ruling when a defensive player knocks the ball out of the quarterback's hand before the quarterback's arm begins moving forward?
   a. Incomplete Pass  
   b. Defensive interference  
   c. Illegal procedure  
   d. Fumble

34. What is another name for a wide receiver?
   a. Halfback  
   b. Tight-end  
   c. Flanker  
   d. Safety

35. How many points are awarded for a field goal?
   a. one  
   b. two  
   c. three  
   d. six
36. Where does the "noseguard" position himself on the line of scrimmage?
   a. In the middle of the defensive line
   b. At the end of the offensive line
   c. In the middle of the defensive secondary
   d. In the offensive backfield

37. What is a "sweep"?
   a. A zone defense
   b. A passing play
   c. A blitz
   d. A running play

38. What is the minimum number of players the offensive team must put on the line of scrimmage?
   a. 4
   b. 5
   c. 6
   d. 7

39. When a punt returner waves his arm, he is signalling for:
   a. time out
   b. new play
   c. fair catch
   d. re-punt

40. What position does a "blocking back" play?
   a. Cornerback
   b. Quarterback
   c. Nickel back
   d. Running back

41. NOT including the quarterback, what is the maximum number of players a team is allowed to line up in the offensive backfield?
   a. 1
   b. 2
   c. 3
   d. 4

42. On most plays, who does the left offensive tackle block?
   a. Right defensive end
   b. Left defensive end
   d. Left defensive tackle

43. Which of the following is an example of a "special teams" play?
   a. A field goal
   b. A double-reverse
   c. A flea-flicker
   d. A blitz

44. Which player is usually responsible for covering a running back on a pass play?
   a. Safety
   b. Defensive End
   d. Linebacker

45. Which player carries the ball on "draw play"?
   a. Cornerback
   b. Wide receiver
   c. Tight end
   d. Running back
Please answer the following questions. Again, darken the corresponding circles on your answer sheet.

46. How much experience do you have playing organized, full-contact football?
   a. none            b. 1-2 years
   c. 3-4 years       d. 5 or more years

47. At all levels of football, what is the frequency with which you attend games?
   a. none            b. 1-2 times per year
   c. 3-4 times per year d. 5-6 times per year
   e. 7 or more times per year

48. Approximately how many hours per week do you spend watching football (collegiate or professional) on TV?
   a. none            b. 1-3 hours per week
   c. 4-6 hours per week d. 7-8 hours per week
   e. more than 8 hours per week

49. How much do you follow football in the newspaper?
   a. not at all       b. a little
   c. a moderate amount d. a good deal
   e. very much

50. How much do you like football?
   a. not at all       b. a little
   c. a moderate amount d. a good deal
   e. very much

51. How many football teams (collegiate or professional) do you actively follow?
   a. none            b. 1-2
   c. 3-4             d. 5-6
   e. more than 6
Appendix C

Sample of Free Recall Measure
Social Security #

Focusing on player number 27, please list everything you can remember from the film. Please number each of your comments.
Social Security # ____________

Focusing on player number 68, please list everything you can remember from the film. Please number each of your comments.
Appendix D

Evaluative Rating Measures
Please read the following questions carefully, and indicate your response in the appropriate space. Make sure your answers are whole numbers. Mark all of your answers directly on the questionnaire, and make sure you only choose ONE answer to each question.

1. Given that a passing grade is "90" for running backs, on a scale from 1-100, rate player No. 43 on his overall performance:

2. Given that a passing grade is "90" for offensive tackles, on a scale from 1-100, rate player No. 27 on his overall performance:
Please read the following questions carefully, and indicate your response in the appropriate space. Make sure your answers are whole numbers. Mark all of your answers directly on the questionnaire.

1. Given that a passing grade is "70" for offensive tackles, on a scale from 1-100, rate player No. 70 on his overall performance:

   ———

2. Given that a passing grade is "70" for offensive tackles, on a scale from 1-100, rate player No. 68 on his overall performance:

   ———
Appendix E

Behavior Frequency Rating Measures
Please read the following questions carefully, and indicate your response in the appropriate space. Make sure your answers are whole numbers. Mark all of your answers directly on the questionnaire, and make sure you only choose ONE answer to each question.

1. How many times, out of the 14 plays you saw, was No. 43 successful?  
   
2. How many times, out of the 14 plays you saw, was No. 27 successful?  
   
3. How many plays did No. 43 carry the ball?  
   
4. How many plays did No. 27 carry the ball?  
   
5. How many plays did No. 43 block?  
   
6. How many plays did No. 27 block?  
   
7. How many plays did No. 43 gain yardage when he carried the ball?  
   
8. How many plays did No. 27 gain yardage when he carried the ball?  
   
9. How many plays did No. 43 lose yardage when he carried the ball?  
   
10. How many plays did No. 27 lose yardage when he carried the ball? 
   
11. How many plays did No. 43 get tackled by a defensive lineman?  
   
12. How many plays did No. 43 get tackled by a linebacker?  
   
13. How many plays did No. 43 get tackled by a defensive back?  
   
14. How many plays did No. 27 get tackled by a defensive lineman?  
   
15. How many plays did No. 27 get tackled by a linebacker?  
   
16. How many plays did No. 27 get tackled by a defensive back?  
   
17. How many times was No. 43 uninvolved in the play?  
   
18. How many times was No. 27 uninvolved in the play?  
   
19. How many plays did No. 43 miss a block?  
   
20. How many plays did No. 27 miss a block?  
   
21. How many plays did No. 43 break a tackle?  
   
22. How many plays did No. 27 break a tackle?
Please read the following questions carefully, and indicate your response in the appropriate space. Make sure your answers are whole numbers. Mark all of your answers directly on the questionnaire, and make sure you only choose ONE answer to each question.

1. How many times, out of the 14 plays you saw, was No. 70 successful?

2. How many times, out of the 14 plays you saw, was No. 68 successful?

3. How many times did the defensive player No. 70 try to block make the tackle?

4. How many times did the defensive player No. 68 try to block make the tackle?

5. How many plays did No. 70 push his defensive opponent backwards?

6. How many plays did No. 68 push his defensive opponent backwards?

7. How many plays did No. 70 miss a block?

8. How many plays did No. 68 miss a block?

9. How many times was No. 70 uninvolved in the play?

10. How many times was No. 68 uninvolved in the play?

11. How many plays did No. 70 block a defensive lineman?

12. How many plays did No. 70 block a linebacker?

13. How many plays did No. 70 block a defensive back?

14. How many plays did No. 68 block a defensive lineman?

15. How many plays did No. 68 block a linebacker?

16. How many plays did No. 68 block a defensive back?

17. How many plays did No. 70 knock his defensive opponent down?

18. How many plays did No. 68 knock his defensive opponent down?

19. How many times did No. 70 make more than one block on a single play?

20. How many times did No. 68 make more than one block on a single play?
Appendix F

Manipulation Check Questionnaire
QUESTIONNAIRE

Please respond to the following scales based on your impressions of the RUNNING BACKS viewed in the film.

1) Your ability to critically observe the performances of the running backs was:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Below</td>
<td>Average</td>
<td>Difficulty</td>
<td>Above</td>
<td>Average</td>
<td>Extremely</td>
</tr>
<tr>
<td>Difficult</td>
<td>Average</td>
<td>Difficulty</td>
<td>Average</td>
<td>Difficulty</td>
<td>Difficulty</td>
<td></td>
</tr>
</tbody>
</table>

2) Your ability to evaluate the performances of the running backs was:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Below</td>
<td>Average</td>
<td>Difficulty</td>
<td>Above</td>
<td>Average</td>
<td>Extremely</td>
</tr>
<tr>
<td>Difficult</td>
<td>Average</td>
<td>Difficulty</td>
<td>Average</td>
<td>Difficulty</td>
<td>Difficulty</td>
<td></td>
</tr>
</tbody>
</table>

3) The overall difficulty of forming impressions of running backs was:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Below</td>
<td>Average</td>
<td>Difficulty</td>
<td>Above</td>
<td>Average</td>
<td>Extremely</td>
</tr>
<tr>
<td>Difficult</td>
<td>Average</td>
<td>Difficulty</td>
<td>Average</td>
<td>Difficulty</td>
<td>Difficulty</td>
<td></td>
</tr>
</tbody>
</table>

4) To what extent did you find the observational purpose clear to you?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Totally</td>
<td>Somewhat</td>
<td>Somewhat</td>
<td>Clear</td>
<td>Very</td>
<td>Clear</td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>Unclear</td>
<td>Clear</td>
<td>Clear</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5) To what degree did you attempt to form an impression of your target players?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>Rarely</td>
<td>Some of</td>
<td>Often</td>
<td>Very</td>
<td>Always</td>
<td>Often</td>
</tr>
<tr>
<td>the time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6) To what degree did you attempt to memorize details while observing your target players?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>Rarely</td>
<td>Some of</td>
<td>Often</td>
<td>Very</td>
<td>Always</td>
<td>Often</td>
</tr>
<tr>
<td>the time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix G

Cronbach's Accuracy Components
Algebraic Difference Score Formulae for the Four Components of Accuracy

\[ E^2 = (\bar{r}_{..} - \bar{r}_{..})^2, \]
\[ DE^2 = \frac{1}{n} \sum_i (r_{ik} - \bar{r}_{i.}) - (\bar{r}_{..} - \bar{r}_{..})^2, \]
\[ SA^2 = \frac{1}{k} \sum_j (r_{.jk} - \bar{r}_{.j}) - (\bar{r}_{..} - \bar{r}_{..})^2, \]
\[ DA^2 = \frac{1/kn}{\bar{r}_{..}} \sum_i \sum_j (r_{ij} - \bar{r}_{i.} - \bar{r}_{.j} + \bar{r}_{..}) - (\bar{r}_{..} - \bar{r}_{..} + \bar{r}_{..})^2, \]

where \( r_{ij} \) and \( t_{ij} \) = rating and true score for raters \( i \) on dimension \( j \), \( \bar{r}_{i.} \) and \( \bar{r}_{.j} \) = mean rating and mean true score for rater \( i \); \( \bar{r}_{ij} \) and \( \bar{r}_{..} \) = mean rating and mean true score for dimension \( j \); and \( \bar{r}_{..} \) = mean rating and mean true score over all raters and dimensions.
Appendix H

Behavior Frequency True Scores
Please read the following questions carefully, and indicate your response in the appropriate space. Make sure your answers are whole numbers. Mark all of your answers directly on the questionnaire, and make sure you only choose ONE answer to each question.

1. How many times, out of the 14 plays you saw, was No. 43 successful? 10
2. How many times, out of the 14 plays you saw, was No. 27 successful? 13
3. How many plays did No. 43 carry the ball? 3
4. How many plays did No. 27 carry the ball? 7
5. How many plays did No. 43 block? 11
6. How many plays did No. 27 block? 4
7. How many plays did No. 43 gain yardage when he carried the ball? 3
8. How many plays did No. 27 gain yardage when he carried the ball? 5
9. How many plays did No. 43 lose yardage when he carried the ball? 0
10. How many plays did No. 27 lose yardage when he carried the ball? 0
11. How many plays did No. 43 get tackled by a defensive lineman? 2
12. How many plays did No. 43 get tackled by a linebacker? 6
13. How many plays did No. 43 get tackled by a defensive back? 1
14. How many plays did No. 27 get tackled by a defensive lineman? 1
15. How many plays did No. 27 get tackled by a linebacker? 3
16. How many plays did No. 27 get tackled by a defensive back? 1
17. How many times was No. 43 uninvolved in the play? 0
18. How many times was No. 27 uninvolved in the play? 3
19. How many plays did No. 43 miss a block? 2
20. How many plays did No. 27 miss a block? 0
21. How many plays did No. 43 break a tackle? 1
22. How many plays did No. 27 break a tackle? 2
Please read the following questions carefully, and indicate your response in the appropriate space. Make sure your answers are whole numbers. Mark all of your answers directly on the questionnaire, and make sure you only choose ONE answer to each question.

1. How many times, out of the 14 plays you saw, was No. 70 successful?  
2. How many times, out of the 14 plays you saw, was No. 68 successful?  
3. How many times did the defensive player No. 70 try to block make the tackle?  
4. How many times did the defensive player No. 68 try to block make the tackle?  
5. How many plays did No. 70 push his defensive opponent backwards?  
6. How many plays did No. 68 push his defensive opponent backwards?  
7. How many plays did No. 70 miss a block?  
8. How many plays did No. 68 miss a block?  
9. How many times was No. 70 uninvolved in the play?  
10. How many times was No. 68 uninvolved in the play?  
11. How many plays did No. 70 block a defensive lineman?  
12. How many plays did No. 70 block a linebacker?  
13. How many plays did No. 70 block a defensive back?  
14. How many plays did No. 68 block a defensive lineman?  
15. How many plays did No. 68 block a linebacker?  
16. How many plays did No. 68 block a defensive back?  
17. How many plays did No. 70 knock his defensive opponent down?  
18. How many plays did No. 68 knock his defensive opponent down?  
19. How many times did No. 70 make more than one block on a single play?  
20. How many times did No. 68 make more than one block on a single play?
VITA

Robert T. Brill

PERSONAL INFORMATION

Born: 27 April 1965

Social Security Number: 207 54 4602

Marital Status: Married, no children

Business Address: Department of Psychology
Moravian College
Bethlehem PA 18018

Business Phone: (215) 861-1561

Home: 2214 Townline Way
Blue Bell PA 19422

Home Phone: (215) 278-0629

EDUCATION

B.A. LaSalle University, Pennsylvania, Magna Cum Laude, 1987
Dual Major: Psychology and Management

M.S. Virginia Polytechnic Institute and State University, 1989
Major field of study: Industrial/Organizational Psychology

Thesis title: Reciprocal Influence of Subordinate Reaction on Ratings: Supervision and Attributions of Supervisors
Major Advisor: Dr. Neil M. A. Hauenstein

Ph.D. Virginia Polytechnic Institute and State University, (expected August 1992)
Major field of study: Industrial/Organizational Psychology

Dissertation title: The Effects of Job Knowledge, Observation Purpose, and Judgment Complexity on Recall and Rating Ability
Major Advisor: Dr. Neil M. A. Hauenstein
EXPERIENCE

Teaching

1991-Present  Instructor
Moravian College, Department of Psychology
  Responsible for teaching three courses (each
  course meets four class hours per week), including
  Social Psychology, I/O Psychology, Work
  Motivation, and Statistics/Methods. Supervise
  field studies and serve as temporary Faculty
  Advisor for Psi Chi, and as Faculty Associate for
  a student resident hall.

Spring 1991  Graduate Teaching Assistant, Virginia Polytechnic
Institute and State University, Department of
Psychology
  Taught a three credit course on Principles of
  Psychological Research.

1988 - 1991  Coordinator of Introductory Psychology, Virginia
Polytechnic Institute and State University, Department of
Psychology
  Coordinated the activities of 1200
  undergraduate students per semester including
  registration, test construction, and
  grading; supervised 18-20 graduate student
  teachers conducting the Introductory Psychology
  labs.

1988 - 1990  Graduate Teaching Assistant, Virginia Polytechnic
Institute and State University, Department of
Psychology
  Instructed lab sessions of introductory
  psychology course.

Research

1/89 - 12/89  Thesis Research
Chairperson: Neil M. A. Hauenstein
Virginia Polytechnic Institute and State
University
  Designed and conducted a laboratory study which
  investigated the potential effects of subordinate
  reactions to appraisals on supervisors subsequent
  ratings, attributions, and amount of supervision.
11/87 - 6/88  Research Assistant
Steven Zaccaro
Virginia Polytechnic Institute and State University
Conducted literature review, data collection and analyses, and aided in research design. This research investigated factors which influence cohesion, and efficacy among groups of people.

Consulting/Work

1991- present Communications Task Force
Lehigh Valley Chapter of the American Red Cross
Upon the merger of three chapters, served as facilitator and advisor on committee to work toward developing and implementing strategies to improve upward and downward flow of communication, as well as increasing inter-departmental education.

Montgomery Regional Hospital contract to Neil M. A. Hauenstein, Virginia Polytechnic Institute and State University
Assisted in developing a pay structure for over one hundred job titles based on a common metric job analysis. Conducted interviews with job family incumbents to revamp performance appraisal system in order to promote merit-based pay raises and improve employee feedback at the task level.

1989 - 1990 Skill-based Selection System
Shenandoah Life Insurance Company contract to Neil M. A. Hauenstein, Virginia Polytechnic Institute and State University
Conducted job analysis interviews with company members to establish interpersonal, problem solving, and communication skills required for superior customer relations. Constructed a task inventory from interview data, validated a selection battery based on standards of customer relations developed from the task inventory, assisted in creating a compensation system for all employment levels of the company.

6/90 - 8/90 Human Factors Internship
International Business Machines Corp.
Assisted in collection of data assessing usability principles within software development labs; organizing data base and conducting data analyses; and preparation of reports to each line-of-business manager. Completed a literature search on human-computer interaction.
HONORS AND AWARDS

1989 - 1991 Virginia Polytechnic Institute and State University Departmental Tuition Waiver
1989 - 1991 Departmental Graduate Student Representative
1983 - 1987 LaSalle University Christian Brother's Scholarship
1987 LaSalle University Student Service Award
1987 LaSalle University Male Scholar-Athlete Award
1986 - 1987 All Scholastic MAAC Conference Team for Soccer
1983 Provincial Medal of Honor, (Northeast Catholic HS)
1983 William McLean Grant-in-Aid Scholarship
1983 St. Joseph's University Presidential Scholarship

PROFESSIONAL ORGANIZATIONS AND ACTIVITIES

Member, American Psychological Association

Member, Society of Industrial/Organizational Psychology

Member, Southeastern Psychological Association

Member/Faculty Advisor, Psi Chi

PUBLICATIONS


PRESENTATIONS


SKILLS

Experience using SPSS and SAS computer packages. Statistical training includes analysis of variance, correlation and regression analysis, analysis of covariance, factor analysis, path analysis, and multivariate analysis. Possess strong organizational abilities, and outstanding communication and interpersonal skills.