

COGNITIVE STRATEGIES IN JUDGMENT: THE EFFECT OF PURPOSE,  
CUE DIMENSIONALITY, AND COGNITIVE COMPLEXITY ON STUDENT  
EVALUATION OF INSTRUCTORS

by

NAND KISHOR

B.Ed., M.A. The University of the South Pacific

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES  
EDUCATIONAL PSYCHOLOGY & SPECIAL EDUCATION

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 22, 1987

© NAND KISHOR, 1987

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Psychology & Special Education

The University of British Columbia  
1956 Main Mall  
Vancouver, Canada  
V6T 1Y3

Date September 22nd., 1987.

## ABSTRACT

This investigation focused on describing cognition in performance judgment of teaching in higher education. The influence of appraisal purpose and cue dimensionality was observed on subjective importance and utilization of information. Information integration strategies were examined in relation to purpose and cognitive complexity. Exploratory analysis focused on the measurement of good instructor schema profiles, and on the effect of cognitive complexity on halo in performance ratings.

Seventy subjects were assigned randomly to two purpose conditions in the experiment: summative and formative judgment. Two questionnaires, two rating tasks, and a Role Construct Repertory grid were administered for data collection. The data were analyzed through regression modeling at the individual level and via analysis of variance procedures at the group level.

The results indicate that the impact of cue dimensions is strong on subjective importance and utilization of information but varies with the purpose of appraisal. Raters valued and utilized trait information more than behavior information in evaluation required for personnel decisions. Where evaluation was feedback on the quality of teaching and expressed the need for improvement, raters utilized behavior information more than trait information. This pattern of information utilization suggests that saliency of information in performance judgment is a function of purpose and cue dimensionality, and that appraisal purpose has an effect on raters' cognition through schematic processing.

The results also show that the use of varied strategies in mentally

integrating dimensions of information is affected by raters' cognitive complexity. Although subjects mainly used compensatory strategies, the complex individuals used noncompensatory strategies as well. Exploratory analysis shows that cognitive complexity also affects halo in rating judgments. The findings seem to support the validity of student rating of instructors, and the utility of cognitive complexity construct in understanding performance judgment.

It is suggested that the influence of schematic processing and cue saliency be addressed in further theorizing and research on performance judgment. As well, the inclusion of purpose of judgment and developmental constructs, such as cognitive complexity, is recommended for theorizing and research on judgment processes.

## TABLE OF CONTENTS

Abstract .....	ii
List of Tables .....	vi
List of Figures .....	vii
Acknowledgements .....	viii
 I. THE NATURE OF THE STUDY .....	 1
A. Introduction .....	1
B. The Problem .....	4
C. The Purpose of the Study .....	8
D. The Context .....	10
E. Significance of the Study .....	11
 II. REVIEW OF THE LITERATURE .....	 13
A. Influences on Judgment Processes .....	13
1. Cognitive Limitations .....	13
2. Mental Models .....	15
3. Judgment Task Environment .....	17
B. Cognitive Views of Performance Judgment .....	19
1. Cognitive Distortion in Performance Judgment .....	20
2. Performance Appraisal as Person Perception .....	21
3. Performance Appraisal as Prototype Matching .....	22
4. Performance Appraisal as Social Perception .....	24
C. Research on Raters' Cognition .....	26
1. Cognitive Effect of Appraisal Purpose .....	26
2. Effect of Cue Dimensionality .....	32
3. Influence of Cognitive Complexity .....	33
D. Evaluation of Teaching .....	38
E. Summary .....	41
 III. HYPOTHESES, QUESTIONS, AND METHOD .....	 44
A. Rationale for Hypotheses and Method .....	44
B. Hypotheses and Exploratory Questions .....	48
1. Importance of Information .....	48
2. Utilization of Information .....	49
3. Information Importance and Utilization Consistency .....	49
4. Information Integration .....	49
5. Exploratory Questions .....	50
C. Methodology .....	50
1. Subjects .....	50
2. Instruments .....	51
a. Importance of Information Measure .....	51
b. Rating Judgment Task A .....	52
c. Cognitive Complexity Measure .....	55
d. Good Instructor Schema Measure .....	57
e. Rating Judgment Task B .....	59

3. Experimental Design and Variables .....	60
4. Data Collection .....	62
IV. ANALYSIS AND RESULTS .....	64
A. Test of the Hypotheses .....	64
1. Importance of Information .....	65
2. Utilization of Information .....	69
3. Subjective Importance and Utilization Consistency .....	73
4. Information Integration .....	75
B. Exploratory Analysis .....	82
1. Measuring A Good Instructor Schema .....	82
2. Cognitive Complexity and Halo .....	84
V. DISCUSSION .....	86
A. Importance and Utilization of Information .....	86
1. Effect of Purpose .....	86
2. Effect of Cue Dimensionality .....	88
3. Interactive Effect of Purpose and Cues .....	90
4. Subjective Importance and Utilization Consistency .....	93
B. Information Integration .....	94
1. Effect of Purpose .....	96
2. Effect of Cognitive Complexity .....	97
C. Findings From Exploratory Analysis .....	98
1. Measurement of Schema .....	98
2. Cognitive Complexity and Halo .....	100
D. Summary of the Findings and Conclusions .....	102
E. Strengths and Limitations of the Study .....	105
F. Implications .....	107
G. Directions for Further Research .....	113
VI. REFERENCES .....	117
VII. APPENDIX .....	132
A. Glossary .....	132
B. Important Information Measure .....	134
1. For Summative Condition .....	134
2. For Formative Condition .....	135
C. Performance Rating Task A .....	136
1. For Summative Condition .....	136
2. For Formative Condition .....	138
3. Coding and Rotation of Profiles .....	140
D. Cognitive Complexity Measure .....	141
E. Good Instructor Schema Measure .....	142
F. Performance Rating Task B .....	147

## *LIST OF TABLES*

Table 1: The Effect of Purpose and Cue Dimensionality on Importance Ratings .....	66
Table 2: Mean Importance of Performance Related Information .....	68
Table 3: Mean, Median and Range of Variance Explained by Regression Models .....	70
Table 4: The Effect of Purpose and Cue Dimensionality on Information Utilization .....	71
Table 5: Mean Regression Weights for Formative and Summative Judgment ....	73
Table 6: Variance Extracted from Original Sets of Variables by Canonical Variates .....	74
Table 7: Mean and Standard Deviation of Variance Explained .....	75
Table 8: Mean Diagnostic Ratios in Good Instructor Schema Profile .....	83

## *LIST OF FIGURES*

Fig. 1: Mean Importance Rating of Cues .....	67
Fig. 2: Mean Weight of Information Utilized .....	72
Fig. 3: Plot of Cell Means for Subjects 3 and 8 (Formative) .....	77
Fig. 4: Plot of Cell Means for Subjects 13, 27, 29 and 30 (Formative) .....	78
Fig. 5: Plot for Cell Means of Subjects 2, 5, 6 and 15 (Summative) .....	79
Fig. 6: Plot for Cell Means of Subjects 19, 31, 32 and 33 (Summative) .....	80



## ACKNOWLEDGEMENTS

I thank my program advisor and research supervisor, Dr. Marshall Arlin, for his continued support, encouragement, and guidance throughout my tenure at UBC. Thanks also to members of my dissertation committee, Dr. Ronald Jarman and Dr. Peter Grimmett, for their helpful comments in the completion of this dissertation.

As well, thanks to Professors Walter Boldt, Seong-Soo Lee, Anne Triesman, Nancy Suzuki, Patricia Arlin, Daniel Kahneman, Robert Conry, Helga Jacobson, and Ian Housego for providing diverse and rich intellectual stimulation that contributed to my development.

During the course of my study I was affiliated to The Center for the Study of Teacher Education (CSTE) at UBC. This dissertation is a result of the cooperation between the Department of Educational Psychology & Special Education and CSTE.

An acknowledgement is offered to CSTE for a Doctoral Fellowship, Educational Research Services & Computing (UBC) for Research Assistantship, Canadian International Development Research Center for a Doctoral Fellowship, and The University of the South Pacific (Fiji) for study leave.

No dissertation would be completed without the care and concern of many friends. To them, I express my sincere gratitude. Two persons need special mention. I am greatly indebted to Dr. Daniel Birch (UBC) and Dr. Diana Kendall (Canberra) for the inspiration and support they provided me to study in this part of the world.

On a more personal note, I am grateful to my parents for instilling in me the belief best summarized in the words "awake, arise, and stop not till the goal is reached." My achievement is dedicated to them. To Lila, Lin, and Kevin I owe too much for their patience, love and care.

## I. THE NATURE OF THE STUDY

### A. INTRODUCTION

Human judgment, the mental act of weighing and combining information from cues or information to make an inference about some criterion, is a phenomenon of every human experience. It includes evaluation, decision making, choice, and the selection of the response after perceiving the stimulus. Human judgment is an inescapable aspect of thinking. As a result, it is being studied by researchers from various disciplines. Within the psychological literature, two lines of inquiry on how we make judgments and decisions are identifiable. One approach has been to apply prescriptive models of choice for predicting human judgment; another approach has been to describe human judgment as constrained by cognitive mechanisms.

Prescriptive models, though successful in describing simpler automatic mental processes, fail to describe judgments that require thoughtful deliberations (Pitz & Sachs, 1984). Prescriptive models of human judgment are derived from probability theory and from Expected Utility theory (von Neumann & Morgenstern, 1947). A hybrid of these theories is the Bayesian decision theory (Edwards, 1968; Raifa & Schlaifer, 1961), widely applied in the study of predictive judgment. Prescriptive models provide a set of rules for combining beliefs and preferences in making a judgment, but as descriptions of human judgment prescriptive models have had setbacks (Einhorn & Hogarth, 1981; Schoemaker, 1982). A number of prescriptive axioms are violated in human induction (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1983).

Another approach has been to describe human judgment as affected by the cognitive functioning of the human mind. Theoretical insights in this line of inquiry have been gained from the recognition that human behavior depends on the nature of the environment, the nature of the organism, and the means the organism has developed for coping with the environment (Brunswik, 1952; Hogarth, 1981; Piaget, 1936/1970; Simon & Newell, 1971). Within this approach, the main thrust has been to uncover how information processing mechanisms of the mind constrain judgments. A number of errors and inconsistencies in human judgment have been identified (for reviews see Einhorn & Hogarth, 1981; Slovic, Fischhoff, & Lichtenstein, 1977). These errors and inconsistencies have been attributed to cognitive limitations, judgmental heuristics, and schematic processing<sup>†</sup> (Hogarth, 1980; Kahneman, et al., 1982, Taylor & Crocker, 1981).

The present study was an attempt to enrich our understanding of how certain factors affect cognitive processing of information in evaluative judgment. According to a conceptualization by Hogarth (1980), judgment takes place within a system composed of three elements. First, is the person; second, is the task environment within which the person makes judgments; and third, are the actions that result from judgment which may subsequently affect both the person and the environment. This study focused on the first two elements. Cognitive complexity or the disposition for processing multidimensional data, was a developmental aspect of the person; cue dimensionality or the nature of information and purpose for judgment were two aspects of the task environment. The task environment was performance judgment.

---

<sup>†</sup>Terms peculiar to the domain of this study are included in the glossary in Appendix A.

Performance judgment plays an important part in human resources management. It is central to important decisions related to staff recruitment, training and development, promotion and career planning. The pressure to evaluate the performance of individuals usually develops from productivity concerns in an industrial setting and accountability concerns in professional occupations like teaching. Not surprisingly, since Thorndike's (1920) study of the halo effect in supervisors' ratings of their subordinates, considerable effort has been devoted to eliminating errors and biases in performance evaluation.

The literature concerning performance evaluation is replete with psychometric studies attempting to improve ratings and rating scales (Borman & Dunnette, 1975; Dickinson & Zellinger, 1980). A large number of performance rating scales and instruments have been devised. A fair number of studies have also focused on rater training in order to find ways to eliminate common rating errors such as leniency-stringency, central tendency, and the halo effect (Bernardin & Pence, 1980).

Despite much interest and effort, there is considerable dissatisfaction with the progress made in the field of performance appraisal research (Borman, 1978; DeCotiis & Petit, 1978; Kane & Lawler, 1979; Landy & Farr, 1980). In a comprehensive review article, Landy and Farr (1980) concluded that previous approaches in resolving problems in performance appraisal have not been appreciably successful, and any further research on improving the rating format and rater training would probably be a futile endeavor. They stated, "It is time to stop looking at the symptoms of bias in rating and begin examining potential causes" (p. 101). They suggested that further research should examine raters'

cognition. Similar suggestions have been made by others as well (DeNisi, Cafferty, & Meglino, 1984; Feldman, 1981; Ilgen & Feldman, 1983).

Researchers in the past have not paid much attention to a rater's cognition in performance judgment. Expressing this concern, Wexley and Klimoski (1984) compared traditional performance appraisal research to a *black box* approach to describing human behavior, where a worker's performance represents the stimuli, the rating represents the response, and errors and biases are seen as weaknesses in the instruments, or lack of rater training. Understanding cognitive processing in performance judgment seems critical to our understanding of the problems in performance appraisal in particular, and human judgment in general.

## B. THE PROBLEM

Most of the available research on judgment processes deals with probabilistic inference in gambles, business and medical decision making, and risk perception. Evaluative judgment such as performance appraisal has not been a popular context for research on judgment processes in the past, as can be seen from major review papers (Einhorn & Hogarth, 1981; Pitz & Sachs, 1984; Slovic, Fischhoff, & Lichtenstein, 1977). However, there is a need to study judgment processes in performance appraisal, because inconsistencies and biases are a pervasive problem in this area (Landy & Farr, 1980).

Previous research on judgment strategies has concentrated on a number of stimulus features in a judgment task. These include cue inter-relationships, set size effects, number and format of cues, cue response compatibility, extremity of

information, cue redundancy, and primacy and recency effects on judgment (for a review see Slovic & Lichtenstein, 1971). More recently, researchers have found that other features of the stimuli such as information load (Payne, 1976), information presentation (Crocker, 1981; Tversky & Kahneman, 1981), and cue dimensionality (Wallsten, 1980) also affect cognitive processes that produce a judgmental response.

Nevertheless, relatively little attention has been paid to the effect of purpose of judgment in the study of mental strategies in judgment. It has been stressed that judgment is primarily exercised to facilitate action (Einhorn & Hogarth, 1978; Hogarth, 1980). Because an action serves a purpose (or purposes), how judgments are *formed* may be influenced by the purpose for which a judgment is required. It is hard to think of a situation where formal evaluative judgment does not serve a purpose. Whether purpose may determine the type of information necessary for a judgment, and thereby have an effect on the utilization of cues needs to be examined.

A common result in research on cue dimensionality is that cue saliency affects the use of information in judgments. Cue saliency in previous research has been considered mainly in terms of number, frequency, and perceptual characteristics of the stimuli. The emerging general theory is open as to what determines cue saliency (Wallsten & Barton, 1982). In certain judgment situations, cue dimensionality may be reflected in information content. For example, cues in performance judgment provide information concerning traits and role behaviors. Hence, research is needed to determine whether cue saliency is a function of information content and the purpose of judgment.

Furthermore, in their review paper Pitz and Sachs (1984) have noted that developmental constructs have been largely ignored in the study of human judgment processes. These authors drew attention to the role of a person's level of moral development, but another factor may be a person's cognitive complexity. Cognitive complexity, a developmental construct, relates to a person's disposition in processing multidimensional data (Bieri, Atkins, Briar, Leaman, Miller, & Tripodi, 1966). Therefore, whether cognitive complexity affects the use of information integration strategies needs examination.

Researchers of human judgment have noted that "human reasoning cannot be adequately described in terms of content-independent formal rules" (Kahneman & Tversky, 1982). For a meaningful investigation of the issues raised above, performance judgment afforded an ideal task environment. Besides, the causes of errors and biases in performance judgment are largely unknown (Landy & Farr, 1980), and researchers have already attempted to study the role of appraisal purpose (Zedeck & Cascio, 1982) and cognitive complexity (Schneier, 1977) on rating judgments. Further, as performance judgment is based on traits and role behaviors (Wexley & Klimoski, 1984), it provided an environment where the effect of cue dimensionality represented by information content could be studied.

Several researchers have proposed information processing approaches to studying performance judgment. Landy and Farr (1980) suggested that important aspects in processing a rating judgment were the manner in which a rater "treats" (mentally integrates) the available information and the purpose of appraisal. Cooper (1981) proposed that cognitive distortion, introduced by a rater's beliefs and implicit theories in the processing stages, was the source of halo in

performance ratings. Ilgen and Feldman (1983) drew attention to a rater's cognitive structures and the prototype matching processes. DeNisi, Cafferty, and Meglino (1984) suggested the importance of a rater's memory, cognitive complexity, and cognitive style on the processing of performance information. In these views, a common emphasis is on the effect of purpose on a rater's cognition, and on the effect of a rater's cognitive complexity on rating properties.

Nevertheless, the available research findings on the effect of purpose on raters' cognition is not only limited, but also unclear (McIntyre, Smith, & Hassett, 1984; Murphy, Balzer, Kellam, & Armstrong, 1984; Williams, DeNisi, Blencoe, & Cafferty, 1985; Zedeck & Cascio, 1982). These studies fail to clarify how cognitive processing in performance judgment is affected by appraisal purpose. Likewise, the findings on the effect of cognitive complexity on performance ratings, are contradictory (Bernardin, Cardy, & Carlyle, 1982; Schneier, 1977; Lahey & Saal, 1981; Sauser & Pond, 1981). Moreover, researchers in the past have focused mainly on psychometric properties of ratings, although cognitive complexity may affect information integration because the construct represents a person's disposition in processing multidimensional data (Bieri, et al., 1966).

Futhermore, neither the cognitively oriented theoretical models (Cooper, 1981; DeNisi et al., 1984; Ilgen & Feldman, 1983; Landy & Farr, 1980), nor the empirical studies on raters' cognition (McIntyre, Smith, & Hassett, 1984; Murphy, Balzer, Kellam, & Armstrong, 1984; Williams, DeNisi, Blencoe, & Cafferty, 1985; Zedeck & Cascio, 1982) have addressed the influence of cue dimensionality in performance judgment. However, cue dimensionality has been found to affect



judgment strategies in other areas (Slovic & Lichtenstein, 1971; Pitz & Sachs, 1984). Important cue dimensions in performance judgment are traits and role behaviors because "what a person is" and "what a person does" make up the appraisal content (cf. Wexley & Klimoski, 1984). There is little research to demonstrate empirically the conditions under which trait and behavior information become salient. Therefore, whether trait and behavior cues bear an influence on the utilization of performance information needs investigation.

In essence, we do not have a clear understanding of influences on the mental processes that lead to a judgment in performance appraisal. Whether purpose determines the type of information necessary for a judgment, and thereby, affects the utilization of cues, lacks evidence. Further, we lack theories that may explain what determines cue saliency in performance judgment, and we lack evidence on whether cognitive complexity affects the way raters mentally combine performance information.

### C. THE PURPOSE OF THE STUDY

This investigation tested hypotheses pertaining to the influence of appraisal purpose, cue dimensionality, and cognitive complexity on how rating judgments are *formed*. Specifically, the effects of these variables were examined on subjective valuation, utilization, and integration of information - the processes that lead to a rating response (Anderson, 1981).

The effects of purpose and cue dimensionality were observed on subjective importance and utilization of performance related information. Rating judgments

were required for the purposes of (a) providing feedback on the quality of performance - formative judgment, and (b) indicating the suitability for promotion - summative judgment. The ratings were an expression of judgment only, and did not include any justification, guidance, and recommendations for improvement; because the main interest in this study was on how rating judgments are formed and not on how evaluations are to be communicated for development. Cue dimensionality was represented by trait and role information. The predictions that purpose and cue dimensionality will influence subjective importance and utilization of information, and that cue utilization will be consistent with subjective importance of cues, were examined.

The effects of purpose and cognitive complexity were observed on the use of cue integration strategies. The information integration strategies were of two broad types: compensatory and noncompensatory. A compensatory strategy is in use when cues are combined additively or by averaging across dimensions; a noncompensatory strategy is in use when cues are combined interactively or multiplicatively across dimensions (Einhorn, 1970, 1971; Hogarth, 1980). The predictions that appraisal purpose, and cognitive complexity will influence the use of information integration strategies were examined.

In addition, this study served two exploratory purposes. One was to explore the measurability of good instructor schema - the mental representation of the characteristics of a good instructor. The other was to explore the effect of cognitive complexity on halo in performance ratings, using correlational techniques to assess halo (Pulakos, Schmitt, & Ostroff, 1986).

#### D. THE CONTEXT

The primary goal of this investigation was to describe cognitive phenomena underlying human judgment in performance appraisal. The goal of a study often imposes its own constraints on design and procedure. To achieve the goal of describing and understanding how certain variables affect cognitive processes in performance judgment, this study was conducted as a laboratory experiment, so that the nature, amount, and presentation of information could be controlled.

The use of a laboratory procedure, however, relies on the assumption that human cognitive processing varies little when task demands are similar. A further assumption is that people have no special reason to distort their mental functioning in a simulated task. The fundamental premise of this study (and of related studies in the literature) is that how a person processes information in a contrived setting can provide important heuristic clues as to the mental strategies underlying performance judgment. These clues then may indicate directions for research within the ecological reality of the phenomena of interest.

Performance judgment takes place in many settings. A suitable setting for this study was appraisal of teaching. Evaluation of teaching takes place at all levels of the educational system. The present study was anchored in the appraisal of teaching in higher education. A majority of the previous studies of performance appraisal which deal with purpose and cognitive complexity have been done on student evaluation of university teaching. Accordingly, it was reasoned that a similar sample and judgment task would facilitate the interpretation of the results.

Furthermore, students in higher education are increasingly required to provide evaluations on their instructors. Interest in student evaluation of teaching at colleges and universities is growing (Dunkin & Barnes, 1986; Marsh, 1984). At the university level, evaluation of teaching is required for the purposes of feedback to the instructor and for tenure/promotion decisions, but much controversy surrounds the reliability and validity of student evaluations of teaching (Centra, 1979; Cohen, 1981).

### E. SIGNIFICANCE OF THE STUDY

The benefits of the present investigation are mainly at the theoretical level. Studying how task features and individual characteristics influence the cognitive processes in judgment of teaching, increases our understanding of how performance judgments are formed. If we know how judgments are produced, we may succeed in reducing the fallibility in human judgment in general, and in the evaluation of teaching in particular.

Although all evaluative judgment may serve specific purposes, we do not know much about the effect of purpose on how performance judgments are mentally processed. The results here may clarify our understanding of the effect of appraisal purpose on performance judgment from a cognitive perspective. In addition, the opportunity to infer cognitive processing of performance information experimentally, may provide clues to the conditions leading to bias and inaccuracy in performance judgment. An accumulation of such knowledge may provide a knowledge base from which may evolve a formal *model of the rater* as theoretical basis for measurement procedures (Feldman, 1981; Krantz et al., 1971).

Several studies on human judgment processes have found that judgment strategies are affected by cue dimensionality. In most of these studies, the dimensions of the stimuli are arrayed from most to least salient, and the effect of cue saliency on judgment is observed (Wallsten, 1980). As cue dimensionality in this study reflected traits and role behaviors, the findings may suggest the importance of conceptualizing cue dimensionality in terms of semantic features of information.

And finally, the neglect of developmental constructs on research in human judgment is unfortunate (Pitz & Sachs, 1984). Investigating how cognitive complexity, a developmental construct, affects the use of cue integration strategies may further add to our knowledge of processes in judgment. The problems people have in using multiple strategies for integrating cues into a judgment, may well be a result of their developmental maturity. Developmental constructs may add a new dimension to theorizing on human judgment.

## **II. REVIEW OF THE LITERATURE**

This chapter provides a review of the pertinent theoretical and empirical literature related to cognitive processes in human judgment and performance appraisal. This is accomplished in five parts. The first part (A) is a discussion of influences on judgment processes. The second part (B) is a discussion of cognitive perspectives on performance judgment. An analysis of research bearing on raters' cognition is presented in part three (C). The fourth part (D) is a discussion concerning evaluation of teaching. The chapter concludes with a summary in part five (E).

### **A. INFLUENCES ON JUDGMENT PROCESSES**

Researchers have studied many factors which may affect human judgment. Generally, the literature indicates that judgment and decision processes are affected by (1) the cognitive limitations of the mind; (2) mental models, schemata or cognitive structures; (3) features of the task environment. These influences on the formation of judgment are discussed in the following three sub-sections.

#### **1. Cognitive Limitations**

The limitations of our mental apparatus affect how we cope with large amounts of information available to our sensory modalities. The human mind has a number of limitations. Attention is a scarce resource and never totally available (Kahneman, 1973). The storage capacity of the working memory is limited (Miller, 1956), and the processing of information is mainly in a serial

manner (Newell & Simon, 1972). Although the working memory is under the immediate control of a person, it allows manipulation of only about seven "bits" (Miller, 1956) or five "chunks" (Simon, 1974) of unrelated information in a serial manner. Serial processing produces recency and primacy bias in extracting meaning from information on which judgments may be based. Furthermore, unless information in working memory is rehearsed, it is rapidly lost (Atkinson & Shiffrin, 1971) because the processing duration is brief, usually less than ten seconds (Murdock, 1961).

Limitations on attention and memory storage make perception selective and anticipatory (Neisser & Becklon, 1975). As a result, what a person expects to see tends to determine what a person does see. It has been found that people have a strong tendency to seek information that is consistent with their expectations, than to seek information that may disconfirm their expectations (Webster, 1964). For instance, researchers of social cognition have found that aspects consonant with personal view are generally given more weight than nonconsonant information (Nisbett & Ross, 1980).

Limited attentional resource, limited storage capacity of the working memory, sequential processing, brief execution times, and selective perception mean that the mind usually can process only a small amount of information at one time. These limitations require an individual to develop and utilize simplifying cognitive processing operations in order to deal with large amounts of data in a judgment task.

## 2. Mental Models

Cognitive processing of large amounts of data is facilitated by categorization of information (Bruner, Goodnow, & Austin, 1956). Research on categorization of objects and events has shown that information is commonly stored and processed in relation to mental models, prototypes, or schemata (Mervis & Rosch, 1981; Smith & Medin, 1981). It is therefore natural and economical that one may judge an object, event, or procedure by the degree to which the observed stimulus set is representative of one's schema (Rumelhart, 1977, 1980; Rosch & Mervis, 1975) or mental models (Holland, Holyoak, Nisbett, & Thagard, 1986)

Mental models or schema are abstract knowledge structures in our mind. To describe knowledge structures, psychologists have used terms like frames (Minsky, 1975), scripts (Abelson, 1976), prototypes (Cantor & Mischel, 1977), nuclear scenes (Tomkins, 1979), and reference frames (Leyton, 1986), in addition to an earlier and more general term, schema (Bartlett, 1932; Piaget, 1936/1970; Rumelhart, 1977, 1980). Taylor and Crocker (1981) defined a schema as a

cognitive structure that consists in part the representation of some defined stimulus domain. The schema contains general knowledge about that domain, including the specification of the relationships among its attributes, as well as specific examples or instances of the stimulus domain. (p. 91)

In a judgment task, mental models, schema or knowledge structures are assumed to be utilized via the effortless, readily available simplifying heuristics. A number of such heuristics commonly used in judgmental tasks have been popularized by Kahneman and Tversky (1972, 1973). These include judgment by



representativeness, judgment by availability, and judgment by anchoring and adjusting. Kahneman and Tversky suggest that although these heuristics are economical and useful, nonetheless, they lead in certain circumstances to systematic errors in human judgment. Cooper (1981) and Feldman (1981) speculated the mediating role of the representativeness heuristic in performance judgment. Judging by representativeness involves an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor, or more generally, between an outcome and a mental model of some sort (Kahneman & Tversky, 1984).

Mental models facilitate cognitive processing, but also introduce bias in judgment. Schema give structure to experience, determine information encoding and retrieval, fill in missing data, and provide the basis for problem solving, planning and anticipating the future (Bruner, 1971; Hastie, 1981; Taylor & Crocker, 1981). Given these functions of schema, schematic processing or prototype matching facilitates judgment, but schematic processing could also lead to faulty inferences when the stimulus configuration is incongruent with one's schema. Taylor and Crocker (1981) cite evidence of faulty medical diagnosis (Elstein, Shulman, & Sprafka, 1978), prejudice among jury members (Davis, Spitzer, Nago, & Stasser, 1980), poor policy decisions (Janis & Mann, 1977), and belief in discredited theory (Kuhn, 1970) as induced by schemata. Although schema is widely used as an explanatory concept, Fiedler (1982) points out that attempts to provide verification by measuring schema are rare.

Similarly, simplifying heuristics cause systematic errors and inconsistencies in human judgment and decision making (Kahneman, Slovic, & Tversky, 1982).

Judgment of prototypicality is made through the use of judgmental heuristics like the representativeness heuristic (Tversky & Kahneman, 1974); the representativeness of a particular stimulus is judged by invoking a schema against which the stimulus configuration is compared (Taylor & Crocker, 1981). However, there is evidence that the representativeness heuristic makes people insensitive to prior information (unless it is causal in nature), leads to misconception of chance, misconception of the regression phenomena, and to an illusion of validity (Kahneman & Tversky, 1972, 1973). An error or bias in judgment may therefore result from a person becoming insensitive to variations in data due to his/her reliance on simplifying heuristics, prototype matching or schematic processing.

### **3. Judgment Task Environment**

Investigators have identified a number of contextual factors that affect strategy use in a judgmental task other than performance judgment. Most of the research has focused on stimulus features. Earlier research mostly relating to gambles, business, and medical judgment investigated cue inter-relationships, cue-response compatibility, set size, extremity of information, redundancy, inter-item consistency, and primacy and recency effects (Slovic & Lichtenstein, 1971). The results have generally demonstrated that presentation format of cues relates to variations in judgment strategies.

More recent research has examined other features of information which may influence judgment processes. Researchers have found that the amount of information or information load affects information search and integration

strategies in judgment tasks. For example, Payne (1976) manipulated the complexity of the judgment task by varying the number of alternatives and the number of cue dimensions. The results showed that the search for information decreased as the total amount of information increased, indicating a trade-off between cognitive load and the complexity of information. Similarly, Shaklee and Fischhoff (1982) found changes in judgment strategies as information load increased demands on memory, and Shaklee and Mims (1982) found a tendency to use simpler but less accurate strategies as memory demands increased.

The manner in which a judgment task is presented and the instructions given to subjects to perform the task affect judgment as well. Tversky and Kahneman (1981) manipulated information presentation. They presented information as negatively and positively framed, and found a strong effect of framing on subjects' inferences in probabilistic judgment. Crocker (1981) suggested that instructions to the subjects is one of the important factors influencing people's judgment of covariation in data.

In the majority of past research on information processing in judgment, the saliency of cues has been found to have a profound impact on cue utilization and integration strategies. For example, the anchoring and adjustment heuristic proceeds from the most salient dimension of the stimulus, and adjustment is made as additional dimensions are considered (Tversky & Kahneman, 1974). Although cue saliency has been an important explanatory variable, there has been little effort to define and manipulate it (Nisbett & Ross, 1980). Consequently, there is no theory addressing how cue saliency may be determined (Wallsten, 1980). However, in a study by Wallsten and Barton (1982), cue

saliency was manipulated by varying perceptual characteristics and the probability of occurrence. They found that despite the prominence of probabilities, subjects were also responsive to perceptual features of the stimuli. Perceptual manipulations intended to affect dimensional salience and processing order generally had the predicted effect, whereby subjects traded off perceptual and probabilistic features.

In the past, the effect of cues or stimulus dimensionality in research on judgment processes seems to have been mainly researched in terms of perceptual characteristics and probability of occurrence of cues, the factors which were assumed to induce differential salience (Wallsten & Barton, 1980; Slovic & Lichtenstein, 1971). However, stimulus information may depend on other factors as well. In perception research, one view is that information is in meaning, extent, time, frequency, and intensity (Kubovy, 1981). Another view is that information is in structure (Garner, 1974); the potential origins of structure are experience, constraints, statistics, analysis, and geometry (Cutting, 1987). Perspectives in social perception and judgment suggest that information is in the concreteness of the stimuli (Nisbett, Borgida, Crandall, & Reed, 1976) and personal implicit theories (Nisbett & Ross, 1980).

## **B. COGNITIVE VIEWS OF PERFORMANCE JUDGMENT**

The current emphasis in performance appraisal research is on the entire performance judgment process of which a rater's cognition is an important aspect. Attention is drawn to a rater's cognitive functioning in the formation of the rating response, with a view to examining the potential causes of biases and

errors in performance judgment. Several theoretical models of performance judgment have been derived from theories and research in cognitive and social cognitive psychology. Generally, it is suggested that performance appraisal be viewed in terms of person and social perception, cognitive categorization, and prototype matching. These perspectives are discussed in the following four sub-sections.

### **1. Cognitive Distortion in Performance Judgment**

One of the earliest process oriented approaches to performance evaluation was outlined by Borman (1978). He viewed performance appraisal as a three step process. The first step was observing work-related behaviors. The second step was evaluating the behaviors in terms of the effectiveness they represent, and the third step was mentally weighting the evaluations to arrive at a single rating on a performance dimension. Although the third step in Borman's model had direct implications for cognitive processing of performance information, he did not explicate processes in terms of a rater's cognitive functioning. He limited his discussion to suggesting that in arriving at a single rating, raters somehow combine the information from multiple dimensions.

Cooper (1981) elaborated Borman's (1978) model, and speculated how cognitive distortion may be introduced in performance judgment. Cognitive distortion refers to the phenomenon of observations being distorted in such a way that raters both lose and add information. The loss of information may occur due to the failure to retrieve observed information stored in memory, whereas addition of information may result from a rater's implicit theories of illusory

covariation of the rating dimensions. He suggested that the reliance on heuristics of judgment (Tversky & Kahneman, 1974) appear to be a factor causing cognitive distortions. However, he stated that because systematic research was lacking, many of the processes could only be hypothesized. Cooper (1981) interpreted the halo error as cognitive distortion, drawing upon implicit personality theory literature (Schneider, 1973), research on covariation judgment (Chapman & Chapman, 1969), and theorizing on people's construction of reality (Kelly, 1955). Addressing halo, Cooper (1981) felt, "Prospects for eliminating it remain slim until there is a better appreciation of the halo-reduction barriers represented by cognitive distortions" (p. 235).

## **2. Performance Appraisal as Person Perception**

Upon reviewing a large body of literature on performance rating, Landy and Farr (1980) conceptualized a coherent representation of the entire performance appraisal process. They proposed that performance rating be examined as a specific phenomenon of person perception, from the perspective of implicit personality theory, and Wherry's psychometric theory of rating (Wherry & Bartlett, 1982).

The major components of Landy and Farr's (1980) model are the context of appraisal and the rating process. The rating process is comprised of the cognitive strategies of the rater and the administrative appraisal system of the organization. In the main, the authors suggest that cognitive characteristics of the rater, the purpose of rating, and organizational variables such as the position being rated have significant bearing on performance judgment. In their model,

special emphasis is placed on raters' cognitive processing of performance information. The authors expressed the concern, "We must learn much more about the way in which potential raters observe, encode, store, retrieve, and record information if we hope to increase the validity of ratings" (p. 100).

Landy and Farr (1980) noted that cognitive characteristics of raters have not been investigated systematically. They proposed that cognitive complexity and the way a rater "treats" or mentally integrates several dimensions of information, could be a potential source of variance in rating judgments. They also stressed the central importance of the purpose of performance judgment.

### **3. Performance Appraisal as Prototype Matching**

An alternative process model of performance appraisal was proposed by Feldman (1981), and later elaborated by Ilgen and Feldman (1983). Their model is based on theories and literature on cognitive categorization and prototype matching (Rosch, 1978), person perception (Cantor & Mischel, 1977, 1979), and on the theory of automatic and controlled processing of information (Schneider & Shiffrin, 1977).

According to Ilgen and Feldman (1983), performance judgment is an outcome of matching an employee with the attributes of prototypes representing categories in a rater's mind. The assignment to a category is either automatic or consciously monitored. When an employee demonstrates behavior similar to the stored prototype, assignment to the category is via automatic processing; the matching process is accomplished automatically with little mental effort. When an

employee's behavior is unlike the prototype in some respect, controlled processing is involved because special cognitive effort and attention are required in the prototype matching process.

Hence, Feldman et al. suggested that the recall of information about an employee may be influenced by how the information was encoded. Haloed ratings result from recall of prototypical information, because distinct behaviors are treated as being similar when classified into a category. In trait rating, the traits are recalled together and covary with category. Thus, "the more prototypical the stimulus person the greater will be the halo effect" (Feldman, 1981, p. 140). Discussing halo, under-evaluation and over-evaluation, Feldman (1981) contended, "neither overt prejudice nor motivational biases are necessary to produce such results, which arise from the nature of the categorization processes itself" (p. 130).

Based on research on human cognition, Feldman (1981) suggested that categorization in performance appraisal may be affected by selective attention, rater expectancies, and memory distortions via the use of simplifying heuristics of judgment (Tversky & Kahneman, 1974). Additionally, Feldman et al. suggested that a rater's cognitive structure and personal construct system or cognitive complexity (Kelly, 1955) may also influence the categorization of ratees. As no direct evidence was available, Feldman (1981) concluded by recommending laboratory based investigations of ratee categorization processes, together with related psychometric and field research.



#### 4. Performance Appraisal as Social Perception

Another cognitive view of performance appraisal process has been presented by DeNisi, Cafferty, and Meglino (1984). Their model is based mainly on Wyer and Srull's (1981) model of category accessibility. DeNisi et al's model portrays performance appraisal as the product of a set of social cognitive operations. These operations include information acquisition, organization, retrieval from memory and the integration of information to form a judgmental rating. DeNisi et al. consider, "performance appraisal is an exercise in social perception and cognition embedded in an organizational context requiring both formal and implicit judgment" (p. 362).

DeNisi et al. (1984) claimed that the purpose for appraisal and the role of the rater as an active information seeker are two distinct features of their model. The purpose for appraisal may predispose the rater to select an internal frame of reference or a schema which guides information search and interpretation. They suggested that schema, the mentally stored prototype of a good worker, guides what information is sought and how that information is interpreted. Acquisition of job-relevant information is considered the primary input, which may also be affected by factors other than one's schema, such as time pressures operating on the rater and the nature of the rating instrument.

DeNisi et al. (1984) emphasized the need for an examination of a rater's information search and integration strategies, given different purposes for evaluation and different rating instruments. The role of a raters's memory accessibility is considered critical. A rater's cognitive complexity and field

dependence-independence dimension of cognitive style are also considered important influences in processing information related to performance.

In essence, the cognitive views of performance judgment discussed above are based largely on related theories and literature dealing with person perception and social cognition. As a result, these theoretical models reflect similar views, but Ilgen and Feldman (1983), and DeNisi et al. (1984) are more explicit in their statement of needed empirical research. These views tend to emphasize information acquisition. Although information acquisition is undoubtedly necessary, it does not complete the judgment process. A person "operates" on the information, that is, mentally weighs and integrates the information gathered, to reach a judgment (Anderson, 1981). Acquisition of information may be highly automatic and distorted by stimulus features alone, as happens in perceptual illusions. Comparatively, weighing and integrating information, that is actually *forming* the judgment is presumably deliberate, and involves controlled (Schneider & Shiffrin, 1977) and deeper processing of information (Craik & Lockhart, 1972).

A common thread in the process models of performance appraisal is the emphasis on raters' cognition (Copper, 1981; DeNisi, et al. 1984; Ilgen & Feldman, 1983; Landy & Farr, 1980). In these views, it is suggested that the purpose of performance judgment and raters' cognitive complexity bear influences on how rating judgments are formed. Although how raters process information is of central importance, these models have not addressed the importance of cues on which judgments may be based. However, researchers in other areas have shown that cue dimensionality is an important variable affecting information use and judgment strategies (cf. Part A.3 in this review).

## C. RESEARCH ON RATERS' COGNITION

Theorizing on cognitive processing of performance information has indicated a number of important areas for investigation in the performance appraisal process. The importance of appraisal purpose and raters' cognitive complexity has been repeatedly stressed. The empirical literature in these areas is reviewed in this section.

### 1. Cognitive Effect of Appraisal Purpose

A number of researchers have suggested the importance of purpose for judgment in the performance appraisal process (DeCotiis & Petit, 1978; DeNisi et al., 1984; Landy & Farr, 1980; Zedeck & Kafry, 1977). Earlier studies reviewed by Landy and Farr (1980) indicated that purpose for evaluation operates as a motivational variable. These studies reported that ratings required for administrative decisions were significantly less lenient than ratings done for the purpose of research studies. However, it is postulated that the purpose for performance evaluation may have effects beyond rater motivation (DeNisi, et al., 1984; Ilgen & Feldman, 1983; Landy & Farr, 1980). The purpose for which an appraisal is conducted may cue raters to search for and utilize certain types of information, and thus serve a cognitive function.

One of the first studies revealing a cognitive effect of appraisal purpose on rating judgments was conducted by Zedeck and Cascio (1982). These researchers examined the effect of appraisal purpose on rater accuracy, discriminability, and information utilization policy. They used policy capturing methodology and

operationalized information utilization in terms of the weighting (regression weights) of the various dimensions of information presented in rater vignettes of 33 supermarket checkers. Ratings on a seven point scale were required for decisions about merit pay increases, the need for development/training, and for retention of employees. Each rater's standard deviation of ratings was used in an ANOVA for testing the hypothesis on discriminability between ratees as a function of appraisal purpose and rater training. Only the effect of appraisal purpose was found significant. The group evaluating for merit pay raises weighted "skill in human relations" most heavily, and the groups evaluating for development and retention purposes relied equally on "organizational ability and bagging skill." The researchers interpreted this information utilization policy in terms of organizational and consumer perspectives because of the appropriateness of the dimensions weighted for different appraisal purposes.

Although Zedeck and Cascio (1982) presented some evidence on the cognitive effect of appraisal purpose on rating judgments, their study is limited to information utilization - they did not examine the effect of appraisal purpose on the use of information integration strategies. Furthermore, they did not proceed with a theoretical framework for interpreting information utilization as a function of appraisal purpose, a criticism that has also been echoed by others (Williams, DiNisi, Blencoe, & Cafferty, 1985). An alternative explanation for information utilization in their study may be offered in terms of the processing of personality information and job behaviors for different purposes. This alternative explanation is based on schematic processing, a theoretical perspective from which predictions could be made about the usage of information across different jobs, when ratings are done for different purposes.

A weak effect of appraisal purpose on psychometric qualities of ratings was reported by McIntyre, Smith, and Hassett (1984). A sample of undergraduates rated videotaped lectures acted by drama students. Ratings on teaching effectiveness were required for three purposes: research, course improvement, and hiring decisions. Subjects were instructed about the purpose of their rating before presenting the videotapes. A post-experimental questionnaire was used to check if the purpose of rating was appropriately perceived by the different groups. Because of unequal cell sizes and heterogeneity of variances for every dependent variable investigated, the researchers presented their results of the main effects at a conservatively adjusted (reduced degree of freedom) alpha level of .10, and concluded a weak effect of appraisal purpose on leniency and accuracy of ratings.

McIntyre et al. (1984) used the final ratings (the product) as the dependent variable to draw inferences about the effect of appraisal purpose on raters' cognition. This prevented the researchers from discussing any aspect of information processing in raters' cognition, which may have been influenced by different purposes of appraisal. Although 15 of their subjects were excluded from the analysis, the authors did not report if the excluded subjects were equally distributed across the three purpose conditions in the study. Comparing their own results with those of Zedeck and Cascio (1982), McIntyre et al. (1984) questioned whether the variable was a purely cognitive one, although their study did not directly address information utilization and integration. However, in discussing the results, McIntyre et al. argued that the perceived purpose of rating may have mainly an emotional effect. One of the groups in their study was rating for the purpose of research (their study), but rating for research is hardly ever a real function of performance evaluation.

A study that has more directly examined the effect of appraisal purpose on raters' cognitive processes was conducted by Williams, DeNisi, Blencoe, and Cafferty (1985) who examined how purpose influenced information acquisition and utilization. In the first experiment, they presented *consistency*, *distinctiveness*, and *consensus* information on eight hypothetical ratees, in light of Kelley's (1971) covariation principles in attribution theory. They investigated whether subjects used the three types of information differently, resulting in different judgments for different purposes. Analysis on mean ratings showed that appraisal purpose had a limited effect on consistency, distinctiveness, and consensus information. Yet, the researchers suggested that raters were "sensitive to covariation information". In their second experiment, they investigated raters' search for different types of information as a function of appraisal purpose. Subjects were required to request distinctiveness, consistency, consensus, and covariation information presented on a micro computer. A MANOVA on preferences (after arcsine transformation) across purposes did not show a significant main effect of appraisal purpose. For all three purposes, distinctiveness information was preferred most, followed by consensus, and consistency information. Discussing the pattern in information requested, the researchers concluded that appraisal purpose appears to serve a cognitive function affecting the type of information searched for, but perhaps not the use of that information, a conclusion which contradicts Zedeck and Cascio's (1982) finding discussed above.

Williams et al. (1985) did not study the interaction between rating purposes and the type of information presented to the subjects. Moreover, a possible logical flaw in their study was that they investigated information utilization followed by information search, but in cognition, information search most likely precedes

information utilization. Therefore, their conclusion that raters may search for but not use different types of information in rating for different purposes, suffers from the illogical order in which information search and utilization were investigated. They seem to make that conclusion on the analysis of information preferences in the second experiment. Preference and use of information may not be the same. Analyzing preferences and addressing information use is perhaps a great inferential leap on their part. Moreover, the fact that they did not seek interaction effects in their first experiment, and did not correlate information preferences with the ratings, leaves a gap in our understanding of the relative contribution of the three types of information to rating judgments. Nevertheless, in comparison to other studies on the topic, their theoretical stance in presenting different types of information is a step forward in studying the effects of purpose on raters' cognition.

The effect of appraisal purpose on accuracy in observing teacher behavior and evaluating teaching performance was investigated by Murphy, Balzer, Kellam, and Armstrong (1984). Ratings were required for research and for making personnel decisions. Forty five student subjects evaluated four videotaped lectures delivered by graduate students. Subjects were informed of the purpose of their rating, and were required to (a) rate each lecturer's performance on standard teacher evaluation forms, and (b) indicate the frequency with which twelve "well-defined" behaviors were observed in each lecturer. The investigators found that appraisal purpose did not affect the accuracy of performance ratings; nor did it affect the accuracy of observing the frequency of the critical behaviors. However, they found that appraisal purpose did influence the relationship between accuracy in observing teacher behavior and the accuracy in evaluating teaching

performance. This latter finding led the experimenters to speculate that appraisal purpose affects the way raters process information, without necessarily affecting the general level of ratings.

In the four studies reviewed above, the pattern of results and certain features of the study procedures are noteworthy. The two studies that did not find a clear and substantial effect of appraisal purpose had "rating for the purpose of research" (their studies) as one of the rating purposes (McIntyre, et al. 1984; Murphy, et al., 1984). Because rating for the purpose of research is hardly ever a true function of performance appraisal, it may not have invoked a particular schema or prototype to provide a basis for judgment. Prototypes and implicit theories may exist in terms of performance effectiveness but not in terms of appraisal for research functions. Therefore, it would have been difficult for raters to engage in prototype matching, and for the researchers to find a reasonably strong cognitive effect of appraisal purpose. Moreover, these studies also required appraisals on only two or four ratees, which did not allow a sampling of a sufficiently large number of occasions. Epstein (1980) has warned about the limitations of not aggregating a reasonable number of occasions in the study of human behavior.

With the exception of one experiment by Williams et al. (1985), and the study by Zedeck and Cascio (1982), the rest of the studies have mainly analyzed the final ratings (the product) and drawn inferences about the effect of purpose on raters' cognition. The effect of purpose on the processes in cognition, that is how information is mentally weighted, utilized, and integrated, has received little attention. Only one of the studies (Williams, et al., 1985) has presented



information using a theoretical rationale, although performance evaluation may not be a case of attribution of causality. And finally, none of the studies provided a wide enough scale for an unrestricted expression of subjective judgments. Researchers of information integration suggest the use of a scale with about 20 points for functional measurement of subjective judgment (Anderson, 1982).

## **2. Effect of Cue Dimensionality**

In past research on human judgment processes, cue dimensionality, mainly the saliency of cues, has been found to have an effect on cue integration strategies (Nisbett & Ross, 1980; Wallsten, 1980). In contrast, the importance of cue dimensionality appears to have been oversighted in theoretical perspectives on cognition in performance judgment (Cooper, 1981; DeNisi, et al., 1984; Ilgen & Feldman, 1983; Landy & Farr, 1980).

Researchers addressing rater cognition appear to have also overlooked the effect of cue dimensionality on information use and integration in processing rating judgments (Cardy & Kehoe, 1984; Murphy, et al, 1984; Murphy & Balzer, 1986; McIntyre, 1984; Zedeck & Cascio, 1982). One study that investigated the use of information of different types is that of Williams et al. (1985), provided that consistency and consensus information can be perceived as aspects of cue dimensions. However, their discussion is limited to suggesting that raters were sensitive to covariation information.

In performance judgment, trait and behavior are the two main dimensions of performance information (Wexley & Klimoski, 1984). Besides, trait and

behavior are naturally occurring dimensions of information in person perception (Cantor & Mischel, 1979; Nisbett & Ross, 1980). Although performance judgment is conceived of as an exercise in person perception (DeNisi, et al., 1984; Ilgen & Feldman, 1983; Landy & Farr, 1980), none of the studies on rater cognition reviewed here sought the effect of trait and behavior information on rating judgments. Furthermore, none of the studies sought the interactive effects of appraisal purpose and cue dimensionality on the subjective importance and actual use of information in the formation of rating judgments.

### **3. Influence of Cognitive Complexity**

Considerable literature has dealt with cognitive complexity as a variable which influences people's perceptions and evaluations of events. Vannoy (1965) states that although various writers have given somewhat different meanings to the construct, it has generally been postulated that some persons are prone to employ few dimensions when they perceive and evaluate stimuli, or are inclined to make only very gross discriminations among dimensions for meaning; other persons are believed to employ many dimensions, and/or to make fine discriminations among the dimensions they utilize.

Cognitive complexity is a construct that emerged from Kelly's (1955) theory of personal constructs, and is commonly defined as a "disposition to view the person-objects in one special environment in a complex or differentiated manner" (Vannoy, 1965, p. 385). A cognitively complex person has a relatively more differentiated system of dimensions for processing the behavior of others than a cognitively simple person (Bieri, Atkins, Briar, Leaman, Miller, & Tripodi, 1966).

Bieri et al state that cognitive complexity is the ability to discriminate *between* dimensions attributed to stimuli (i.e. differentiation) and the ability to discriminate *within* each dimension (i.e. articulation).

The construct of cognitive complexity has previously been examined as a moderater variable in studies of leadership behavior (Mitchell, 1970), team performance (Kennedy, 1971), and decision making (Menasco, 1976). Theorists of the performance appraisal process have also emphasized the role of a rater's cognitive complexity (DeNisi, et al., 1984; Ilgen & Feldman, 1983; Landy & Farr, 1980). Raters' cognitive complexity may describe the way in which they organize and integrate their thoughts, and reflect the relative number of dimensions they use to describe what they perceive.

The effect of cognitive complexity on performance appraisal was first found by Schneier (1977). Schneier defined cognitive complexity as "the degree to which a person possesses the ability to perceive behavior in a multidimensional manner" (p. 541), and measured the variable using the Role Construct Repertory (REP) grid. In his study, cognitively complex raters demonstrated less halo, were less lenient, and used a wider range on behaviorally anchored rating scales (BARS) than did the cognitively simple raters. On these findings, Schneier (1977) suggested that to the degree there is compatibility between a rater's cognitive complexity and the cognitive demands of the appraisal process, there will be an increase in the psychometric quality of the resultant ratings.

Following Schneier's (1977) findings, a number of reviewers suggested that cognitive complexity of raters may relate to effective performance appraisal

(DeCotiis & Petit, 1978; Dunnette & Borman, 1979; Jacobs, Kafry, & Zedeck, 1980). In the cognitive process models discussed earlier in this review (Part B), Feldman (1981), Landy and Farr (1981), and DeNisi et al. (1984) have also pointed out the importance of a rater's cognitive complexity in making performance judgments. As a result, several researchers have empirically tested the relationship of cognitive complexity to performance rating effectiveness (Bernardin, Cardy, & Carlyle, 1982; Lahey & Saal, 1981; Sauser & Pond, 1981), but surprisingly, Schneier's findings have not been confirmed in any of these investigations.

Lahey and Saal (1981) investigated the cognitive compatibility hypothesis using three different cognitive complexity measures and four different rating scales. Cognitive complexity of undergraduate students was assessed with a REP grid, factor analysis of the REP grid data, and a sorting task. Performance ratings were obtained on a seven point BARS, three point mixed standard rating scales, seven point graphic rating scales, and simple "alternate" three point rating scales. No differences in leniency, halo, or range restriction emerged either as a function of cognitive complexity, or from the interaction of cognitive complexity with scale format.

As Lahey and Saal (1981) used multimethod assessments of cognitive complexity and rating properties, they considered their study a comprehensive test of the cognitive compatibility hypothesis. Nevertheless, some procedures in their study may have reduced the chances of obtaining the expected results. In all their analyses, they used transformed ratings as data points. The ratings on the seven point scales were transformed to three points in order to equate the

metric for the repeated measures ANOVA. For example, in the analysis of leniency effect, a composite rating was obtained first by transforming seven point scales to three and then by calculating the mean across dimensions. The transformations would have reduced the variability in the data, which perhaps diminished the effect of cognitive complexity. Furthermore, they assessed halo by calculating the standard deviation of ratings across the rating dimensions for each rater-ratee combination, which is an inadequate measure of halo according to a recent Monte Carlo study by Pulakos, Schmitt, and Ostroff (1986).

Bernardin, Cardy, and Carlyle (1982) re-examined the role of cognitive complexity as a predictor of appraisal effectiveness in a series of experiments. In their first experiment, 28 undergraduates rated three of their psychology instructors on two separate scales. One scale was a BARS consisting of five performance dimensions. The same dimensions were represented on a second three point graphic rating scale. Cognitive complexity was measured by a REP grid, and halo was indexed by the standard deviation of each rater's ratings on five dimensions for each ratee. The data on BARS were transformed to a three point scale to examine rating errors as a function of cognitive complexity and scale format. There was no significant effect of complexity. The analysis procedure and the results were similar in the second experiment, in which the effect of cognitive complexity was examined on the accuracy of ratings given to two hypothetical instructors. In the third experiment, 31 police sergeants evaluated two patrol officers from a pool of 65 (selection criteria is not stated) on 11 dimensions using two rating scales. The analysis and the results were the same as in the first two experiments. Based on the results of previous studies as well as their own, Bernardin et al. concluded, "the plethora of null findings

certainly casts doubt on at least the generalizability of the cognitive compatibility theory, if not also its validity."

The null findings in the study by Bernardin et al. (1982) resulted possibly from some of the procedures they adopted in their experiments. The transformation of ratings from a ten point to a three point scale might have reduced the variability in the data, and thereby, perhaps diminished the effect of cognitive complexity. Furthermore, the fact that each sergeant appraised only two of the 65 patrol officers means that the raters' evaluations were being compared, although they evaluated different officers (random selection cannot be assumed because the selection criteria is not stated).

The discussion above suggests that previous research on the influence of cognitive complexity on performance judgment suffered from a number of methodological weaknesses. One of the weaknesses was the lack of face validity of the REP grid because all of the studies reviewed above used the REP measure with the same *elements* and *constructs* on the grid as introduced by Bieri et al. (1966) in a study of clinical judgment. In performance judgment research, these elements (e.g. father) and constructs (e.g. shy-outgoing) may lack face validity. Another common limitation in the studies considered above is in the assessment of halo. The standard deviation as an index of halo is now known to be problematic (Pulakos, et al., 1986). Additionally, the transformation of ratings probably resulted in a loss of variability, and thus, diminished the effect of cognitive complexity. Because transformation of data reduces nonadditivity, it might have masked interaction effects as well (Winer, 1971). And finally, the concern with psychometric properties of ratings, has perhaps prevented the study

of the information integration strategies which complex and simple raters use in performance judgment.

However, some support relating to the cognitive compatibility hypothesis has come from other areas. Researchers in education, for example, have used "conceptual level", a concept similar to the notion of cognitive complexity, to study supervisor-supervisee interactions. Thies-Sprinthall (1980) reported that supervisors identified as conceptually differentiated were more flexible, responsive, and recognized a wider range of teaching behaviors than the supervisors identified at a lower level of conceptual development. Likewise, Grimmatt (1984) found that "abstract" supervisors, that is, supervisors at a higher level of conceptual development, engaged in conjoint appraisal of teaching-learning episodes with the supervisees, asked more open ended questions, and elicited ideas from the supervisees more than the "concrete" supervisors.

#### **D. EVALUATION OF TEACHING**

Thorndike (1920) initially characterized the halo error in a study on evaluation of teaching. Even today, appraisal of teaching performance (like appraisal of performance in other occupational settings) is jeopardized by problems in rating. The question of reliability and validity of ratings of teaching performance at elementary and secondary school levels continues to be a matter of great concern (Evertson & Holley, 1981; Hawley, 1982; Medley, 1982). Although student evaluation of teaching at colleges and universities has been widely endorsed in recent years (Centra, 1979), similar problems are present at this level as well (Cohen, 1980, 1981; Marsh, 1984; Marsh & Overall, 1980).

Current theorizing and research on teacher evaluation appears still to be dominated by a concern with instrument development (Peterson, Micceri, & Smith, 1985), and with the problems of determining criteria and techniques for evaluation (McGreal, 1983; Millman, 1981; Stiggins & Bridgeford, 1985). Relatively little attention is given to the rater's ability to draw inferences and to the cognitive processes that may underlie performance judgments. Not surprisingly, it has been pointed out that the potential contribution of theories from cognitive psychology to research on teaching and teacher education has not yet been widely realized (Shavelson, 1985).

Investigation of bias in student ratings of instructors has focused mainly on the "Dr. Fox effect" or "educational seduction". Researchers of the Dr. Fox effect suggested that expressive or enthusiastic behavior was as important as the content taught in arousing reactions toward instructors, and questioned the validity of student ratings (Naftulin, Ware, & Donnelly, 1973).

One of the Dr. Fox studies investigated the effect of the purpose of evaluation (Meier & Feldhusen, 1979). The stated purpose of evaluation had no effect on any of the rating measures; nor did purpose interact with expressiveness and lecture content. Expressiveness had a significant effect on all five rating factors but not on student achievement. The authors concluded that "an expressive lecturer can generate a halo effect which influences others' evaluation of him" (p. 343). Similar effect of expressiveness or enthusiasm was observed in almost all of the Dr. Fox studies. In a meta-analysis by Abrami, Leventhal, and Perry, (1982), the proportion of rating variance accounted for by expressiveness had a weighted mean of .293 across twelve studies.



The early Dr. Fox studies have been criticized for lack of external as well as internal validity (Frey, 1978). In their meta-analysis, Abrami et al. (1982) found that although there were methodological flaws in several of the studies and disagreement on the issue of validity of student ratings, expressiveness typically had a large impact on ratings, and lecture content typically had a large impact on student achievement. Even if the Dr. Fox studies are considered methodologically sound, these studies address bias as affective phenomena, and do not consider the validity of student ratings from a cognitive information processing perspective.

The cognitive processing of information in evaluation of teaching may vary depending on one's conception of teaching. According to Shulman (1986), "the normative conception of teacher effectiveness is derived from one's theory or ideology and requires a judgment of correspondence between the conception and the exemplar" (p. 28). It can be assumed from Shulman's point of view that those who judge teaching likely have exemplars, prototypes, or mental models of a good teacher and good teaching. The good teacher schema may be in terms of traits and behaviors because desirable personal traits and use of effective methods are important characteristics of good teaching (Medley, 1979). In evaluation of instructors researchers have found instructor enthusiasm, sociability, warmth, resourcefulness, and leadership are important traits (Cohen, 1981; Kulik & McKeachie, 1975; Marsh, 1983), and planning, presentation clarity, grading, communication, and research activity are important behaviors (Cohen, 1981; Frey, Leonard, & Beatty, 1975, Marsh, 1983, 1984).

## E. SUMMARY

This review started with a discussion on how the cognitive limitations of the mind and certain features of the task affect human judgment. The literature indicates that to cope with the limitations of attention and memory, we use schemata and simplifying heuristics. However, schematic processing may introduce bias and inconsistencies in our judgment. Schemata are assumed to exist and operate as theorized, and although schema is widely used as an explanatory concept, little effort has been devoted to its measurement. Certain features of the task such as the amount of information, cue saliency, and the manner in which information is presented, also affect how we form judgments. Researchers have found that cue saliency has a strong effect on judgment strategies. Cue saliency may be determined by perceptual characteristics, structure and content of information, and personal implicit theories.

The cognitive perspectives on performance judgment were reviewed as well. These views indicate a number of variables that may influence how raters process information. Two of the variables suggested as having important influences on how raters form rating judgments are cognitive complexity and the purpose for appraisal. However, research findings concerning the effect of purpose on raters' cognition are limited and equivocal. Likewise, the findings on the effect of cognitive complexity on the halo effect, are contradictory.

Studies which sought the effect of purpose on raters' cognition were examined, but no clear results emerged. The findings are mixed, and fail to clarify how the formation of rating judgments are affected by purpose.

Researchers have mainly analyzed the ratings and not the processes that lead to a rating response, which include information utilization and integration. One study that investigated information utilization lacked a theoretical basis for interpreting the results; another study that addressed information utilization did not examine the interaction between purpose and information utilization. Moreover, the purpose conditions used in some of the studies were unrealistic, and the number of ratees evaluated by the subjects quite small for the study of raters' judgment strategies.

Studies that investigated the effects of cognitive complexity on performance judgment, were also examined. The conclusions of most of these studies are pessimistic, particularly in regard to the effect of cognitive complexity on psychometric qualities of ratings, but the pessimistic conclusions may be a result of assessing halo by inappropriate means. Furthermore, not only some of these studies have methodological flaws, they have focused exclusively on psychometric properties of ratings. The studies reviewed here did not consider the effect of cognitive complexity on the use of varied information combination strategies.

Researchers of human judgment processes have found that cue dimensionality has a strong effect on information utilization and integration strategies. The literature reviewed indicates an absence of a clear theory on what determines cue saliency. The studies reviewed here did not seek the effect of cues on raters' cognition. Nor is the importance of cue dimensionality addressed in the prevalent cognitive perspectives on performance appraisal.

Consequently, there are a number of matters that need to be resolved. We need to study the effect of appraisal purpose on cognitive processes that lead to

a rating judgment. Specifically, whether purpose affects information valuation, utilization, and integration needs to be examined, so that the effect of purpose on raters' cognition could be clarified. Further, we need to determine whether purpose and cue dimensionality conjointly influence judgment processes, so that what affects cue saliency in performance judgment could be identified. Research is also needed to study the influence of cognitive complexity on the use of information integration strategies, so that we learn more about individual variables that affect performance judgment. Additionally, the relationship between cognitive complexity and halo needs to be explored with halo assessed by correlational means. And finally, given the importance of schemata in human judgment, procedures to measure schema at the individual level need to be explored as well.

The hypotheses and methodology of the study are presented in the next chapter.

### III. HYPOTHESES, QUESTIONS, AND METHOD

This chapter outlines the hypotheses, exploratory questions, and research methodology. In the first part (A) is the rationale for the hypotheses and the methodology. The hypotheses and exploratory questions are presented in the second part (B). The design and data collection procedures are described in the third part (C).

#### A. RATIONALE FOR HYPOTHESES AND METHOD

It has been theorized that appraisal purpose has an effect on raters' cognition (DeNisi, et al., 1984; Landy & Farr, 1980; Zedeck & Cascio, 1982). In the jargon of cognitive psychology, the purpose for appraisal becomes a *priming stimulus*. A priming stimulus *activates* knowledge of a category or schemata, and hence, facilitates processing of the input (Loftus & Loftus, 1974). Many theorists discuss this phenomena in terms of *spreading activation*. Activation of a concept by priming, makes that concept and related knowledge in memory more accessible for processing the input (Collins & Loftus, 1975; Ortony, 1978). Thus, the perceived purpose of appraisal may activate prototypes, mental models, or generally the good worker schema, and thereby, provide *cognitive readiness* (Bruner, 1957) for the interpretation and utilization of performance information.

The different conceptions of the term schema could be classified into broad categories. Hastie's (1981) classifications include central tendency, procedural, and template schemata. Taylor and Crocker (1981) suggest that schemata are of three types: person schema, event schema, and role schema. Relevant to

performance judgment seem to be the notions of role and person schemata. Role schema include prototypes or mental models of people in terms of the roles or behavior in particular occupations like professor or fireman (Taylor & Crocker, 1981). Person schema consist of prototypes of people in terms of traits (Cantor & Mischel, 1977). Appraisal content also comprises trait and behavior information (Wexley & Klimoski, 1984). Good teaching, particularly, depends upon desirable traits and effective methods or role behaviors (Medley, 1979). Therefore, trait and behavior cues in the evaluation of college instructors may interact in the formation of rating judgments for different purposes.

The formation of a rating judgment involves mental transformaion of information through the processes of *valuation*, and *integration*. This conceptualization is based on the information integration theory proposed by Anderson (1981). Valuation involves the process of extracting salient and pertinent information from stimuli in working memory, and determining their weighting. Valuation is subjective, and therefore, directly susceptible to the influence of one's mental models or schemata. In performance judgment, the schema of a good worker - a combination of role and person schemata - may have an influence on the valuation of information, which may affect the subjective importance and utilization of performance information. The concept of integration refers to thought processes that combine the differentially weighted stimuli according to some rule to produce a response. The rules of combining information into a response, when modeled mathematically, reflect the information integration strategies; the weightings of cues in the rules reflect the utilization of information. Thus, in performance judgment, a rater may obtain information through direct observation of a person in action, retrieve information from

memory, or use observations recorded by others, but the information obtained is subjectively weighted and transformed into a rating response.

However, there is some controversy as to whether people can report the importance of information in their judgment, which may reflect their subjective weighting policy. Nisbett and Wilson (1977) have suggested that people tell more than they know, which casts doubt on people's ability to report, retrospectively, the importance of information in their own judgment. Further, in social judgment research, Brehmer (1976) found that although individuals generally weighted cues fairly accurately, they failed to apply this knowledge consistently. Similar findings have been reported by others as well (Slovic & Lichtenstein, 1971; Schmitt & Hunter, 1977). On the other hand, Ericson and Simon (1978), and Surber (1985) have provided evidence to the contrary. Therefore, it would be interesting to find out whether raters in performance judgment show consistency in expressing the subjective importance of cues and the actual utilization of similar cues.

Furthermore, cue integration strategies have rarely been studied in relation to developmental aspects of an individual (Pitz & Sachs, 1984). A developmental construct that might affect the use of varied information integration strategies seems to be a person's cognitive complexity. Theoretically, cognitive complexity refers to a person's disposition to view the task environment in a complex or differentiated manner (Vannoy, 1965). A cognitively complex person has a relatively more differentiated system of dimensions for processing information, than a relatively less complex person (Bieri, et al., 1966). Therefore, a cognitively complex person may employ more strategies of integrating information than a cognitively simple person.

Researchers have studied different kinds of information utilization and integration models (Anderson, 1981, Einhorn, 1970). These models fall into two basic categories, compensatory and noncompensatory models (Hogarth, 1980). Compensatory models entail the linear additive and averaging strategies involving trade-offs between dimensions of information. The integration of information in compensatory models is not interactive. Noncompensatory models on the other hand, refer to integration strategies that involve interactive use of cues (Billings & Marcus, 1983; Einhorn, 1970, 1971; Hogarth, 1980). In a noncompensatory strategy, such as conjunctive or elimination-by-aspects (Tversky, 1972), the amount of information utilized per alternative is variable because a low score on one dimension may not be compensated by a high score on another, causing the elimination of certain information and alternatives.

The strategies of information utilization and integration are traditionally assessed by regression models or policy capturing analysis (e.g. Borko & Cadwell, 1982; Cadwell & Jenkins, 1986; Einhorn, 1970, Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Norman, 1986; Slovic & Lichtenstein, 1971; Zedeck & Cascio, 1982). However, models of judgment may not be isomorphic with the processes they represent (Hoffman, 1960). Einhorn et al. (1979) compared process-tracing protocols and linear regression models. They concluded that both models capture the same underlying processes but at different levels of generality, and argued that linear regression models do capture the interactive and contingent processes. Moreover, the success of the linear regression models in a wide variety of judgment tasks strongly indicates that some fundamental characteristics of human judgment is captured (Goldberg, 1968).



## B. HYPOTHESES AND EXPLORATORY QUESTIONS

In light of the discussion above and the discussion in the previous two chapters, this study attempted to test hypotheses concerning (a) the effects of *purpose* and *cue dimensionality* on **subjective importance** and **utilization** of information, (b) the **consistency** between *subjective importance* and *utilization* of information, and (c) the effect of *purpose* and *cognitive complexity* on the **use of information integration strategies** in performance judgment. The research hypotheses are stated in sub-sections one to four below, and the exploratory questions are presented in the fifth sub-section.

### 1. Importance of Information

**Hypothesis 1.A:** Appraisal purpose will affect subjective importance ratings of trait and behavior information in performance judgment.

**Hypothesis 1.B:** Cue dimensionality will affect subjective importance ratings of trait and behavior information in performance judgment.

**Hypothesis 1.C:** Appraisal purpose and cue dimensionality will conjointly affect the subjective importance ratings of trait and behavior information in performance judgment.

## 2. Utilization of Information

**Hypothesis 2.A:** Appraisal purpose will affect utilization of trait and behavior information in performance judgment.

**Hypothesis 2.B:** Cue dimensionality will affect utilization of trait and behavior information in performance judgment.

**Hypothesis 2.C:** Appraisal purpose and cue dimensionality will conjointly affect the utilization of trait and behavior information in performance judgment.

## 3. Information Importance and Utilization Consistency

**Hypothesis 3:** Cue utilization in performance judgment will be consistent with subjective importance of cue dimensions.

## 4. Information Integration

**Hypothesis 4:** In comparison to formative judgment, in summative judgment raters will combine cue dimensions using a noncompensatory strategy in addition to a compensatory strategy.

**Hypothesis 5:** In comparison to cognitively simple raters, cognitively complex raters will combine cue dimensions using a noncompensatory strategy in addition to a compensatory strategy.

## 5. Exploratory Questions

The current study also sought answers for two questions of exploratory interest. These questions arise from the following two points in the literature reviewed in chapter two. First, although schema has been used as an important explanatory concept in several studies, attempts to quantitatively measure it are lacking (Fiedler, 1982). Second, the use of standard deviations in assessing halo is inappropriate (Pulakos, et al., 1986). Hence, the questions of exploratory interest were as follows:

1. Can a person's good instructor schema profile be measured quantitatively?
2. What is the effect of cognitive complexity on halo, when halo is measured by correlational techniques?

## C. METHODOLOGY

### 1. Subjects

Seventy students enrolled in the Faculty of Education programs at The University of British Columbia voluntarily served as subjects in this study. The sample of seventy was considered sufficiently large in order to detect medium to large effects of the independent variables. A consideration in choosing the sample was the familiarity with the judgment task. Students in the education programs at this university are required to evaluate their instructors, and were assumed to comprehend the performance judgment task as intended. They were also assumed to have sufficient knowledge of teaching to be able to make judgments on instructors. Six subjects in the sample were graduate students, 23 were in the

fourth year, and 42 were in the third year of their teacher education programs. There were 11 males and 69 females.

## 2. Instruments

Five measures were administered in this study: two questionnaires, two performance rating tasks, and a Role Construct Repertory grid. One of the questionnaires was a self-report measure of the importance subjects attached to cues or dimensions of information related to the performance of a university instructor. The other questionnaire was a measure of subjects' good instructor schema. Performance rating Task A contained 27 hypothetical ratee profiles; the other Task B was a single vignette described in sentences. The reliabilites of these instruments in the present study are reported in the next chapter. All of these measures are described below in separate sections, and included in the Appendix.

### *a. Importance of Information Measure*

Dimensionality of cues was reflected in trait and behavior items of information because appraisal content mainly comprises trait and role information (Wexley & Klimoski, 1984). The subjective importance individuals attached to traits and behaviors related to performance was measured using a questionnaire that listed ten items of information. Five of these items concerned traits and five concerned teaching behaviors. The trait and behavior items were listed alternatively.

The selection of the ten cues included in the instrument was based on their importance as identified in the instructor evaluation literature (Cohen, 1981; Frey, Leonard, & Beatty, 1975; Hildesbrand, Wilson, & Dienst, 1971; Kulik & McKeachie, 1975; Marsh, 1983, 1984). These researchers have commonly found that most important instructor traits are enthusiasm, sociability, warmth, resourcefulness, and leadership; most important behaviors are planning, presentation clarity, grading, communication, and research activity. The Importance of Information Measure is included in Appendix B.

Subjects rated the importance of each item of information on separate seven point interval scales. Three points on the scales were anchored as follows: 1 = least important, 4 = important, and 7 = most important type of information. The mean ratings on trait and behavior items were used in the analyses as measures of subjective importance of trait and behavior dimensions.

#### *b. Rating Judgment Task A*

A methodological limitation in a majority of previous studies was the use of a small number of ratees, usually two to eight. In the present study, performance judgment Task A consisted of 27 hypothetical ratee profiles on four dimensions at three levels. Hypothetical ratee profiles limited the amount of detail that could be included, but were judged to be suitable because it is a procedure that allows control over the amount of information which was essential in this study. Moreover, the use of hypothetical profiles were justified because in a real situation students normally evaluate on general impressions and not on specific details (cf. Cadwell & Jenkins, 1985).

Keeping in mind the limitations in human mental capacities (Chapter 2, Part A.1), it was reasoned that including four dimensions in the ratee profiles would not result in an information overload, thus, allowing thoughtful rating judgments. Additionally, a greater number of dimensions would have resulted in a huge number of possible profiles, and the cues in a fractional replicate would have occurred at a smaller number of times at each of the chosen levels, which could have adversely affected the impact of the cues on raters evaluations.

Each profile presented information on the two most salient traits and the two most salient behaviors from the ten included in the *Importance of Information Measure*. The traits were enthusiasm and resourcefulness; and the behaviors were presentation clarity and grading-marking. These traits and behaviors were chosen because of their popularity and explanatory power (factor loadings) of these dimensions in the existing evaluation instruments which have been developed and construct validated through research (Frey, Leonard, & Beatty, 1975; Hildesbrand, Wilson, & Dienst, 1971; Kulik & McKeachie, 1975; Marsh, 1983).

The four information dimensions or cues, each at three levels (above average, average, and below average) make 81 different combinations of ratee profiles ( $3^4$ ). To expect the participants to judge 81 profiles and to respond to four other measures would have been unrealistic in terms of their time, concentration, and motivation. Therefore, to maintain subjects' concentration for thoughtful judgments, they were required to rate only 27 of all possible profiles.

The 27 ratee profiles presented to the subjects were obtained by using a fractional replicate procedure (Winer, 1971). Choosing a one-third fractional

replicate ensured that each dimension was expressed as either above average, average, or below average the same number of times. In the 27 ratee profiles, each performance dimension was expressed nine times at each of the three levels. Hence, each dimension or cue had the same chance of being used in the formation of the rating judgments.

The 27 replicates were obtained from the fractional factorial tables developed by Conner and Zelen (1959). These tables have been prepared so that: (a) no main effects are aliased with other main effects, or aliased with two factor interactions; (b) as few two factor interactions as possible are aliased with other two factor interactions; (c) two factor interactions which are only aliased with higher order interactions are termed measurable (Connor & Zelen, 1959, p. 2). In the one-third replication chosen for the present study, six first-order interactions (AB, AC, AD, BC, BD, CD) were measurable.

The arrangement of the dimensions within a profile was on a Latin squares pattern. This was done to ensure an even distribution of any recency or primacy effects across all dimensions. The set of 27 ratee profiles was collated into booklets following Latin squares rotation as well. This was necessary because there is some evidence that the serial position of ratees has an effect on performance ratings, although no general pattern has emerged to date (Landy & Farr, 1980). The rotation of ratee profiles balanced out rater fatigue and "observer drift" as well, if there was any. Observer drift refers to shifting criteria at different points in the rating task, which can cause carry over effects, leniency or stringency in ratings. The factorial combination and Latin squares arrangement pattern are included in Appendix C.

A methodological limitation in previous studies was the use of response scales that do not allow sufficient variability, and thereby, restrict the expression of subjective judgments. In this study, an eighteen point scale was provided for a numerical rating judgment on each ratee profile, because researchers studying information integration suggest the use of a scale with about 20 points (Anderson, 1982). The rating response or judgment was required in terms of suitability for promotion (summative condition) and in terms of feedback on the quality of performance (formative† condition). As the same set of profiles was used for two different appraisal purposes, neutral anchors (very poor, average, outstanding) were used to mark three points on the scale.

After reading the purpose of appraisal, which was intended as a priming stimulus, subjects rated each of the profiles in one session. They indicated their rating judgment by marking a point on the scale for each profile. Each profile was presented on a separate page in order to inhibit comparative ratings. Care was taken to code each profile, so that the factorial combination and arrangement of its dimensions could be traced for the purposes of data analyses. The instructions and two typical profiles, one for each appraisal purpose, are included in Appendix C.

### *c. Cognitive Complexity Measure*

Cognitive complexity was measured using a version of the Role Construct Repertory (REP) grid introduced by Kelly (1955) and revised by Bieri et al.

---

† The use of the term *formative* should not imply specific recommendations for improvement (as it may connote in teacher evaluation literature), because for the purposes of this study, the ratings required were an expression of judgment only.



(1966). The REP grid measure has been found to be valid and reliable across a number of samples. For example, test-retest reliabilities of .68, .86 and .82 have been reported in other studies (Schneier, 1979).

However, unlike previous studies, the same *elements (roles)* and *constructs* as in Bieri et al. (1966) were not used in the present study because these lack face validity in a study on performance judgment. Instead, following the guidelines provided by Easterby-Smith (1980, 1982), more relevant roles (e.g. male teacher) and constructs (e.g. decisive-indecisive) were chosen to increase the face validity of the measure. The constructs chosen were those which reflect important characteristics of a teacher (cf. Hawley, 1982; Medley, 1982, Millman, 1981). The modification of roles and constructs is not likely to have affected the reliability of the instrument because such modification provides an alternate form of the measure, a procedure which has been previously used to establish the reliability of the REP grid measure of cognitive complexity (Schneier, 1979). However, the reliability of the REP measure in the present study is reported in the next chapter.

The 10 by 10 (roles by constructs) grid listed roles horizontally on the top and the constructs vertically on the right. The constructs were calibrated on a six point bipolar scale. The subjects decided the degree to which each construct applied to each person inserted for the roles. In this measure, the use of many degrees of constructs in describing each role person as opposed to using one or a few degrees, indicates complexity (Bieri, et al., 1966). Subjects were asked to provide a rating in each cell of the grid. The grid was scored for the number of different judgments made, and because the scoring was reversed (as is usually

done) the smaller number of different judgments made represented the most complexity. The REP grid used in this study is included in Appendix D.

*d. Good Instructor Schema Measure*

This measure was used for the puposes of exploring the measurement of a good instructor schema. A technique for quantitatively measuring stereotypes at an individual level has been offered by McCauley and Stitt (1978,) and McCauley, Stitt, and Segal (1980). This procedure, based on Bayesian probability estimation, was adapted for the measurement of a subject's good instructor schema.

In drawing generalizations about classes of people, Bayes rule states:  $p(\text{characteristic } B/\text{group } A) = p(\text{characteristic } B) \text{ times the } p(\text{group } A/\text{characteristic } B) \text{ divided by } p(\text{group } A)$ . In other words,  $p(B/A) = p(B) \times LR$ , where LR is the likelihood ratio  $p(A/B)/p(A)$ . LR is called the "diagnostic ratio" (DR), because it is the measure of the degree by which the occurrence of *A* revises the probability of *B*. When the DR is 1.0, the occurrence of *A* describes nothing about the probability of *B*, that is, the occurrence of *A* has no diagnostic value.

Using a series of studies, McCauley et al. (1978, 1980) concluded that a diagnostic ratio of greater or less than 1.0 indicates a stereotype quantitatively. The departure of the diagnostic ratio from 1.0 indicates the strength of an attribute of the stereotype. These authors claim that their Bayesian technique is the first quantitative measure of stereotypes or schema, and it is superior to other existing group measures. A definite advantage of this procedure is that the

diagnostic ratio allows schema measurement at an individual level in quantitative terms.

Several other studies have since measured stereotypes via a probability estimating procedure. Rasinski, Crocker, and Hastie (1985) constructed a Bayesian normative criterion based on subjects' own stereotypes in a study analyzing social perceiver's use of subjective probabilities. McCauley, Durham, Copley, and Johnson (1985) used probability estimates to measure stereotypes quantitatively in a study analyzing the impact of personal experience on population predictions.

The procedure presented by McCauley et al. (1978, 1980) for measuring stereotypes was adapted for measuring subjects' good instructor schema in the present study. Probability estimates were obtained on eight related and two unrelated attributes of a university instructor. For each attribute, subjects estimated four probabilities. The following is an example of the probability statements for the behavior "presents the subject matter with clarity".

1.  $p(\text{behavior})$ :- What percentage of instructors present the subject matter with clarity?
2.  $p(\text{behavior}/\text{group})$ :- What percentage of **good** instructors present the subject matter with clarity?
3.  $p(\text{group}/\text{behavior})$ :- What percentage of **all** instructors who present the subject matter with clarity **are** good instructors?
4.  $p(\text{group})$ :- What percentage of all instructors are **good** instructors?

Each of the four parts of the probability estimation questionnaire was presented on a separate page. Subjects were asked to provide their best

estimates, as they had no way of knowing the exact answers. The first two probabilities were required to encourage subjects to engage in Bayesian reasoning. The second two percentage figures were used to compute a diagnostic ratio for each attribute, which reflected the saliency of an attribute in relation to the others. The diagnostic ratios were computed by dividing  $p(\text{group/behavior})$  by  $p(\text{group})$ . The good instructor schema measure is included in Appendix E.

*e. Rating Judgment Task B*

Performance rating Task B was presented in order to explore the effect of cognitive complexity on halo in ratings. It has been suggested that a methodological weakness in the majority of previous studies is the inappropriate measure of halo (Pulakos, et al., 1986). Previous studies relied on the standard deviation of dimensional ratings across ratees, but Pulakos, et al. (1986) recommended the use of correlation between dimensional ratings.

In this study, halo was assessed by correlation techniques but in a modified way. Rating Task B consisted of a description of an instructor in a number of sentences. Although the rating scale included eight dimensions, information related to performance was supplied on four dimensions only. In other words, data were supplied for four dimensions: preparation, presentation, sociability, and dependability; and deliberately withheld for the other four dimensions: grading-marking, enthusiasm, communication, and scholarship. This increased the possibility of halo in the ratings. Rating Task B is included in Appendix F.

The subjects were neither informed of the information withheld, nor

instructed to provide a rating on all dimensions. They rated each dimension on a seven point scale. From the ten decomposed ratings, two ratings were computed for each subject: the average rating on the dimensions with missing information - rating "x," and the average rating on the dimensions with information included in the vignette - rating "y." The two ratings (x,y) for each subject in the cognitively complex and simple rater groups, were used to compute a correlation reflecting the degree of halo in the ratings of complex and simple raters.

### 3. Experimental Design and Variables

In testing the hypotheses, the present study had three independent variables. These were **appraisal purpose** (formative and summative conditions), **cue dimensionality** (trait and behavior), and **cognitive complexity** (complex and simple). The dependent variables were subjective **importance ratings**, the **utilization** (regression weights) of trait and behavior information, and compensatory and noncompensatory **information integration strategies**.

In the exploratory analysis concerning halo in ratings, cognitive complexity was the independent variable. For schema measurement, the dependent measures were diagnostic ratios in the good instructor schema profiles.

Equal numbers of participants were assigned to the two rating purpose conditions by random distribution of the questionnaire booklets. The cognitively complex and simple rater groups were created by eliminating the middle twenty percent of the subjects, ranked by their cognitive complexity scores.

The data analysis was performed on two levels. Individual level analysis

was performed to extract measures to represent dependent variables for between-groups analysis of interest.

For the examination of information utilization, there are two paradigms, factorial ANOVA and regression (Slovic & Lichtenstein, 1971). The regression procedure, known as policy capturing or len's modeling, was used in the present study. This choice was based on the wide use of policy capturing analysis in previous related research on performance appraisal (Cadwell, 1985; Zedeck & Cascio, 1982; Zedeck & Kafry, 1977), and in various other areas where information utilization and integration is of interest (Borko & Cadwell, 1982; Cadwell & Jenkins, 1986; Norman, 1986).

In policy capturing analysis, a multiple linear regression model is utilized as a descriptive tool, capturing the various aspects of vicarious functioning (Einhorn, et al., 1979; Shavelson, Webb, & Burstein, 1986). A structural multiple linear regression analysis was performed for each subject. The equation was as follows:  $Y_1 = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$ . In this equation,  $Y_1$  represented the vector of rating responses from rating judgment Task A, and  $X_1$  to  $X_4$  represented the four cues or information dimensions. The cue dimensions were coded 2, 1, and 0 for above average, average, and below average performance, respectively. A subject's rating responses, for the 27 ratee profiles in rating Task A, were regressed on the precoded values of the performance dimensions. The unstandardized regression coefficients indicated how much influence each dimension had in a subject's rating, and thus, reflected the subject's information weighting or utilization policy (Einhorn, 1970, 1971; Einhorn, et al., 1979; Slovic & Lichtenstein, 1971; Zedeck & Kafry, 1977).

In the present investigation, policy capturing was used to generate data for hypothesis testing (Zedeck & Cascio, 1982). The policies of apriori groups or clusters (raters in summative and formative conditions) were compared using ANOVA procedures. The regression weights and  $R^2$ s were used as data for testing hypotheses concerning information utilization and integration (Anderson, 1977; Zedeck & Cascio, 1982; Zedeck & Kafry, 1977)

Information integration strategies were identified by comparing regression models. The amount of variance explained by the linear component in the regression model, reflected the use of a compensatory strategy. The use of a noncompensatory strategy was inferred from the amount of variance accounted for by the nonlinear component in regression modeling (Billings & Marcus, 1983; Einhorn, 1970, 1971, Weldon & Gargano, 1985).

#### **4. Data Collection**

After approval from the university ethics committee on research, student subjects were sought for voluntary participation in the study. The experimenter administered the instruments to the subjects individually and also in groups of 11, 13, and 21. Every person was administered the same instruments in the same order with one exception. The exception was that one half of the subjects were primed to make rating judgments as feedback (formative condition); the other half were primed to make rating judgments for promotion decisions (summative condition).

After a brief introduction about the general nature of the project, each

subject received a questionnaire booklet. This booklet contained the following in this order: (1) a letter of consent; (2) a page outlining the content of the questionnaire booklet and the purpose of rating; (3) the Importance of Information Measure; (4) the Good Instructor Schema Measure; (5) Rating Task A with the instructions including the purpose of rating as priming stimulus; (6) Rating Task B; (7) the REP grid. All subjects completed the instruments, in the order presented, in about 30 to 40 minutes. As the subjects had some familiarity with the judgment task, no difficulties were encountered in the administration of the instruments.

The analyses of the data and the results are presented in the next chapter.



## IV. ANALYSIS AND RESULTS

This chapter presents the results in two parts. The tests of the hypotheses are presented in the first part (A). The second part (B) includes the results of the exploratory analyses.

Most of the analysis was done using SPSS:X (Nie, 1983) and BMDP (Dixon, 1983) statistics software. The assumptions underlying the statistical analyses performed were examined throughout. Where the classical approach to analysis of variance has been employed, the strength of association between the variables is expressed by "eta squared". Because in the case of a nested ANOVA,  $\eta^2$  tends to be upwardly biased, a conservative estimate was obtained by dividing the  $SS_{\text{effect}}$  by  $SS_{\text{total}}$  (Hays, 1981; Keppel, 1973; Vaughn & Corballis, 1969).

### A. TEST OF THE HYPOTHESES

The research hypotheses were presented in the previous chapter (Part 3.B); they are presented here as well for ease of reference. Although the research hypotheses were directional, they were cast into null form for statistical testing. The hypotheses are translated into statistical terms (effects) corresponding to the statistical models of analysis. The criterion for rejection of the hypotheses was a two-tail test at the conventional alpha level of .05. As the evaluation of the hypotheses required different approaches, supporting statistical information concerning the variables involved are included.

## 1. Importance of Information

Hypothesis 1.A stated that Appraisal Purpose will affect subjective importance ratings of trait and behavior information in performance judgment. Hypothesis 1.B stated that Cue Dimensionality will affect subjective importance ratings of trait and behavior information in performance judgment. Hypothesis 1.C stated that Appraisal Purpose and Cue Dimensionality will conjointly affect subjective importance ratings of trait and behavior information in performance judgment.

The subjective importance given to trait and behavior dimensions of information was measured by the Importance of Information Measure which had reliabilities (Cronbach alpha) of 0.81 and 0.73 in formative and summative conditions, respectively. Subjects, primed with either a formative or a summative appraisal purpose, indicated the importance of each item on a seven point scale. The average importance for trait items, and the average importance for behavior items formed the two dependent variables which were treated as a repeated measure of Cue Dimensionality.

A repeated measures analysis of variance (SPSS:X ANOVAR) was performed with Cue Dimensionality as the within-subjects factor and Appraisal Purpose as the between-subjects factor. The assumptions underlying the analyses were met (Winer, 1971). The results are presented in Table 1, and the relationship between the variables is graphically displayed in Figure 1.

As can be seen in Table 1, the between-subjects effect, that is, the effect of Appraisal Purpose was not significant,  $F(1,68)=0.238$ . Therefore, the null

hypothesis 1.A was not rejected. The within-subjects effect of Cue Dimensionality was significant,  $F(1,68)=39.692$ ,  $p<.05$ , and so was the interaction effect between Appraisal Purpose and Cue Dimensionality,  $F(1,68)=13.683$ ,  $p<.05$ . Consequently, null hypotheses 1.B and 1.C were rejected. As estimated by  $\eta^2$ , the strength of association between Cue Dimensionality and importance ratings was 0.118. These results indicate a significant effect of Cue Dimensionality, and a significant interaction effect of Purpose and Cue Dimensionality on the subjective importance ratings of trait and behavior information in performance judgment.

Table 1

**The Effect of Purpose and Cue Dimensionality on Importance Ratings**

Source	DF	Mean Sqs.	F Ratio	$\eta^2$
A-Appraisal Purpose	1	0.257	0.238	0.002
S-Within	68	1.081		
B-Cue Dimensionality	1	13.578	39.692*	0.118
AB	1	4.681	13.683*	0.041
BS-Within	68	0.342		

\*  $p<.05$ .

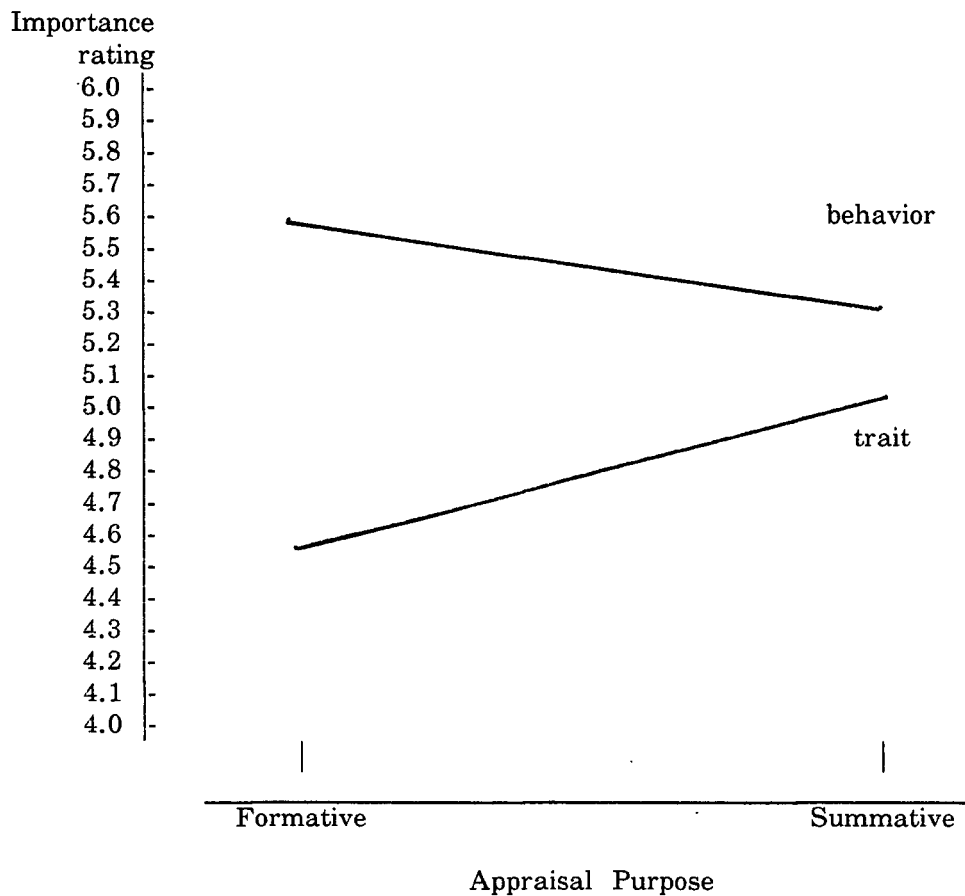


Fig. 1: Mean Importance Rating of Cues

Further analysis was undertaken to identify traits and behaviors which contributed to the group difference on Cue Dimensionality and the interaction between Appraisal Purpose and Cue Dimensionality. The average importance ratings for each dimension are reported in Table 2. The *t* values (independent test) in Table 2 indicate that the two appraisal groups differed significantly on three behavior dimensions (planning-preparation, grading-marking, communication) and one trait dimension (enthusiasm). The means on the three behavior dimensions are lower for the summative group, suggesting that raters making

judgments for promotion did not consider these behavior dimensions as important as did the raters evaluating for feedback. In the trait category, although enthusiasm was considered most important by both groups, it was significantly more important for the summative group. Leadership was more important for summative judgment, but the difference between the groups was not significant. Except for research activity, the mean ratings show that the formative group considered behavior dimensions more important than trait dimensions, although the difference between the groups was not statistically significant on all dimensions.

Table 2

**Mean Rated Importance of Performance Related Information**

Cue Dimension	Formative		Summative		t
	M	SD	M	SD	
<b>Behavior</b>					
Planning-preparation	6.11	1.2	5.62	0.7	1.79*
Lecture presentation	6.20	1.0	6.09	0.7	0.56
Grading-marking	5.35	1.3	4.71	1.5	1.91*
Communication	6.40	0.9	5.94	1.3	1.76*
Research activity	3.85	1.7	4.14	1.7	-0.70
<b>Trait</b>					
Enthusiasm	5.20	1.1	5.86	1.3	-2.32*
Sociability	3.89	1.4	4.31	1.5	-1.28
Resourcefulness	4.94	1.2	5.37	1.2	-1.49
Leadership	4.77	1.3	5.03	1.4	-0.77
Warmth	4.17	1.4	4.66	1.3	-1.52

\*  $p < .05$ . Note: N=35 in all groups. Scale was 7 points.

## 2. Utilization of Information

Hypothesis 2.A stated that Appraisal Purpose will affect utilization of trait and behavior information in performance judgment. Hypothesis 2.B stated that Cue Dimensionality will affect utilization of trait and behavior information in performance judgment. Hypothesis 1.C stated that Appraisal Purpose and Cue Dimensionality will conjointly affect utilization of trait and behavior information in performance judgment.

The relative utilization of trait and behavior dimensions of information in the ratee profiles was measured by regression modeling or policy capturing procedure described in chapter 3 (Part C.3). A regression model was computed for each person to obtain the subject's information utilization policy. The four cue dimensions in the profiles were regressed on the vector of ratings given to the 27 profiles in rating Task A. The reliabilities were 0.74 and 0.84 (Cronbach alpha) for summative and formative raters, respectively. As the coded vectors were uncorrelated, the regression weights were treated as an index of the relative utilization of the different types of information (Pedhazur, 1982).

The linear regression models estimated were significant for every subject except for subject 3 in the summative group. Statistics on the proportion of variance explained by the main effects in formative and summative judgment are presented in Table 3. As shown in Table 3, it appears that slightly more rating judgment variance may be explained in formative judgment. Also, raters in the formative group appear to be more linearly consistent in their rating judgments,  $R^2 = .81$ , than the raters in the summative group,  $R^2 = .76$ .

Table 3**Mean, Median, and Range of Variance Explained by Regression Models**

Statistic	R <sup>2</sup> Formative	R <sup>2</sup> Summative
Mean	.81	.76
Median	.84	.77
Highest	.94	.92
Lowest	.60	.62*

\* Would be .21 if one outlier (Subject 3) is included.

The average regression weight for trait items and the average weight for behavior items for each subject (subject 3 was included because one of the weights was significant) formed the two dependent variables in testing the null hypotheses 2.A, 2.B, and 2.C (Zedeck & Kafry, 1977). The two average weights for behavior and trait items were treated as repeated measures of cue dimensions. A repeated measures ANOVA was performed with Cue Dimensionality as the within-subject factor and Appraisal Purpose as the between-subjects factor. The assumptions underlying the analysis were examined, and as the heterogeneity assumption was met, the summation of policy capturing data was possible (cf. Borko & Cadwell, 1982). The results are summarized in Table 4 and illustrated in Figure 2.

The between-subjects effect, that is, the effect of Appraisal Purpose was not statistically significant,  $F(1,68)=1.802$ . Therefore, null hypothesis 2.A was not rejected. The within-subjects effect of Cue Dimensionality was significant,

$F(1,68)=55.481$ ,  $p<.05$ , and so was the interaction between the Appraisal Purpose and Cue Dimensionality,  $F(1,68)=23.25$ ,  $p<.05$ . Consequently, null hypotheses 2.B and 2.C were rejected. As estimated by  $\eta^2$ , the strength of association between Cue Dimensionality and utilization of information was 0.296.

Table 4

**The Effect of Purpose and Cue Dimensionality on Information Utilization**

Source	DF	Mean Sqs.	F Ratio	$\eta^2$
A-Appraisal Purpose	1	0.323	1.802	0.006
S-Within	68	0.179		
B-Cue Dimensionality	1	17.151	55.481*	0.296
AB	1	7.187	23.250*	0.124
BS-Within	68	0.309		

\*  $p<.05$ .

These results indicate a strong influence of Cue Dimensionality, and a significant interactive effect of Appraisal Purpose and Cue Dimensionality on utilization of trait and behavior information in performance judgment. It can be seen in Figure 3 that unlike formative evaluation, trait and behavior information contributed almost equally in the formation of summative judgments.



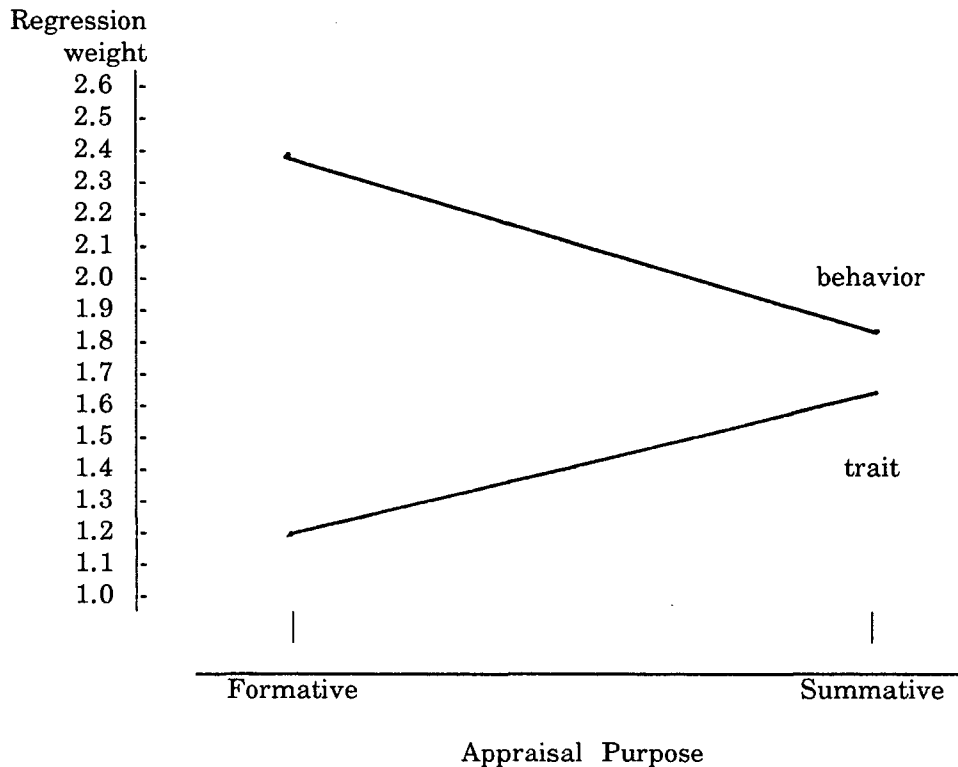


Fig. 2: Mean Weight of Cues Utilized

In further analysis, the utilization of each information dimension was determined for both summative and formative groups. As the policy capturing data were not heterogeneous when tested for in the ANOVA, group average main effects of the information dimensions, as reflected by the regression weights, were tested for significance by comparing the means with zero via a  $t$  test (Norman, 1986). The average regression weight per group is presented in Table 5.

As shown in Table 5, all main effects were significant for summative as well as for formative judgment, which indicates that all information dimensions were effectively utilized. As reflected in the mean weights, presentation clarity

was the most heavily weighted dimension for both appraisal purposes, followed by grading-marking for formative judgment and enthusiasm for summative judgment. Least weighted dimensions were resourcefulness in formative judgment and grading-marking in summative judgment.

Table 5

**Mean Regression Weights for Formative and Summative Judgment**

	Formative	Appraisal	Summative	Appraisal
Main Effects	Mean	t	Mean	t
b1-enthusiasm	1.43	15.23*	1.90	15.41*
b2-present. clarity	3.17	21.11*	2.47	12.67*
b3-resourcefulness	1.04	10.64*	1.29	15.60*
b4-grading-marking	1.61	13.68*	1.21	11.23*

\*  $p < .05$ .

### 3. Subjective Importance and Utilization Consistency

Hypothesis 3 stated that cue utilization in performance judgment will be consistent with subjective importance of cue dimensions. Support for this hypothesis can be gleaned from the pattern of results for hypotheses 1 and 2 above. Nevertheless, for a more direct test of the hypothesis a substantial correlation between the mean importance ratings and the mean regression weights for trait and behavior information was predicted.

A canonical correlation analysis was used to test the null hypothesis of no consistency between subjective importance ratings and utilization of cues as reflected by regression weights. The two sets of variables were (1) the mean subjective **importance** (ratings) of trait and behavior information, and (2) the mean **weight** given to trait and behavior information for each subject. The linearity of relationship between variables, normality of their distribution, and within set multicollinearity were found to be satisfactory in the examination of scatter plots and distributional statistics (Tabachnik & Fidell, 1983).

The analysis showed a significant and substantial relationship between the two sets of variables. Two canonical correlations were significant by Bartlett's test for eigenvalues. The first canonical correlation was 0.55,  $X^2(4)=29.99$ ,  $p<.05$ ; the second was 0.30,  $X^2(1)=6.14$ ,  $p<.05$  with the first canonical correlation removed. As a result, null hypothesis 3 was rejected, and it can be seen in Table 6, that the canonical variates accounted for large amounts of variance in the original variables.

Table 6

**Variance Extracted from Original Sets of Variables by Canonical Variates**

Variate	Cann. Corr.	Original Sets of Variables	
		Importance Rating	Regression Weight
1	0.55	27.3%	59.1%
2	0.30	72.7%	40.9%

#### 4. Information Integration

Hypotheses 4 and 5 addressed the use of compensatory and noncompensatory information integration strategies in rating judgments. The analysis of these hypotheses involved estimating linear and nonlinear mathematical models of information use. In a mathematical model, use of a noncompensatory strategy or configural use of information produces significant interactions among cue dimensions (Billings & Marcus, 1983). Given the fractional factorial design of the rating task in this study, the six first-order interaction effects (AB, AC, AD, BC, BD, CD) were measureable.

Two models were estimated for each subject. One was the main effects model (ME) - the linear regression model developed in the policy capturing analysis. The other was a regression model including both the main and the interaction effects (MEI). For each subject, subtracting  $R^2$ -ME from  $R^2$ -MEI left the proportion of variance explained by all two-way interaction effects,  $R^2$ -INTER (since the error variance is common to both models).

The proportion of variance explained by  $R^2$ -INTER ranged from .01 to .13 and .01 to .24 for formative and summative ratings, respectively. The means and standard deviations of the amounts of variance accounted for by the interaction terms are presented in Table 7. It was assumed that the greater interactive use of cues would result in a greater amount of explained variance (Billings & Marcus, 1983).

In order to determine whether a subject integrated cues interactively in addition to using a linear additive strategy, hierarchical regression was performed,

Table 7**Mean and Standard Deviation of Variance Explained**

$R^2_s$	Formative	Group	Summative	Group
	M	SD	M	SD
$R^2$ -MEI	.86	.07	.83	.09
$R^2$ -ME	.81	.08	.76	.12
$R^2$ -INTER	.05	.03	.07	.05

where the main effects and interaction effects were predictor variables entered in two blocks. The main effects were entered in the first block (linear component), and the interaction terms were entered in the second block (nonlinear component). The decision rule was that if the increment in variance explained due to the second block was significant, the subject was considered to be combining the information dimensions multiplicatively, that is, using a noncompensatory integration strategy in addition to a compensatory strategy.

Hypothesis 4 stated that in comparison to formative judgment, raters in summative judgment condition will combine cue dimensions using a noncompensatory strategy in addition to a compensatory strategy. To test null hypothesis 4, the amount of variance due to the interaction terms ( $R^2$ -INTER) was used in an ANOVA (Anderson, 1977). The effect of Appraisal Purpose was not significant  $F(1,68)=2.34$ , and the null hypothesis was not rejected.

At the individual level, eight subjects in the summative (22.9%) and six subjects in the formative group (17.1%) were identified as combining cues using both compensatory and noncompensatory strategies. The configural impact of cue dimensions is illustrated in Figures 3 and 4 for the 6 subjects in the formative group, and in Figures 5 and 6 for the 8 subjects in the summative group. A severe non-parallelism depicts the use of an interactive, noncompensatory strategy (Anderson, 1982). The graphs portray that in addition to interactive use of the four cues, subjects also used two pairs of cues interactively. For example, subject 3 (Fig. 3, left panel) combined the pair A and B, and the pair C and D, interactively. Similar pattern is noticeable for the rest of the subjects.

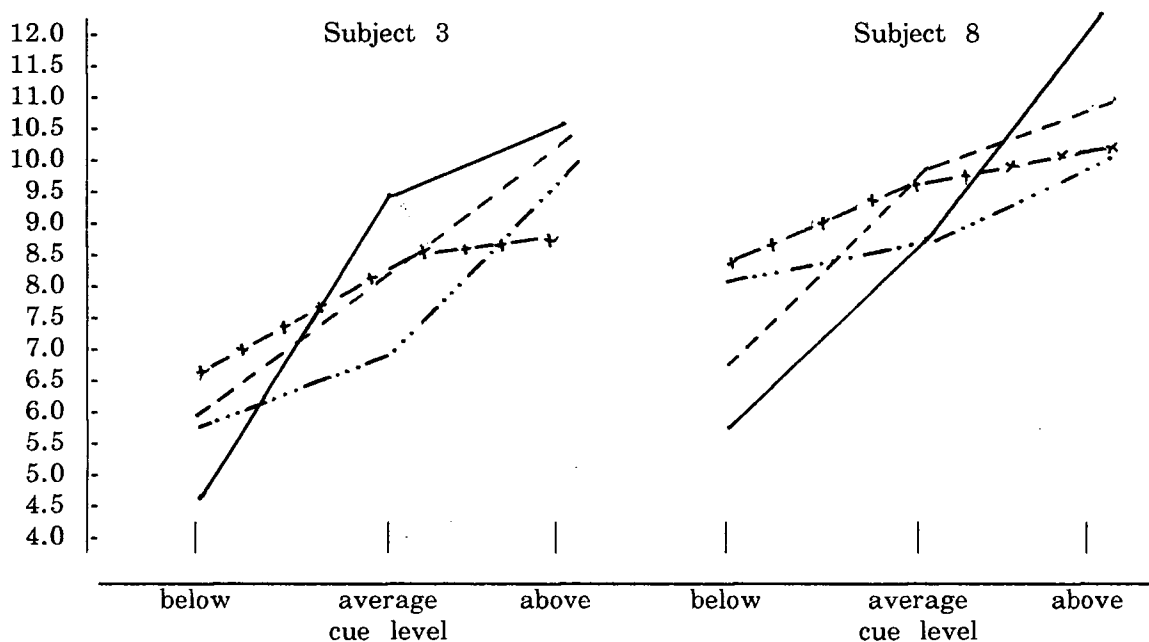


Fig. 3: Plot of Cell Means for subjects 3 and 8 (Formative)

Cues:

- A - enthusiasm
- B - presentation clarity
- + - x - x - x - C - resourcefulness
- ..... D - grading-marking

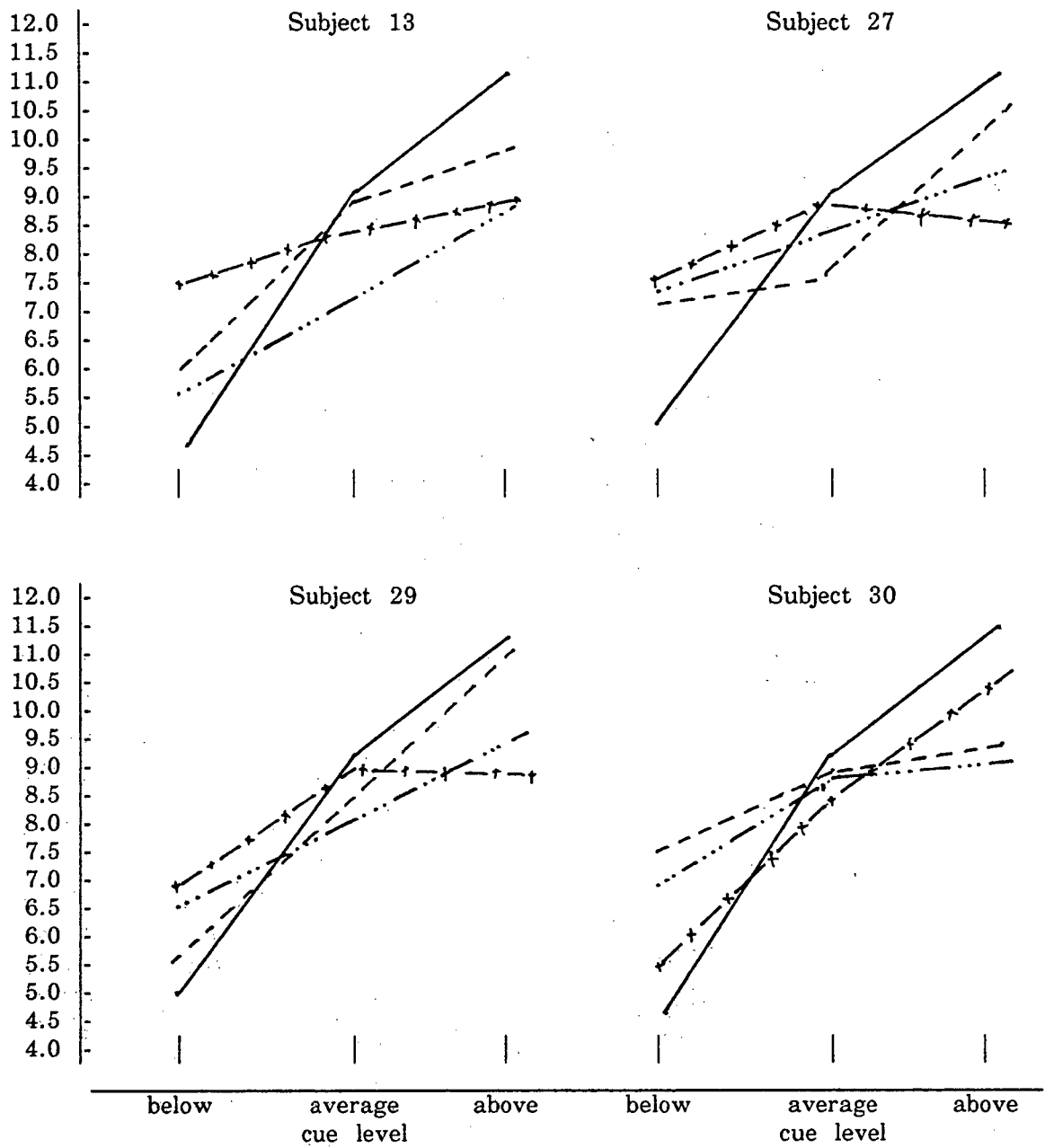


Fig. 4: Plot of Cell Means for Subjects 13, 27, 29, and 30 (Formative)

Cues:

- A - enthusiasm
- B - presentation clarity
- - - C - resourcefulness
- ..... D - grading-marking

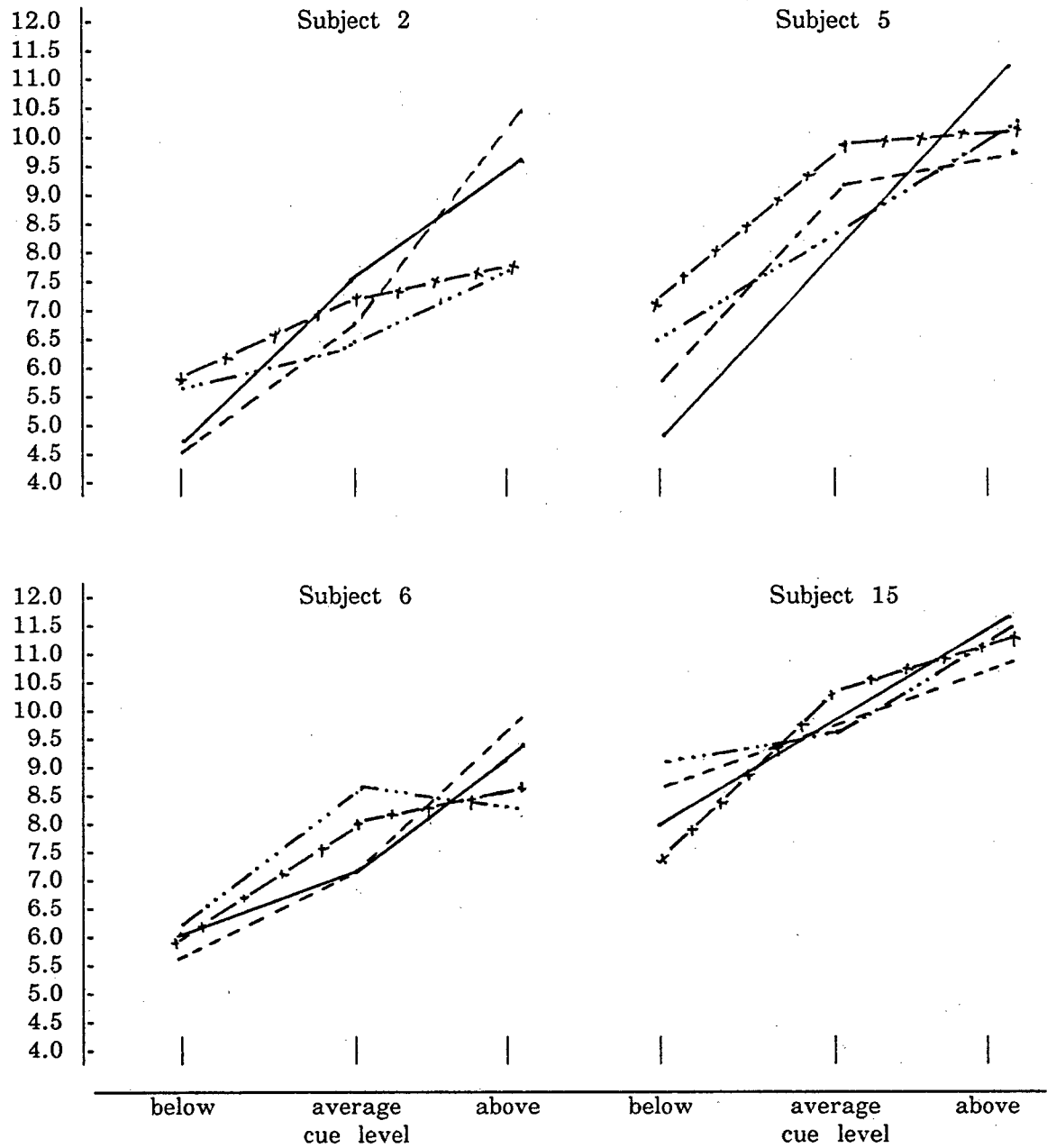


Fig. 5: Plot of Cell Means for Subjects 2, 5, 6, and 15 (Summative)

Cues:

- A - enthusiasm
- B - presentation clarity
- +—+—+ C - resourcefulness
- ..... D - grading-marking



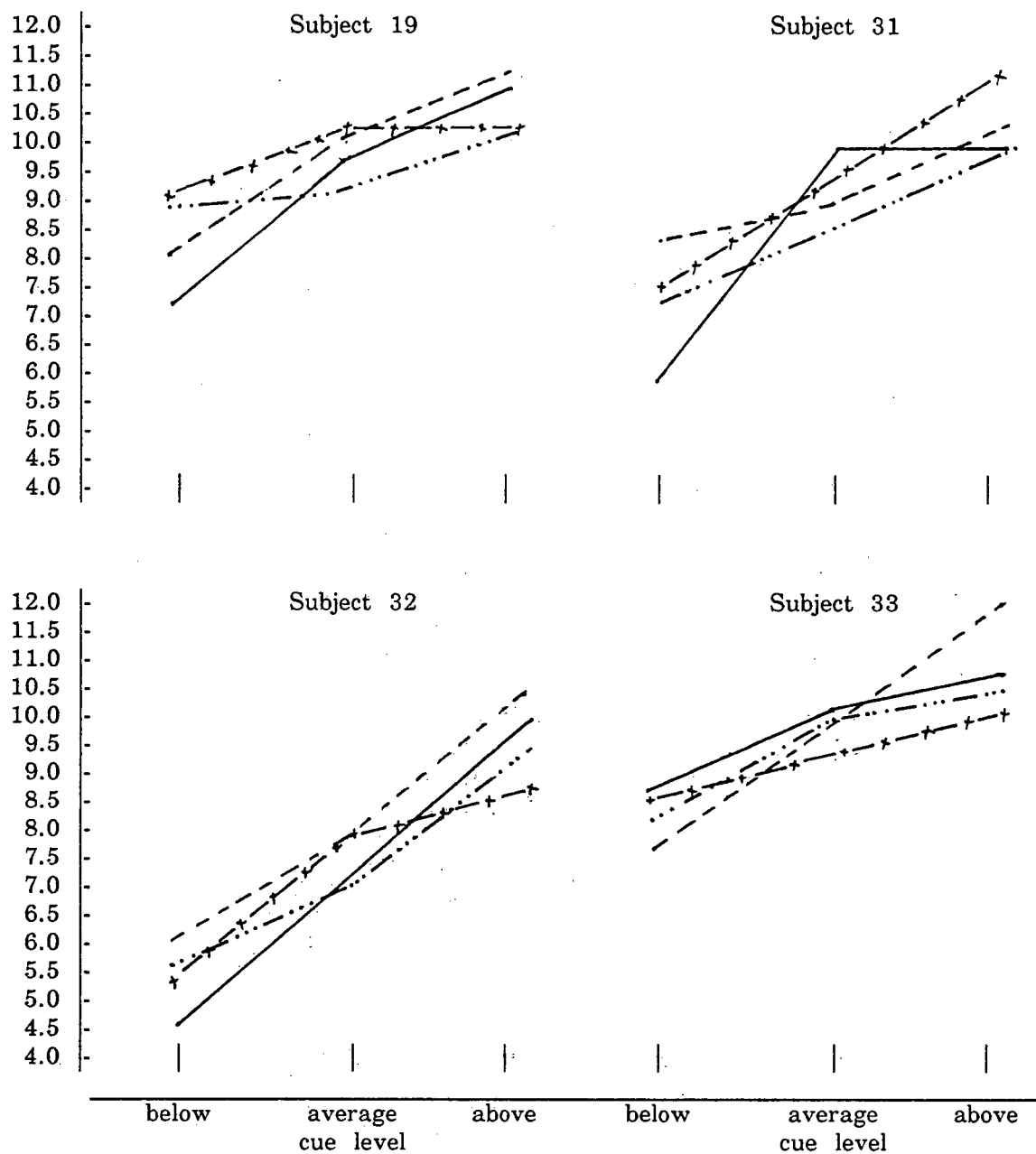


Fig. 6: Plot of Cell Means for Subjects 19, 31, 32, and 33 (Summative)

Cues:

- A - enthusiasm
- B - presentation clarity
- x-x-x- C - resourcefulness
- .-.- D - grading-marking

Hypothesis 5 stated that in comparison to cognitively simple raters, complex raters will combine cue dimensions using a noncompensatory strategy in addition to a compensatory strategy. Cognitive complexity was measured by the REP grid which was scored for the total number of different pairs of responses across the constructs. The reliability (Cronbach alpha) of the whole grid (100 items) was 0.94, using the ratings on ten constructs. As the scoring was reversed, the smallest total score reflected the most complexity. The smallest score was 67, the median was 109, and the largest score was 188.

In order to create two groups (cognitively simple and cognitively complex), the middle 20 percent ( $N=14$ ) of the subjects were excluded from the analysis. Cognitive complexity scores ranged from 66 to 102, and cognitive simplicity scores ranged from 116 to 188, with 28 subjects in each group. The two groups were significantly different on cognitive complexity scores,  $t(54)=13.52$ ,  $p<.05$ , ( $M=91.32$ ,  $SD=9.8$  and  $M=146.89$ ,  $SD=19.4$  for complex and simple groups, respectively).

To test null hypothesis 5, the variance due to the nonlinear component, that is,  $R^2$ -INTER was used as the dependent variable for each subject in an ANOVA (Anderson, 1977). The effect of Cognitive Complexity was significant  $F(1,54)=7.45$ ,  $p<.05$ . Hence, the null hypothesis was rejected. The means of  $R^2$ -INTER indicate that cognitively complex raters ( $M=6.2$ ,  $SD=3.3$ ) combined dimensions interactively more than the cognitively simple raters ( $M=4.0$ ,  $SD=2.5$ ), in addition to using a compensatory strategy.

## B. EXPLORATORY ANALYSIS

There were two questions of exploratory interest in the present study. One related to the measurement of good instructor schema, and the other concerned the effect of cognitive complexity on halo in ratings. The analysis and results pertaining to these questions are presented below in two separate sub-sections.

### 1. Measuring A Good Instructor Schema

The impact of a person's schema has been highlighted in several information processing conceptions of performance evaluation (DeNisi, et al., 1984; Feldman, 1981). However, measurement of schema has not received much attention (Fiedler, 1982). Therefore, it was of interest to explore if a good instructor schema profile could be measured. A measure of stereotype developed by McCauley et al. (1978, 1980) was adapted to measure quantitatively the good instructor schema profile held by the raters.

It was expected that the attributes irrelevant to the good instructor schema would not hold a diagnostic value. According to the Bayesian procedure by McCauley et al. (1978, 1980), a diagnostic ratio (DR) indicates the strength of an attribute in the schema to the extent the ratio differs from 1.0. Therefore, diagnostic ratios were computed from the probability estimates ( $p(\text{group/behavior})/p(\text{group})$ ) probability estimates for each of the ten attributes included in the questionnaire. The mean DRs for the entire sample were tested for difference from 1.0 by *t* tests, and the results are reported in Table 8. An examination of the pattern in the DRs indicates that because the DRs for the

irrelevant attributes are not significantly different from 1.0 and are also much lower in comparison to the relevant attributes, the Bayesian technique used in this study seems a valid procedure for measuring a good instructor schema profile.

Table 8

**Mean Diagnostic Ratios in Good Instructor Schema Profile**

Dimension	M	SD	t
Social work†	1.31	1.3	1.92
Travelling†	1.42	1.8	1.91
Enthusiasm	2.45	4.2	2.90*
Presentation	2.67	4.4	3.15*
Outgoing	2.09	3.6	2.53*
Resourcefulness	2.57	5.0	2.66*
Preparation	2.81	5.1	2.97*
Grading	2.53	4.5	2.83*
Leadership	2.18	3.6	2.74*
Communication	2.67	5.3	2.65*

\*  $p < .05$

† irrelevant attributes

The identification of a good instructor schema permitted further exploration into the effect of appraisal purpose on attribute strength in the schema profile.

Trait attributes were expected to be more salient than behavior attributes in summative condition. In formative condition, behavior attributes were expected to be more salient than trait attributes. In statistical terms, an interaction between purpose, and trait and behavior schema scores (DRs) was expected.

The mean DR for trait items and the mean DR for behavior items were treated as the two levels of a within-subject factor, Schema Profile, in a repeated measures ANOVA. Appraisal Purpose was the between-subjects factor. The main effect of Appraisal Purpose was not significant,  $F(1,68)=1.30$ . The effect of the Schema Profile was significant,  $F(1,68)=5.57$ ,  $p<.05$ . Also, there was a significant interaction between Appraisal Purpose and the Schema Profile,  $F(1,68)=4.65$ ,  $p<.05$ . The significant interaction shows that the saliency of traits and behaviors in the good instructor Schema Profile was influenced by the purpose for judgment.

## 2. Cognitive Complexity and Halo

Exploratory analysis was also done to examine the degree of halo in the ratings of complex and simple raters. In rating Task B, subjects provided decomposed ratings on ten different dimensions. Rating Task B had a reliability of .71 (Cronbach alpha). The vignette evaluated, contained information relating to only five of the dimensions; information pertaining to the other five dimensions was missing. Two average ratings were computed for each subject. One rating was the average of ratings made on dimensions for which information was supplied, and the other was the average of ratings for the dimensions on which information was withheld, resulting in a pair of ratings (x,y) for each subject.

The ratings (x,y) of the 28 subjects in the complex group and the 28 subjects in the simple group were correlated separately to obtain the degree of halo. The correlations were 0.48,  $p < .05$ , for the cognitively complex raters and 0.77,  $p < .05$  for the cognitively simple raters. Although the correlations were significant in both groups, the halo effect, indicated by the strength of the correlations, was weaker (lower correlation) for cognitively complex raters. The correlations for the two groups differed significantly,  $z = 1.76$ ,  $p < .05$ , when tested by Fisher's  $z$  transformation of  $r$  (Guilford & Frutcher, 1978). These results show that cognitively complex individuals rated with less halo compared to the cognitively simple raters.

In summary, the data did not support a main effect of Appraisal Purpose, either for subjective importance ratings or for cue utilization. Nevertheless, there was a significant effect of Cue Dimensionality, and also a significant interactive effect of Purpose and Cue Dimensionality on both subjective importance and utilization of information. Purpose did not influence information integration strategies but cognitive complexity did. In exploratory analysis, a good instructor schema was quantified, the effect of appraisal purpose on schema profiles was detected, and the cognitively complex group rated with less halo than the simple group. These findings are discussed in the following chapter.

## **V. DISCUSSION**

This final chapter provides a review of the findings and their interpretation in relation to theories and issues in research considered in earlier chapters. The interpretation of the findings from confirmatory analysis is presented in the first (A) and second (B) parts. Exploratory results are discussed in the third part (C). A summary of the findings and the conclusions are presented in part four (D). The strengths and limitations of the study are discussed in part five (E). In the sixth (F) and seventh (G) parts, respectively, are the implications and directions for further research.

### **A. IMPORTANCE AND UTILIZATION OF INFORMATION**

#### **1. Effect of Purpose**

It was speculated that purpose of performance judgment will influence the subjective importance and actual utilization of performance information. The data did not support a main effect of appraisal purpose, either on importance ratings, or on cue utilization.

Lack of support for a cognitive effect of appraisal has also been the case in some of the previous studies (Murphy, et al. 1984; McIntyre, et al., 1984), but certain differences between these studies and the present are worth noting. These researchers analyzed rating responses (the product), whereas the dependent variables in the present study were subjective importance and weighting of information (the processes). Because how information is mentally weighted reflects

cognitive strategies, the investigation of rater cognition may be considered more direct in the present study than in others (Murphy, et al., 1984; McIntyre, et al., 1984). However, the main effect of appraisal purpose on rater cognition was not observed in the current study either.

Previous findings on the effect of appraisal purpose on raters' cognition were inconsistent (Murphy et al., 1984; McIntyre et al., 1984; Williams, et al., 1985, Zedeck & Cascio, 1982). In an investigation of raters' information utilization policies, Zedeck and Cascio (1982) found a significant effect of purpose, but Williams, et al. (1985) suggested the opposite. This mixed findings were from studies that used different methodologies. As did Zedeck and Cascio (1982), policy capturing methodology was used to investigate information utilization in the current study as well. However, the findings in the present study are not in agreement with that of Zedeck and Cascio's (1982). Therefore, even though the present study does not lead to a complete resolution, it does suggest that different methodologies, that is policy capturing versus analysis of ratings, are not underlying the inconsistent findings of previous research.

The interaction effect of purpose and cue dimensions was significant. When the interaction effects are significant in an analysis of variance, the main effects cannot be interpreted in isolation (Kirk, 1982). As the predicted interaction of purpose and cues was significant, the effect of purpose is considered further in the discussion on the interactive effect of purpose and cue dimensionality.



## 2. Effect of Cue Dimensionality

It was speculated that cue dimensionality will influence the subjective importance and actual utilization of performance information. The data supported the main effect of cue dimensionality. The effect of cue dimensionality on the subjective importance of information and utilization of similar information was uniform, but the effect was stronger in information utilization than in importance ratings. As estimated by  $\eta^2$ , 11.8% of the variance in importance ratings and 29.6% of the variance in cue utilization was associated with cue dimensions.

One explanation for the stronger effect of cue dimensionality in information utilization lies in the level at which information may have been processed mentally (Craik & Lockhart, 1972). It is conceivable that actually making the rating judgments involves deeper processing and mental concentration than reporting the subjective importance of information. The formation of judgments involves prototype matching processes, whereby the incoming information is given meaning and then mentally transformed into a rating response. Therefore, the stronger effect of cue dimensionality in information utilization allows the speculation that if schema-induced biases enter performance judgment, they are likely to be more pronounced when one actually forms the judgment, than when one decides what kinds of information would be valuable.

As discussed in chapter two, there are few studies that address the effect of cue dimensionality on information processing in performance judgment. After all, it is the content of information on which judgments are based, and the strong influence of cue dimensionality in this study bears it out. Besides, the

effect of cue dimensionality on information utilization is a common finding in previous literature on human judgment research (Nisbett & Ross, 1980; Wallsten & Barton, 1982). As cue dimensions in this study were traits and behaviors, one aspect of cue salience appears to be the nature of the information - how concretely the information can be mentally represented (Nisbett, et al. 1976).

An examination of the dimensional use of information showed that in both summative and formative conditions, raters gave most weight or priority to presentation clarity in arriving at their rating judgments. In the second place was grading-marking for formative appraisal, but enthusiasm for summative appraisal. In fact, grading-marking was least utilized in summative appraisal. The raters were student teachers, and therefore, the clarity of presentation may have been important to them.

The finding that enthusiasm was not the most weighted dimension for either purpose of evaluation is surprising, when we consider the conclusions of studies in the Dr. Fox tradition (Naftulin, et al., 1973). Researchers of the Dr. Fox effect have suggested that expressive behavior portraying instructor enthusiasm is an important source of variation in student ratings (Abrami, et al., 1982). On such conclusions, the validity of students' ratings of instructors is doubted, even though many of the Dr Fox studies lacked internal validity (Frey, 1978). In the present study, enthusiasm was important but not to the exclusion of presentation clarity. However, in making this comparison we should keep in mind that in Dr. Fox studies enthusiasm was manipulated behaviorally by using video tapes of acted lectures, whereas in the present study enthusiasm was presented as a concept and students attached their subjective meaning to it.

Another finding in this study runs against a popular belief that casts doubt on the validity of students' ratings of instructors for summative evaluation. The "grading-satisfaction" hypothesis is that students base their ratings on the grades and marks they receive from the instructors (Cohen, 1981; Marsh & Overall, 1980). The finding in the present investigation is to the contrary - grading-making was considered of least importance in summative evaluation.

Cue dimensionality had a strong effect on subjective importance and utilization of information, but the predicted interactive effect of cue dimensionality and purpose was significant as well. Therefore, the effect of cue dimensionality is further discussed in the next section.

### **3. Interactive Effect of Purpose and Cues**

As predicted, the data supported the interactive influence of purpose and cue dimensionality on both subjective importance and actual utilization of performance information. The main effect of cue dimensionality accounted for 11.8% of the variance in importance ratings and 29.6% of variance in cue utilization, as estimated by  $\eta^2$ . Comparatively, the interactive effect of purpose and cue dimensionality accounted for 4.1% of the variance in importance ratings and 12.4% of variance in cue utilization. Although the interactive effect was relatively weaker, it is the significance of this effect that leads to greater insight into the nature of information processing in performance judgment, because human reasoning has multiple causes.

It was found that raters in the formative judgment condition gave more importance to behavior information; on the other hand, raters in the summative judgment condition gave more importance to trait information. In an identical analysis, similar results were obtained on information utilization, but the effect was stronger. The stronger interactive effect (like the stronger main effect of cue dimensionality) on information utilization may be a result of the depth of processing (Craik & Lockhart, 1972). Actually making the rating judgments may involve deeper processing and mental concentration than reporting the subjective importance of information.

Various theoretical points of view can be drawn upon for the interpretation of the interactive effect of appraisal purpose and cue dimensionality. It has been suggested that appraisal purpose orients the rater to select an internal frame of reference or schema, which guides the interpretation of performance information (DeNisi et al., 1984). This conception implies that purpose operates as a priming stimulus (Loftus & Loftus, 1974), and the frame of reference is the set of mentally stored prototypes or schemata (Taylor & Crocker, 1981). Ilgen (1981) and Ilgen and Feldman (1983) suggested that rating judgments result from a prototype matching or schema based interpretation and utilization of information. Schematic processing is considered the basic cognitive mechanism in human judgment (Bruner, 1971; Cantor & Mischel, 1977; Nisbett & Ross, 1980; Taylor & Crocker, 1981). Schema profiles comprise traits embedded in person schema (Cantor & Mischel, 1977) and behaviors embedded in role schema (Taylor & Crocker, 1981). Appraisal content, too, comprises traits and behaviors (Wexley & Klimoski, 1984).

Therefore, it seems that purpose activates a schema profile in terms of person and role schemata, which provides the basis for evaluating or processing the information related to performance. The importance given to trait and behavior information may be induced by person and role schemata, respectively. Performance information that matches the "initialized" schema profile becomes more salient than other information. Hence, it can be speculated that subjective importance and utilization of that information in rating judgments is indirectly affected by appraisal purpose. The cognitive effect of appraisal purpose appears to be through schematic processing, and not as direct as envisaged by some researchers (Landy & Farr, 1980; Murphy et al., 1984; McIntyre et al., 1984; Williams, et al., 1985).

The effect of purpose on information utilization in the study by Zedeck and Cascio (1982) may be interpreted in terms of schema utilization as well. They did not offer any theoretical interpretation, and restricted their discussion suggesting that subjects utilized information from manager and consumer perspectives. The utilization of "bagging skill" and "skill in human relations" may have resulted from schematic processing because these dimensions of information relate to role and person schemata, respectively. If this interpretation of their study is possible, then taken together, the findings would support the pervasiveness of schema driven judgment in performance appraisal as well.

Certain findings in the current study (discussed in subsequent sections) support the above interpretation of the interactive effect of cue dimensionality and purpose. If it can be assumed that subjective importance reflected mental models or implicit theories, then the finding of consistency between subjective importance

and cue utilization endorses the schematic processing explanation. Likewise, the exploratory result showing that good instructor schema profiles varied with appraisal purpose supports the priming function of purpose and the schematic processing interpretation of the interactive effect of cue dimensions and purpose.

#### **4. Subjective Importance and Utilization Consistency**

The data supported the hypothesis concerning consistency between subjective importance ratings and utilization of similar information. Canonical correlation analysis revealed that in making the rating judgments, raters actually used the information they reported as important prior to performing the rating task. An indication of a substantive relationship between subjective importance and objective weighting of the information dimensions is revealed by the fact that two significant variates were extracted. Unless subjective importance ratings and the weights depicting cue utilization had a substantial relationship, obtaining two significant variates when there were only two variables in each set of original variables would have been difficult (Marascuilo & Levin, 1983; Pedhazur, 1982; Tabachnik & Fidell, 1983).

The close relationship between subjective importance ratings and objective weights, that is, the consistency between information reported as important and information actually utilized, is interesting because some researchers have concluded that people are unable to report the importance of information in their own judgment (Nisbett & Wilson, 1977; Slovic & Lichtenstein, 1971; Schmitt & Levine, 1977). The substantive relationship between importance ratings and utilization of information in this study does not support such a conclusion. The

finding in this study is similar to that reported in other studies, which substituted self reports in regression equations and found median correlations around .60 between the actual and predicted judgments (Blazer, Rohrbaugh & Murphy, 1983; Cook & Stewart, 1975; Hoepfl & Huber, 1970). Some of these studies have been reviewed by Surber (1986). Surber's subjects were also able to self-report the importance of information in judgment of children's achievement, which made him question the validity of extreme conclusions regarding people's inability to report the importance of factors in their judgment.

## **B. INFORMATION INTEGRATION**

Psychologists have long studied judgment and decision processes (for reviews see Einhorn & Hogarth, 1981; Pitz & Sachs, 1984; Slovic & Lichtenstein, 1971). Most of these studies extract mathematical models of judgment made in a multiattribute judgment task and infer information integration from regression analyses. In this tradition, the present study investigated how dimensions of information were mentally combined in performance judgment.

The use of two broad categories of information integration strategies was examined. One was the compensatory strategy where a person may trade-off high and low levels of information between dimensions in order to simplify the judgment task. The other was the noncompensatory strategy where a person chooses to use multiple cut-off strategies to combine multiple dimensions of information (Hogarth, 1980). The existing literature suggests that compensatory strategies hold in a wide variety of judgment tasks (Dawes, 1979; Einhorn & Hogarth, 1981; Pitz & Sachs, 1984; Slovic & Lichtenstein, 1971).

Mathematical models, specifically multiple regression models, were developed for individuals to infer their information integration strategies. A linear additive equation fits the compensatory strategy, whereas a nonlinear main and interaction effects model could reflect a noncompensatory strategy (Billings & Marcus, 1983; Einhorn, 1970, 1971; Einhorn & Hogarth, 1981). A comparison of the amount of variance explained by the two models indicates which model better describes the strategies being used (Einhorn, et al., 1979).

Overall, only 20 percent of the subjects in this study combined cues interactively, that is, used some form of a nonlinear, noncompensatory strategy in addition to using a compensatory strategy. This is not a surprising finding for compensatory strategies or linear additive/averaging models hold in a large number of judgment situations (Anderson, 1981; Dawes & Corrigan, 1974; Slovic & Lichtenstein, 1971). Also, decision theorists recommend the use of linear, compensatory strategies for optimal judgments (Edwards & Tversky, 1967).

The regression modeling also addresses the linear consistency in information processing behavior. Linear consistency was indicated by the  $R^2$  for each individual. The mean  $R^2$  values were .86 and .83 for formative and summative raters, respectively. These high levels of consistency does not support Brehmer's (1976) suggestion that individuals may weigh cues fairly accurately but fail to apply this knowledge consistently. However, linear consistency is likely to be a result of the number of cues and the correlation between the cues.

Two hypotheses were tested concerning the use of integration strategies. The results are discussed in the next two sub-sections.



### 1. Effect of Purpose

The data did not support the hypothesis that the use of information integration strategies will vary with purpose of judgment. Eight subjects in the summative condition and six in the formative condition were identified as users of both compensatory and noncompensatory strategies. For summative as well as formative purposes, subjects generally used a compensatory strategy.

The results may be different if the number of information dimensions in the ratee profiles are varied. There is evidence that people's judgment strategies vary with the cognitive demand and presentation format of the judgment task (Crowder, 1976; Einhorn & Hogarth, 1981; Payne, 1976; Wright, 1974). When Zedeck and Kafry (1977) presented nine dimensions, only four were effectively utilized on the average, in their study concerning appraisal of nurses. In another study concerning selection of graduate school applicants, significant differences existed between the two, four, and six cue conditions (Einhorn 1971).

Further, linear integration of cues is not always the case. There are situations in which configural or nonlinear, noncompensatory cue integration is present (Birnbaum & Stegner, 1981; Einhorn, 1970, 1971; Janis & Mann, 1977; Norman & Louviere, 1974; Stumpf & London, 1981; Wallsten & Budescu, 1981). For example, Stumpf and London (1981) found that configural or nonlinear models were needed to account for the way student and manager subjects evaluated job applicants. Moreover, it has been argued that the wide application of the linear model may in part be a result of an artifact of regression analysis itself (Simon, 1976). Linear multiple regression models are very robust with

respect to departures from linearity (Dawes, 1979; Pedhazur, 1982).

Nevertheless, an interesting observation was that the 14 subjects who used both compensatory and noncompensatory strategies, differed on cognitive complexity scores from the rest of the subjects. The subjects who used both compensatory and noncompensatory strategies were markedly more complex, and fell in the top quartile (67 to 97) of the cognitive complexity scores. The range for the entire sample was 67 to 188 (lower scores indicate greater complexity). The effect of cognitive complexity on the use of information combination strategies is discussed next.

## **2. Effect of Cognitive Complexity**

The hypothesis that cognitively complex raters would use a noncompensatory information integration strategy in addition to a compensatory strategy, was supported. An analysis of the amount of variance accounted for by interaction terms in the regression model revealed a significant difference between the complex and simple rater groups. The mean variance due to interaction terms was significantly greater for complex raters, which suggests that cognitively complex raters seem to have a tendency to use both compensatory and noncompensatory strategies.

Cognitive complexity is a developmental construct. The effect of developmental constructs on cue integration has been generally neglected in research on human judgment processes (Pitz & Sachs, 1984). Although cognitive complexity has been a variable of interest in performance judgment, the emphasis

in the past has been mainly on psychometric properties of ratings. Few studies, if any, have investigated the use of information integration strategies by complex and simple raters. Yet, the finding in the present study is intriguing because of its congruence with cognitive complexity theory (Bieri et al., 1966; Kelly, 1955; Vannoy, 1965).

## C. FINDINGS FROM EXPLORATORY ANALYSIS

### 1. Measurement of Schema

Numerous authors have highlighted the role of schema and schematic processing of information in performance appraisal (DeNisi, et al. 1984; Ilgen & Feldman, 1983). For example, DeNisi et al. suggested that a good worker schema may determine information search and interpretation in performance judgment.

Although widely accepted as a theoretical construct, measurement of schema and its utilization has received little attention (Fiedler, 1982). Most of the studies to date have taken for granted that raters apply their good worker schema in making appraisals. This deficiency in research is perhaps due to the difficulty associated with measuring schema. Consequently, the present study attempted to explore the extent to which schema could be measured quantitatively.

McCauley et al. (1978, 1980) outlined a procedure for the measurement of individually held stereotypes. In this measure, a characteristic related to the stereotype holds a diagnostic value if its score (diagnostic ratio) differs

significantly from one. This procedure was adapted to measure quantitatively the good instructor schema held by subjects in this study. Since schema, as do stereotypes, have related and unrelated dimensions (Hastie, 1981; Taylor & Crocker, 1981; Wyer & Srull, 1981), it was expected that dimensions irrelevant to a good instructor schema would not hold diagnostic values, that is, would not be considered strong attributes of the schema (McCauley, et al. 1978; 1980).

The results confirmed the expectation. Professional travelling and social (community) work behavior were of lesser relevance to subjects' good instructor schemas as indicated by the pattern of diagnostic ratios in the schema profiles. This affirmed the validity of the procedure used to measure schema. It should be mentioned, however, that the relevant attributes of a good instructor schema were not an exhaustive list, although the dimensions included in the measure were the ones commonly found in performance rating scales. The irrelevant attributes were selected on an ad hoc basis.

As the data provided support for the measurement of a good instructor schema in quantitative terms, further exploratory analysis was undertaken to examine if different schema profiles were activated by different appraisal purposes. The composition of schema profiles was expected to be different in terms of trait and behavior dimensions, and so it was for the two groups of raters. The saliency of traits relative to behavior dimensions was notable in the schema profiles of raters in the summative condition; the saliency of behaviors relative to trait dimensions was notable in the schema profiles of raters in the formative condition. This finding supports the inference drawn earlier in this discussion regarding the influence of appraisal purpose being mediated by the

schemata activated. However, firm empirical confirmation for this line of reasoning will have to await a search for a direct link between schema and information utilization. It was not possible to establish the direct link between schema profiles and information utilization in the present study because all dimensions in the schema measure were not included in the information.

## **2. Cognitive Complexity and Halo**

Since Schneier's (1977) exploratory research, the impact of cognitive complexity on performance appraisal has been emphasized by many authors (Cooper, 1981; DeNisi, et al. 1984; Dunnette & Borman, 1979; Ilgen & Feldman, 1983; Landy & Farr, 1980). Following Schneier's findings, numerous researchers investigated the relationship between cognitive complexity and halo in performance ratings (Bernardin, et al., 1982; Cardy & Carlyle, 1982; Lahey & Saal, 1981; Sauser & Pond, 1981). These studies generally failed to confirm that cognitive complexity affects the amount of halo in ratings.

The findings show that the skeptical conclusions of earlier researchers about the importance of cognitive complexity in performance evaluation, may be premature. In comparison to the cognitively simple raters, the complex raters rated with less halo. This is a finding consistent with Schneier's (1977) results, but quite contrary to the findings of others (Bernardin, et al., 1982; Cardy & Carlyle, 1982; Lahey & Saal, 1981; Sauser & Pond, 1981). The results of the present study tend to support the predictive power of the cognitive complexity construct with respect to appraisal effectiveness (Cooper, 1981; DeNisi, et al. 1984; Dunnette & Borman, 1979; Ilgen & Feldman, 1983; Landy & Farr, 1980).

In many respects, the present study was similar to those studies which did not find a significant effect of cognitive complexity on halo. As did the other studies, the present study also used the Role Construct Repertory (REP) grid as a measure of cognitive complexity (Bieri, et al., 1966). Moreover, the descriptive statistics for cognitive complexity scores compare well with the norms provided by Schneier (1979), and the descriptive data reported in other studies<sup>†</sup>. For example, 96 college subjects in Sauser and Pond's (1981) study had a range of 66 to 195, with the median at 96. Similarly, the 70 subjects in this study had a range of 66 to 188, with the median at 103 on cognitive complexity scores (lower scores indicate complexity). However, an improvement in this study was that the REP grid had roles (e.g. male teacher) and constructs (e.g. critical-uncritical) that gave the measure greater face validity for use in a performance appraisal task.

A major refinement in this study that may have brought about the positive result was the measure of halo effect. In previous studies that reported negative results, halo effect was indexed by the standard deviation of dimensional ratings. The inappropriateness of this index of halo has been discussed by Pulakos, et al., (1986), who noted that the majority of published studies they scrutinized, used the questionable standard deviation as an index of halo. As recommended by Pulakos et al., halo effect in the present study was measured using correlations. Additionally, as described in chapter 3, halo was measured by the correlation between ratings for dimensions on which necessary information was either supplied or withheld, thus, creating a situation where halo was highly

---

<sup>†</sup> Not all studies report the descriptive statistics for the cognitive complexity scores of their samples

probable. Subjects were not forced into rating every dimension, but they did so on the basis of what was known about the ratee. However, the finding here should be interpreted cautiously because, for exploratory purposes, only one ratee vignette (Rating Task B) was used in assessing the halo effect.

#### **D. SUMMARY OF THE FINDINGS AND CONCLUSIONS**

This investigation began with the primary objective of examining how appraisal purpose, cue dimensionality, and cognitive complexity affect the subjective importance, utilization, and integration of information in judgment. The task environment was performance judgment of teaching in higher education. The effects of purpose and cue dimensionality were observed on subjective importance and utilization of trait and role information. The use of cue integration strategies was examined in relation to purpose and cognitive complexity. Exploratory analysis focused on the measurement of good instructor schema profiles, and on the effect of cognitive complexity on halo in performance ratings. The findings and the conclusions that can be drawn from these findings, are presented below.

1. There was no appreciable effect of appraisal purpose on subjective importance and utilization of trait and behavior information in performance judgment of teaching. Nor did purpose bear an influence on how performance information was mentally combined. These findings suggest that appraisal purpose does not have a direct impact on raters' mental processing of information in performance judgment. Its effect, however, may be mediated by other factors because purpose did interact with the cues.

2. Cue dimensionality had a strong impact on both subjective importance and utilization of information; the effect was stronger in information utilization than in importance ratings. It is the cues or information that provide the data on which judgments are based. As a result, the nature of information seems an important factor affecting the utilization of information in performance judgment. Cue saliency may be a function of information content. However, the impact of cues may vary with purpose because an interaction between cue dimensionality and purpose was observed.
3. Appraisal purpose and cue dimensionality conjointly influenced subjective importance and utilization of trait and behavior information. On the average, raters valued (subjective importance) and utilized trait information more than behavior information in judgments required for a summative purpose such as personnel decisions. For formative judgments, where the rating provided feedback on the quality of teaching, raters utilized behavior information more than trait information. This pattern of information utilization suggests that saliency of information is a function of purpose as well, and that appraisal purpose has an effect on raters' cognition but but through the schema it activates.
4. Information dimensions were weighted differently for different purposes, but presentation clarity, an aspect of behavior information, was given the most attention in both summative and formative judgments. Grading-marking, the other dimension of behavior information was least weighted in summative judgment. Enthusiasm, a trait dimension, was important but not to the



exclusion of other dimensions of information. These findings suggest that student evaluation of instructors is rational, and may not be necessarily affected by rewards in terms of grades, or haloed by an instructor's enthusiasm.

5. There was consistency between what raters subjectively considered important information and their utilization of similar information in making the rating judgments. This finding suggests that people's judgments are consistent with their subjective values, and that people do have the ability to report what factors they may consider in making judgments.
6. Compared to the cognitively simple raters, complex raters made use of varied strategies in mentally combining dimensions of information related to performance. Although the subjects mainly used compensatory strategies, the complex individuals used noncompensatory strategies as well. This finding indicates that cognitive complexity, the disposition to view multidimensional stimuli in a differentiated manner, a development construct, affects the use of strategies in mentally integrating performance information.
7. A lower degree of halo effect was observed in the ratings of cognitively complex subjects. Given their disposition to view multidimensional stimuli in a differentiated manner, the cognitively complex individuals seem to be less prone to halo error. Hence, cognitive complexity may also affect the psychometric characteristics of performance ratings, especially when halo is indexed by correlational techniques.

8. The validity of a schema measure of a good instructor profile was endorsed. As expected, items not related to the schema profile turned out to be nondiagnostic. This finding indicates that a Bayesian procedure for quantitatively measuring stereotypes (McCauley, et al., 1978, 1980) has the potential to be developed as a measure of schema.

Like all research, the present investigation has some strengths and limitations. The conclusions drawn above should therefore be entertained in light of the strengths and limitations of the study discussed in the next part.

#### **E. STRENGTHS AND LIMITATIONS OF THE STUDY**

In order to test hypotheses of theoretical significance, it was necessary to exercise control over the information provided to the subjects. Therefore, an experimental procedure was chosen so that some of the extraneous variables (e.g. the amount of information) could either be controlled. Consequently, a strength of the study was its internal validity. However, internal validity of a study may compromise its external validity. The controlled setting limits the generalizability that can be given the results. In a normal appraisal setting the raters may have more information about the instructor, usually obtained from many sources and different occasions. Nevertheless, as a means for investigating questions of theoretical significance, a simulated task offered certain advantages. If the present study were to be conducted *in vivo*, only a crude examination of the raters' information processing would have been possible, because the nature and amount of information is usually difficult to control in a normal situation.

The concern for external validity of the present study need not be a serious one, because findings obtained in laboratory studies of information integration are not only meaningfully related to, but also, under certain conditions, predictive of real life behavior (Levin, Louviere, & Schepanski, 1983). Levin et al. have reviewed evidence of external validity of laboratory studies on juror judgments, occupational choice, and hiring decisions. They concluded, "The controlled laboratory setting is then the ideal place to study how the relevant factors are evaluated and integrated to determine judgment and decisions that affect our daily lives" (p. 191).

Besides, psychological research can pursue two different goals: the goal of predicting behavior, and the goal of understanding behavior. As these goals may be incompatible, "attempts to pursue both goals within one study will usually require compromises in procedure that compromise the results, rendering them unsatisfying for either goal" (Anderson, 1981, p. 91). Moreover, the issue of external validity should be raised in relation to the purpose of research (Mook, 1983). It should be reiterated that the focus of this study was on information processing, and not on estimating population values on student evaluation of teaching. Description and understanding, rather than prediction, was the primary goal. The basic intention was to clarify the effect of appraisal purpose, cue dimensionality, and cognitive complexity on the formation of rating judgments. However, if the findings here may vary from an actual performance evaluation, one may dismiss the results as lacking in external validity, or adopt a more progressive philosophy and search for conditions that would account for the differences (cf. Simon, 1968).

The confidence in the results of a study depends on the reliability of the instruments used to collect the data. The reliabilities of the instruments in this investigation were respectable and ranged from moderate (.7) to high (.94).

It should be pointed out that the cues and the purpose conditions in this study were sampled systematically in order to include those of most significance, and the analysis followed a fixed effects rather than a random effects model. Therefore, the results here may be due to the specification of the variables, and generalization to other purposes and cues would require due caution.

## F. IMPLICATIONS

The findings in this study have several implications. These implications relate to the theoretical points of view that informed this investigation, relate to research and policy on performance judgment, and relate to certain issues in performance appraisal and human judgment in general.

The results indicate that appraisal purpose and cue dimensionality interactively influence the subjective importance and utilization of performance information. This finding has implications for some of the cognitively oriented theoretical models of performance judgment (Ilgen & Feldman, 1983; Landy & Farr, 1980). Although these models seem credible, the effect of purpose on raters' cognition is not as direct as suggested by the authors. The effect of purpose on a rater's cognition could be viewed in terms of schematic processing. Purpose may operate only as a priming stimulus activating particular schemata for processing performance information (DeNisi, et al., 1984). Furthermore, in

future theorizing, the effect of purpose could be discussed in relation to cue saliency.

Cue saliency in past research on human judgment has been manipulated by varying perceptual features, frequency, and the order of information. In the current study, trait and behavior cues had an effect on both information valuation (subjective importance) and utilization. Therefore, theoretical developments addressing cue saliency should consider information structure in terms of semantic dimensions as important influence on information use in judgment. Information content may determine how concretely the cues could be represented mentally.

The concept of schema utilization as a heuristic for organizing and retrieving information from memory is well established in the cognitive and social psychological literature. Schema and the like concepts of stereotypes and implicit personality theory have become key constructs in performance appraisal models as well (DeNisi et al, 1984; Cooper, 1981; Ilgen & Feldman, 1983). Schema are assumed to exist and operate as theorized. The present study is perhaps the first to provide some evidence for this assumption. From the exploratory analysis on schema measurement and utilization, we have tentative evidence that schema profiles could be quantified, and that appraisal purpose appears to activate specific schemata. It is speculated that schema, activated by appraisal purpose, guide the utilization of performance information. There is evidence that people seek information mostly to confirm their theories than to explore others (Shaklee & Fischhoff, 1982). If performance judgment is schema driven, causes of systematic biases and errors in performance appraisal may be better understood as a product of schematic processing.

The cognitive effect of appraisal purpose, even if mediated through schematic processing as speculated, has implications for accuracy in performance judgment. The results here show that raters may actually require different information to make evaluations for different purposes. Therefore, it will be necessary to ensure that raters have access to the appropriate information to make accurate appraisals. Accuracy in performance judgment may be dependent on accessibility of relevant information. As there are many purposes of performance evaluation documented by Bernardin and Beatty (1984), the types of information utilized for different purposes should be clarified for the formulation of prescriptive principles, and for the design of appropriate rating instruments.

As a study of raters' cognition, this investigation focused on information valuation (subjective importance), utilization, and integration. Information integration theory (Anderson, 1981) provided this perspective. Systematic differences emerged between individuals and groups, on how they valued and utilized information in arriving at their rating responses. As a result, analyzing how rating judgments are formed seems more informative than analyzing the final ratings or products, as done in the past. Lopes (1982), for example, has shown that judgments can be improved if one can identify how the judgments are produced. Thus, studying how rating judgments are formed, specifically, identifying information utilization and integration strategies, may accumulate knowledge for prescriptions to improve rating judgments. From such knowledge might flow implications for rater training, which in the past has mainly focused on how raters could avoid psychometric errors (Bernardin & Pence, 1980; McIntyre, et al., 1984).

If systematic biases can be identified in judgment strategies, specifically in

information valuation, utilization, and integration, they might be reduced, if not eliminated (Fischhoff, 1982). Investigating the impact of mental models, schemata, and implicit theories on performance evaluation will provide a better understanding of the cognitive biases or cognitive distortions affecting judgments. Moreover, it may contribute toward the question as to whether factor structures of performance rating instruments reflect implicit theories or dimensions of actual teaching behaviors (Abrami, Leventhal, & Dickens, 1981; Larson, 1979; Whitely & Doyle, 1976).

The results of this study also tend to support the validity of cognitive complexity theory, given the finding that complex raters made use of both compensatory and noncompensatory strategies. The results of the exploratory analysis of the effect of cognitive complexity on performance appraisal provide some support for the cognitive compatibility proposition (Schneier, 1977), for halo in ratings was stronger in cognitively simple than in cognitively complex raters. Thus, the importance of cognitive complexity of raters seems to be rightly stressed in the process oriented models of performance appraisal (Cooper, 1981; DeNisi et al. 1984; Landy & Farr, 1980). However, this conclusion is tenuous because of the use of a single rating task in assessing halo. Although further research is needed to reinforce the results in this study, the conclusion of some researchers that cognitive complexity may not be a useful variable in performance appraisal research seems premature (Bernardin, et al., 1982; Cardy & Carlyle, 1982; Lahey & Saal, 1981; Sauser & Pond, 1981).

If cognitive complexity of the rater is an important variable in performance judgment, it may be used in rater selection. One could also attempt to enhance

raters' cognitive complexity through rater training. Although growth in a person's cognitive complexity may not be easy to achieve, there is some evidence that it is possible (Sprinthall & Thies-Sprinthall, 1983).

Moreover, Pitz and Sachs (1984) pointed out that there has been little integration of research on judgment and decision processes with developmental aspects, such as moral development (Rest, 1979), which may affect a person's ability to treat multidimensional stimuli. The findings in the present study imply that the difficulty people have in using different strategies may well be a result of their cognitive complexity, a developmental construct (Bieri, et al., 1966; Kelly, 1955).

The patterns of information use found in this study address certain controversial issues concerning the validity of student evaluation of instructors. Arguments have been made for and against the validity of student evaluations (Centra, 1979). The chief concern among instructors is that students may be overly biased by how the instructors grade and mark students work, that is, the "grading-satisfaction hypothesis" (Cohen, 1981; Marsh & Overall, 1980). In the current study, grading-marking was not given the highest priority, either in formative or in summative evaluation. In fact, grading-marking was the least important dimension in summative evaluation. Nor was enthusiasm, a factor in "educational seduction," utilized to the exclusion of presentation clarity. Hence, the dismay among instructors about the validity of student evaluations seems to be overstated.

The finding of differential use of trait and behavior information in



performance judgment raises some concern regarding evaluation of teaching at the school level. Educators advocate that teacher evaluation ought to serve summative and formative function at the same time, acknowledging the difficulties in achieving the functions from organizational behavior perspectives (Darling-Hammond, Wise & Pease, 1983; Millman, 1981). If information is utilized in relation to the purpose of appraisal, then conducting an appraisal for both summative and formative purpose at once poses a dilemma from a cognitive perspective as well. Neither purpose may be well served because a supervisor might provide feedback on behaviors, but consider traits equally or even more important in deciding the summative rating. Achieving both functions in one judgment means greater cognitive load. Researchers have found that simplifying heuristics which may cause erroneous judgment are used more often when task demands are difficult (Payne, 1976, 1982). Thus, the twin function of teacher evaluation may be difficult to accomplish without increasing the chances of less reasoned and unbiased evaluations.

The consistency between subjective importance of information and utilization of similar information is a finding that addresses an issue in human judgment in general. Some researchers have reached pessimistic conclusions concerning peoples' ability to report the importance of information in judgment (Nisbett & Wilson, 1977; Slovic & Lichtenstein, 1971; Schmitt & Levine, 1977). Other researchers have questioned the validity of such conclusions (Ericson & Simon, 1980; Surber, 1985). The findings in this study raise doubts about the validity of the extreme conclusions concerning people's inability to report the importance of information in their own judgment. Another implication of this finding is that if people are trained for performing performance evaluation, the effect of the training is likely

to transfer positively, provided that such training has an impact on raters' subjective values, schemata, or implicit theories.

The present study transferred theories and literature from the psychology of person perception, judgment and decision making to performance appraisal, and the results confirmed certain hypothesized relationships. Thus, the theories and research on person perception, and judgment and decision making may be useful in other contexts as well, where evaluative judgment is called for. The results here also reinforce the idea that an attempt to understand raters' cognition, particularly information utilization and integration, is a potentially rich route to unravel some of the causes of problems in performance judgment, or even in other areas where rating judgments are required. Further, "lens modeling", specifically, policy capturing methodology may be a valuable tool in such research, and can be used for hypothesis testing.

The findings and the implications which have been drawn from the findings, suggest additional inquiry. Some areas of additional research are discussed in the next section.

## **G. DIRECTIONS FOR FURTHER RESEARCH**

As discussed by Bernardin and Beatty (1984), performance judgment serves several purposes. The present study used only two of the purposes in a fixed effects model. Therefore, how other appraisal purposes interact with cue dimensionality and influence utilization of information is yet to be determined. Similar theoretic perspectives as used in this study may be drawn upon to

develop testable propositions. The present study addressed cue dimensionality from the perspective of role and person schemata, and the results were in the predicted direction. Other theoretical bases for presenting performance information should be explored as well. For example, a further policy capturing study could be developed from attribution theory to study the utilization of consensus, consistency, and distinctiveness information (Kelley, 1971).

As judgment strategies are quite sensitive to changes in task format, content, and demand (Einhorn & Hogarth, 1981), whether a greater number of information dimensions would yield different results remains to be determined. Few studies have been done on the effect of varying numbers of cues in performance judgment (Anderson, 1977). Ideally, the design will have to be such that the amount of information is not confounded with the number of information dimensions. As information load affects judgment processes (Payne, 1976, 1980, 1982), further research may explicate the interaction between appraisal purpose and cognitive load, because performance judgment is often conducted under time pressures (DeNisi, et al., 1984).

The order in which information is presented usually has primacy and recency effects on judgment. In the current study, the order effect was neutralized by rotating (Latin squares) the information dimensions in the ratee profiles. Information presented first may also have effects on subsequent information. We may therefore determine if trait cues dilute the diagnostic value of behavior cues, and vice versa. Such research will have implications for theory concerning cue saliency and for the development of rating instruments.

The cognitive effect of appraisal purpose on information utilization was tested as a between-subjects factor in this study. The findings seem to imply that despite the recommendation of many educators, performing both formative and summative teacher evaluation at once may be cognitively so demanding that errors in judgment may creep in inadvertently. A further study would be to address this issue by using appraisal purpose as a within-subject factor.

The speculation that the effect of appraisal purpose is mediated through schematic processing needs empirical verification. It is suggested that schema "fill in the gaps" in the information we receive (Taylor & Crocker, 1981). Therefore, we may ask if schema and halo are the same phenomenon, or mutually exclusive but which operate simultaneously. Hence, a further study would be to compare the cognitive aspects of halo with schema utilization. Such research would be contingent upon developing a measure of schema at an individual level.

Increasing rating accuracy is a prime goal in performance appraisal. We may expect greater accuracy in the ratings if information is carefully scrutinized, compared, eliminated and weighted. Hence, a further study could examine whether rating accuracy is a function of the manner in which the cues are mentally combined by the rater. We may test the hypothesis that raters who use both compensatory and noncompensatory strategies, make more accurate ratings than those who rely merely on compensatory strategies.

This is perhaps one of the first studies to explore information integration strategies in relation to cognitive complexity. An obvious next step would be to replicate the findings in different judgment situations with varying tasks. Such

research would not only establish how well-founded is the theory of cognitive complexity, but also reveal the extent to which developmental constructs may relate to processes in human judgment.

Finally, as the results from the exploratory analysis in this investigation were positive, the Bayesian measure of schema used in this study needs to be refined through further research. The effect of cognitive complexity on halo should also be studied further using correlational techniques to index halo.

In conclusion, it is appropriate to note that if performance judgment in particular and human judgment in general is to be improved, we have to learn more about how people arrive at their judgments. The findings in this investigation speak to the need for incorporating purpose for judgment, cue dimensionality, and the construct of cognitive complexity in research and theorizing on judgment processes. An understanding of what information is utilized and when, and what factors affect the integration of that information into judgments, may provide a knowledge base from which conditions for decreasing the fallibility in our judgments could be determined.

## VI. REFERENCES

- Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne, (Eds.), *Cognition and social behavior* (pp. 33-45). Hillsdale, NJ: Erlbaum.
- Abrami, P. C., Dickens, W. J., Perry, R. P., & Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, 72, 107-118.
- Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational seduction. *Review of Educational Research*, 52, 446-464.
- Anderson, B. L. (1977). Differences in teachers judgment policies for varying numbers of verbal and numerical cues. *Organizational Behavior and Human Performance*, 19, 68-88.
- Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control processes of short-term memory. *Scientific American*, 224, 82-90.
- Balzer, W. K., Rohrbaugh, & Murphy, K. R. (1983). Reliability of actual and predicted judgments across time. *Organizational Behavior and Human Performance*, 32, 109-123.
- Fischhoff, B. (1982). Debiasing (pp. 422-444). In D. Kahneman, P. Slovic, & A. Tversky, (Eds.), *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Bartlett, F. C. (1932). *Remembering*. Cambridge, MA: Cambridge University Press.
- Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston: Kent.
- Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. *Journal of Applied Psychology*, 65, 60-66.
- Bernardin, H. J., Cardy, R. L., & Carlyle, J. J. (1982). Cognitive complexity and appraisal effectiveness: Back to the drawing board. *Journal of Applied Psychology*, 67, 151-160.
- Bieri, J., Atkins, A. L., Briar, S., Leaman, R. I., Miller, H., & Tripodi, T. (1966). *Clinical and social judgment: The discrimination of behavioral information*. New York: Wiley.

- Billings, R. S., & Marcus, S. A. (1983). Measures of compensatory and noncompensatory models of decision behavior: Process tracing versus policy capturing. *Organizational Behavior and Human Performance*, 31, 331-352.
- Birnbaum, M. H., & Stegner, S. E. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ as a function of hereditary and environment. *Journal of Experimental Social Psychology*, 17, 159-182.
- Borko, H., & Cadwell, J. (1982). Individual differences in teachers' decision strategies: An investigation of classroom organization and management decisions. *Journal of Educational Psychology*, 74, 598-610.
- Borman, W. C., & Dunnette, M. D. (1975). Behavior based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology*, 60, 561-565.
- Borman, W. C. (1978). Exploring upper limits of reliability and validity in performance ratings. *Journal of Applied Psychology*, 63, 135-144.
- Brehmer, B. (1976). Social judgment theory and the analysis of interpersonal conflict. *Psychological Bulletin*, 83, 985-1003.
- Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64, 123-152..
- Bruner, J. S. (1971). *Beyond the information given: Studies in the psychology of thinking*. New York: W. W. Norton & Co.
- Bruner, J. S., Goodnow, J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Brunswik, E. (1952). *Conceptual framework of psychology*. University of Chicago Press.
- Cadwell, J., & Jenkins, J. (1985). Effects of the semantic similarity of items on student ratings of instructors. *Journal of Educational Psychology*, 77, 383-393.
- Cadwell, J., & Jenkins, J. (1986). Teachers' judgments about their students: The effect of cognitive simplification strategies on the rating process. *American Educational Research Journal*, 23, 460-475.
- Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 35, 38-49.
- Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 3-52). New York: Academic Press.

- Cardy, R. L., & Kehoe, J. F. (1984). Rater selective attention ability and appraisal effectiveness: The effect of cognitive style and the accuracy of differentiation among ratees. *Journal of Applied Psychology*, 69, 589-594.
- Centra, J. A. (1979). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.
- Cohen, P. A. (1980). Effectiveness of student rating feedback for improving college instruction: A meta-analysis. *Research in Higher Education*, 13, 321-341.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multi-section validity studies. *Review of Educational Research*, 51, 281-309.
- Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82, 240-247.
- Connor, W. S., & Zelen, M. (1959). *Fractional factorial experiment designs for factors at three levels (Applied mathematics series, No. 54)*. Washington DC: US Department of Commerce.
- Cook, R. L., & Stewart, T. R. (1975). A comparison of seven methods of obtaining subjective descriptions of judgmental policy. *Organizational Behavior and Human Performance*, 13, 31-45.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Crocker, J. (1981). Judgement of covariation by social perceivers. *Psychological Bulletin*, 90, 272-292.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Cutting, J. E. (1987). Perception and information. *Annual Review of Psychology*, 38, 61-90.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53, 285-328.
- Davis, J. H., Spitzer, C. E., Nagao, D. H., & Stasser, G. (1978). In H. Brandstatter & J. H. Davis, & H. Schuler, (Eds.), *Dynamics of Group Decisions* (pp. 33-52). Beverly Hills, CA: Sage.



- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- Dawes, R., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.
- DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. *Academy of Management Review*, 3, 369-373.
- DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance*, 33, 360-396.
- Dickinson, T. L., & Zellinger, P. M. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. *Journal of Applied Psychology*, 65, 147-154.
- Dixon, W. J. (Ed.), (1983). *BMD biomedical computer programs*. Berkley: University of California Press.
- Dunkin, M. J., Barnes, J. (1986). Research on teaching in higher education. In M. C. Wittrock, (Ed.), *Handbook of research on teaching* (3rd. ed., pp. 754-777). New York: Macmillan.
- Dunnnette, M. D. & Borman, W. C. (1979). Personnel selection and classification systems. *Annual Review of Psychology*, 30, 477-525.
- Easterby-Smith, M. (1980). How to use repertory grids in HRD. *Human Resource Development*, 4(2), 3-32.
- Easterby-Smith, M. (1982). *The design, analysis and interpretation of repertory grids: CSML working paper*, (pp. 49). Lancaster, UK: CSML working paper, University of Lancaster.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz, (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley.
- Edwards, W., & Tversky, A. (1967). *Decision making*. Baltimore: Penguin Books.
- Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 73, 221-230.
- Einhorn, H. J. (1971). The use of nonlinear, noncompensatory models as a function of the task and amount of information. *Organizational Behavior and Human Performance*, 7, 86-106.
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 85, 395-416.

- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Process of judgment and thought. *Annual Review of Psychology*, 32, 53-88.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86, 465-485.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge: Harvard University Press.
- Epstein, S. (1980). The stability of behavior: Implications for psychological research. *American Psychologist*, 35, 790-806.
- Ericson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Evertson, C. M., & Holley, F. M. (1981). Classroom observation. In J. Millman (Ed.), *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 2, 127-148.
- Feldman, K. A. (1983). The seniority and instructional experience of college teachers as related to the evaluations they receive from their students. *Research in Higher Education*, 18, 93-124.
- Fiedler, K. (1982). Causal schema: Review and criticism of research on a popular construct. *Journal of Personality and Social Psychology*, 42, 1001-1013.
- Frey, P. W. (1978). A two dimensional-analysis of students ratings of instruction. *Research in Higher Education*, 9, 60-91.
- Frey, P. W., Leonard, D. W., & Beatty, W. W. (1975). Student ratings of instruction: Validation research. *American Educational Research Journal*, 12, 327-336.
- Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.
- Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgment. *American Psychologist*, 23, 483-496.
- Grimmett, P. P. (1984). The supervision conference: An investigation of supervisory effectiveness through analysis of participants' conceptual functioning. In P. P. Grimmett, (Ed.), *Research in teacher education: Current problems and future prospects in Canada* (pp. 131-166). Vancouver, B.C: University of British Columbia.

- Guilford, J. P., & Frutcher, B. (1978). *Fundamental statistics in education and psychology*. Tokyo: McGraw-Hill.
- Hamilton, D. L. (1979). A cognitive-attributitional analysis of stereotyping. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 53-84). New York: Academic Press.
- Hammond, K. R. (1980). *The integration of research in judgment and decision theory*. Boulder, Colorado: Center for Research in Judgment Policy, University of Colorado.
- Hastie, R. (1981). Schematic principles in human memory. In E. I. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The ontario symposium* (Vol. 1, pp. 39-88). Hillsdale, NJ: Erlbaum.
- Hawley, R. C. (1982). *Assessing teacher performance*. Amherst, MA: Educational Research Associates.
- Hays, W. L. (1981). *Statistics* (3rd. ed.). New York: Holt, Rinehart & Winston.
- Hildebrand, M., Wilson, R. C., & Dienst, E. R. (1971). *Evaluating university teaching*. Berkeley: Center for Research and Development in Higher Education, University of California.
- Hoepfl, R. T., & Huber, G. P. (1970). A study of self-explicated utility models. *Behavioral Science*, 15, 408-414.
- Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57, 116-131.
- Hogarth, R. M. (1980). *Judgement and choice: The psychology of decision*. New York: Wiley.
- Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, 90, 197-217.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference learning and discovery*. Cambridge, MA: MIT Press.
- Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. In B. M. Staw, & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 5, pp. 141-197). Greenwich, CT: JAI Press.
- Jacobs, R., Kafry, D., & Zedeck, S. (1980). Expectations of behaviorally anchored rating scales. *Personnel Psychology*, 33, 595-640.
- Janis, I. & Mann, L. (1977). *Decision making*. New York: Free Press.
- Kahneman, D. (1973). *Attention and effort*. New York: Prentice Hall.

- Kahneman, D., Slovic, P., & Tversky, A. (1982). (Eds.). *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39, 341-350.
- Kane, J. S., & Lawler, E. E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B. Staws (Ed.), *Research in organizational behavior* (Vol. 1, pp. 425-478). Greenwich, CT: JAI Press.
- Kelley, H. H. (1971). *Attribution of social interaction*. Morristown, NJ: General Learning Press.
- Kelly, G. A. (1955). *The psychology of personal constructs* (Vol. 1). New York: Norton Press.
- Kennedy, J. L. (1971). The system approach. A preliminary exploratory study of the relation between composition and financial performance in business games. *Journal of Applied Psychology*, 55, 46-49.
- Keppel, G. (1973). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall.
- Kirk, R. E. (1982). *Experimental design: Procedure for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks/Cole.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
- Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy, & J. R. Pomerantz, (Eds.), *Perceptual Organization* (pp. 55-98). Hillsdale, NJ: Erlbaum.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. Kerlinger, (Ed.), *Review of Educational Research* (Vol. 3, pp. 210-240). Itasca, IL: Peacock.
- Lahey, M. A., & Saal, F. E. (1981). Evidence incompatible with a cognitive compatibility theory of rating behavior. *Journal of Applied Psychology*, 66, 706-715.

- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72-107.
- Larson, J. R. (1979). The limited utility of factor analytic techniques for the study of implicit theories in student ratings of teacher behavior. *American Educational Research Journal*, 16, 201-211.
- Levin, I. P., Louviere, J. J., Schepanski, A. A., & Norman, K. L. (1983). External validity tests of laboratory studies of information integration. *Organizational Behavior and Human Performance*, 31, 173-193.
- Loftus, G. R., & Loftus, E. F. (1974). The influence of one memory retrieval on a subsequent memory retrieval. *Memory and Cognition*, 3, 467-471.
- Lopes, L. L. (1982). *Procedural debiasing. Technical Report No. 15*. Madison, WI: Wisconsin Human Information Processing Program.
- Leyton, M. (1986). Principles of information structure common to six levels of the human cognition system. *Information Science*, 38, 1-120.
- Marascuilo, L. A., & Levin, J. R. (1984). *Multivariate statistics in the social sciences: A Researcher's guide*. Monterey, CA: Brooks/Cole.
- Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, 75, 150-166.
- Marsh, H. W. (1984). Student's evaluation of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754..
- Marsh, H. W., & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology*, 72, 468-475.
- McCauley, C., Durham, M., Copley, J. B., & Johnson, J. P. (1985). Patients' perceptions of treatment for kidney failure: The impact of personal experience on population predictions. *Journal of Experimental Social Psychology*, 21, 138-148.
- McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, 36, 929-940.
- McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin*, 87, 195-208.
- McGreal, T. C. (1983). *Successful teacher evaluation*. Alexandria, Virginia: Association of Curriculum Development.

- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology*, 69, 147-156.
- Medley, D. (1979). The effectiveness of teachers. In P. Paterson & H. Walberg, (Eds.), *Research on teaching: Concepts, findings and implications*. Berkely, CA: McCutchan.
- Medley, D. M. (1982). *Teacher competency testing and teacher education*. Charlottesville, VI: Association of Teacher Educators and the Bureau of Educational Research, University of Virginia.
- Menasco, M. B. (1976). Experienced conflict in decision making as a function of level of cognitive complexity. *Psychological Reports*, 39, 923-933.
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89-115.
- Meier, R. S., & Feldhusen, J. F. (1979). Another look at Dr. Fox: Effect of stated purpose for evaluation, lecturer expressivness and density of lecture content on student ratings. *Journal of Educational Psychology*, 71, 339-345.
- Miller, G. A. (1956). Magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Millman, J. (1981). *Handbook of teacher evaluation*. London: Sage.
- Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw Hill.
- Mitchell, T. R. (1970). Leader complexity and leadership style. *Journal of Personality and Social Psychology*, 16, 166-174.
- Mook, D. J. (1983). In defense of external validity. *American Psychologist*, 38, 379-387.
- Murdock, B. B. (1961). The retention of individual items. *Journal of Experimental Psychology*, 62, 618-625.
- Murphy, K. R., & Balzer, W. K. (1986). Systematic distortions in memory in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. *Journal of Applied Psychology*, 71, 39-44.
- Murphy, K. R., Balzer, W. K., Kellam, K. L., & Armstrong, J. G. (1984). Effects of the purpose of rating on accuracy in observing teacher behavior and evaluating teaching performance. *Journal of Educational Psychology*, 76, 45-54.

- Naftulin, D. H., Ware, J. E. Jr., & Donnelly, F. A. (1973). The Dr. Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.
- Neisser, U., & Becklon, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7, 480-494.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.
- Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nie, N. H., (1983), (Ed.). *SPSS: Statistical package for the social sciences*. New York: McGraw-Hill.
- Nisbett, R., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not necessarily informative. In J. W. Payne, (Eds.), *Cognition and social behavior* (pp. 113-133). Hillsdale, NJ: Erlbaum.
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250-256.
- Norman, K. L. (1986). Importance of factors in the review of grant proposals. *Journal of Applied Psychology*, 71, 156-162.
- Norman, K. L., & Louviere, J. J. (1974). Integration of attributes in bus transportation: Two modeling approaches. *Journal of Applied Psychology*, 59, 753-758.
- Ortony, A. (1978). Remembering, understanding, and representation. *Cognitive Science*, 2, 53-69.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.
- Payne, J. W. (1980). Information processing theory: Some concepts applied to decision making. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 95-115). Hillsdale, NJ: Erlbaum.
- Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, 92, 382-401.
- Pedhazur, E. J. (1983). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart & Winston.

- Peterson, D., Micceri, T., & Smith, B. O. (1985). Measurement of teacher performance: A study in instrument development. *Teaching and Teacher Education*, 1, 63-67.
- Piaget, J. (1936/1970). *La naissance de l'intelligence chez l'enfant*. [The growth of intelligence in children, 1970]. Paris: Delachau et Niestle.
- Pitz, G. G., & Sachs, N. J. (1984). Judgment and decision: Theory and application. *Annual Review of Psychology*, 35, 139-163.
- Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within ratees to measure halo. *Journal of Applied Psychology*, 71, 29-32.
- Raifa, H. & Schlaifer, R. (1961). *Applied statistical decision theory*. Mass: Harvard Business School.
- Rasinski, K. A., Crocker, J., & Hastie, R. (1985). Another look at sex stereotypes and social judgments: An analysis of social perceiver's use of subjective probabilities. *Journal of Personality and Social Psychology*, 49, 317-326.
- Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis: University of Minnesota Press.
- Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in internal structure of categories. *Cognitive Psychology*, 7, 753-605.
- Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 265-303). Hillsdale, NJ: Erlbaum.
- Rumelhart, D. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer, (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum.
- Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99-135). Hillsdale, NJ: Erlbaum.
- Sauser, W. I., & Pond, S. B. (1981). Effects of rater training and participation on cognitive complexity: An exploration of Schneier's cognitive reinterpretation. *Personnel Psychology*, 34, 563-577.
- Schank, R., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.



- Schmitt, N., & Levine, R. L. (1977). Statistical and subjective weights: Some problems and proposals. *Organizational Behavior and Human Performance*, 20, 15-30.
- Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, 79, 294-309.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Schneier, C. E. (1977). Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology*, 62, 541-548.
- Schneier, C. E. (1979). Measuring cognitive complexity: Developing reliability, validity, and norm tables for a personality instrument. *Educational and Psychological Measurement*, 39, 599-612.
- Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, 20, 529-563.
- Shaklee, H., & Fischhoff, B. (1982). Strategies of information search in causal analysis. *Memory and Cognition*, 10, 520-530.
- Shaklee, H., & Mims, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 208-224.
- Shavelson, R. (1985, April). *Schemata and teaching routines: A historic perspective*. Paper presented in the symposium on "Teacher Thinking: Relationships among Schemata, Routines, and Teaching Effectiveness", Annual meeting of the American Educational Research Association, Chicago.
- Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teaching. In M. C. Wittrock, (Ed.), *Handbook of Teaching* (3rd. ed., pp. 50-91). New York: Macmillan.
- Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3-37). New York: Macmillan.
- Simon, H. A. (1968). On judging the plausibility of theories. In J. F. Staal & B. Van Rootselaar, (Eds.), *Logic, methodology and philosophy of science. The International Congress for Logic, Methodology and Philosophy of Science, III*. Amsterdam: New Holland.
- Simon, H. A. (1974). How big is a chunk? *Science*, 183, 482-488.

- Simon, H. A. (1976). Discussion: Cognition and social behavior. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 253-267). Hillsdale, NJ: Erlbaum.
- Simon, H., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, 26, 145-159.
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches in the study of human processing in judgment. *Organizational Behavior and Human Performance*, 6, 649-744.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28, 1-39.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Sprinthall, N. A., & Thies-Sprinthall, L. (1983). The need for theoretical development in educating teachers: A cognitive development perspective. In K. R. Howey & W. E. Gardner, (Eds.), *The education of teachers: A look ahead*. New York: Longman.
- Stiggins, R. J., & Bridgeford, N. J. (1985). Performance assessment for teacher development. *Educational Evaluations and Policy Analyses*, 7, 85-97.
- Stumpf, S. A., & London, M. (1981). Capturing rater policies in evaluating candidates for promotion. *Academy of Management Journal*, 24, 752-766.
- Surber, C. F. (1985). Measuring the importance of information in judgment: Individual differences in weighting ability and effort. *Organizational Behavior and Human Decision Processes*, 35, 156-178.
- Tabachnik, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. Higgins, C. Herman, & M. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 89-134). Hillsdale, NJ: Erlbaum.
- Thies-Sprinthall, L. (1980). Supervision: An educative or mis-educative process. *Journal of Teacher Education*, 21(4), 17-20.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Tomkins, A. (1979). Script theory: Differential magnification of affects. In H. E. Howes, & R. A. Dienstbier (Eds.), *Nebraska symposium on motivation* (Vol. 26). Lincoln, Neb: University of Nebraska Press.

- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281-299.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristic and biases. *Science*, 185, 1124-1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.
- Vannoy, J. S. (1965). Generality of cognitive complexity-simplicity as a personality construct. *Journal of Personality and Social Psychology*, 2, 385-396.
- Vaughn, G., & Corballis, M. C. (1969). Beyond tests of significance: Estimating strengths of effects in selected ANOVA designs. *Psychological Bulletin*, 72, 204-213.
- von Neumann, J. & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd. ed.). NJ: Princeton University.
- Wallsten, T. S. (1980). Processes and models to describe choice and inference behavior. In T. S. Wallsten, (Ed.), *Cognitive processes in choice and decision behavior* (pp. 216-237). NJ: Lawrence Erlbaum & Associates.
- Wallsten, T. S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 363-384.
- Wallsten, T. S., & Bedescu, D. V. (1981). Additivity and nonadditivity in judging MMPI profiles. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1096-1109.
- Webster, W. C. (1964). *Decision making in the employment interview*. Montreal: Industrial Relations Center, McGill University.
- Weldon, E., & Gargano, G. M. (1985). Cognitive effort in additive task groups: The effect of shared responsibility on the quality of multiattribute judgments. *Organizational Behavior and Human Decision Processes*, 36, 348-361.
- Wexley, K. N., & Klimoski, R. (1984). Performance appraisal: An update. *Research in Personnel and Human Resources Management*, 2, 35-79.
- Whitley, S. E., & Doyle, K. O. (1976). Implicit theories in student ratings. *American Educational Research Journal*, 13, 241-253.

- Williams, K. J., DeNisi, A. S., Blencoe, A. C., & Cafferty, T. P. (1985). The role of appraisal purpose: Effects of purpose on information acquisition and utilization. *Organizational Behavior and Human Decision Processes*, 35, 314-339.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw Hill.
- Wright, P. (1974). The harrassed decision maker: Time pressures, distractions and the use of evidence. *Journal of Applied Psychology*, 59, 555-561.
- Wyer, R. S., & Srull, T. K. (1981). Category accessibility: Some theoretical and impirical issues concerning the processing of social stimulus information. In E. Higgins, C. Herman, & M. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 161-197). Hillsdale, NJ: Erlbaum.
- Zedeck, S., & Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. *Journal of Applied Psychology*, 67, 752-758.
- Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational Behavior and Human Performance*, 18, 269-294.

## VII. APPENDIX

### A. GLOSSARY

***appraisal purpose*** - the function which a performance judgment is intended to serve. There are two common functions of performance evaluation of a university instructor: summative and formative. Summative judgment is used in making personnel decisions such as promotions. Formative judgment provides feedback to an instructor on his/her quality of teaching. In this study formative judgment was an expression of the need to improve, and excluded guidance and recommendations to the instructor for improvement.

***central tendency error*** - rating toward the middle of the scale.

***cognitive complexity*** - a person's disposition to view behavior in a multidimensional manner, as measured by a modified version of the Role Construct Repertory grid (Bieri, et al., 1966).

***compensatory strategy*** - a category of mental strategies of integrating different dimensions of information by trading-off between dimensions using an additive or averaging rule. The amount of variance explained by the linear component in the regression model indicated the use of a compensatory strategy (Billings & Marcus, 1983; Einhorn, 1970, 1971; Weldon & Gargano, 1985).

***cue dimensionality*** - the nature of an item of information presented in a performance profile in terms of a trait or behavior. A trait dimension comprised an item of information concerning a personality characteristic; a behavior dimension comprised an item of information concerning a role (teaching) behavior.

***decomposed rating*** - a numerical rating on one of the separate seven point interval scales included in rating Task B.

***good instructor schema*** - the abstract mental distribution of behavior and trait attributes in the prototype of a good university instructor. The good instructor schema was measured by a Bayesian procedure, adapted from the stereotype measure outlined by McCauley et al. (1978, 1980). The attributes or the characteristics encoded in the good instructor schema were indexed by the diagnostic ratios in the schema measure.

**halo effect** - an effect of the perceived similarity between dimensions of information that causes the similarity in decomposed ratings. In the present study, halo was indexed by the correlation between decomposed ratings of separate rater groups.

**information integration** - the mental combination of items of information in the mind into a final judgment, indexed by a compensatory or a noncompensatory strategy.

**information utilization** - the weight assigned to dimensions of information when formulating the rating judgments. The regression weights in the policy capturing analysis portrayed a subject's information utilization policy. (Slovic & Lichtenstein, 1971; Zedeck & Cascio, 1982; Zedeck & Kafry, 1977).

**leniency error** - rating on the side of leniency or favourableness.

**noncompensatory strategy** - a category of mental strategies of interactively combining different dimensions of information by determining cut-off levels and using multiplicative rules. The amount of variance explained by the nonlinear component in the regression model indicated the use of a noncompensatory strategy (Billings & Marcus, 1983; Einhorn, 1970, 1971; Weldon & Gargano, 1985).

**performance profile** - a profile description of a hypothetical university instructor comprising items of performance related information in terms of traits and behaviors.

**rating judgment** - an overall numerical rating given to a profile description of a hypothetical university instructor on an 18 point interval scale, indicating suitability for promotion or need for improvement.

**stringency error** - rating on the side of severity or unfavourableness.

**schematic processing** - schema based interpretation and utilization of information (Taylor & Crocker, 1981)

## B. IMPORTANT INFORMATION MEASURE

### 1. For Summative Condition

Different types of information can be obtained in order to make an evaluation of a university instructor. The information may reflect different dimensions listed below. In making your rating judgment, you may like more information on some dimensions than others. Because your ratings will be used in making *promotion decisions*, indicate for each dimension how important will it be for you to receive information of a particular type. Circle a number on the scale provided to the right of each dimension. On these scales 1 = least important, 4 = important , and 7 = most important.

---

Planning, preparation	1	2	3	4	5	6	7
Enthusiasm	1	2	3	4	5	6	7
Lecture presentation	1	2	3	4	5	6	7
Sociability	1	2	3	4	5	6	7
Resourcefulness	1	2	3	4	5	6	7
Grading, marking	1	2	3	4	5	6	7
Leadership	1	2	3	4	5	6	7
Communication	1	2	3	4	5	6	7
Warmth	1	2	3	4	5	6	7
Research activity	1	2	3	4	5	6	7

---

## 2. For Formative Condition

Different types of information can be obtained in order to make an evaluation of a university instructor. The information may reflect different dimensions listed below. In making your rating judgment, you may like more information on some dimensions than others. Because the purpose of your rating is to express a need for improvement - *provide feedback*, on the quality of teaching, indicate for each dimension how important will it be for you to receive information of a particular type. Circle a number on the scale provided to the right of each dimension. On these scales 1 = least important, 4 = important , and 7 = most important.

---

Planning, preparation	1	2	3	4	5	6	7
Enthusiasm	1	2	3	4	5	6	7
Lecture presentation	1	2	3	4	5	6	7
Sociability	1	2	3	4	5	6	7
Resourcefulness	1	2	3	4	5	6	7
Grading, marking	1	2	3	4	5	6	7
Leadership	1	2	3	4	5	6	7
Communication	1	2	3	4	5	6	7
Warmth	1	2	3	4	5	6	7
Research activity	1	2	3	4	5	6	7

---



## C. PERFORMANCE RATING TASK A

### 1. For Summative Condition

There are 27 profiles<sup>†</sup> of instructors presented here, one per page. You are asked to rate each one of these profiles on the scale at the bottom of the page. Try not to compare one profile with another - it is important that you rate each profile on its own merit. Do the ratings on *your subjective criteria*, and use the same criteria for all the profiles.

Each profile is comprised of observations made on 4 dimensions related to teaching. The observations are recorded at three levels: below average, average, and above average. It is suggested that you complete the rating of these profiles in one session. You may take as long as you wish.

*It is important that you keep in mind the function your rating will serve. Remember that your rating is required to make **promotion decisions** on the instructors whose profiles are presented here. Because promotions are crucial decisions affecting the institution as well as the individual, evaluative ratings become imperative. In considering these instructors for promotion, the heads of the departments and the deans will use your ratings in making their decisions. Promotion to a higher rank means granting pay increases and perhaps tenure. Therefore, you are asked to evaluate these instructors very thoughtfully.*

Please turn over the page and begin the **Rating Task A**.

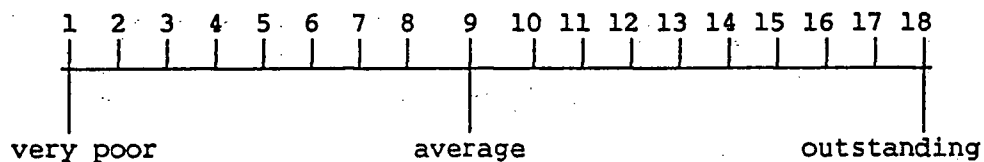
---

<sup>†</sup> Only one is included here.

Instructor P13

Information Dimension	<u>Observation Recorded as "XX"</u>		
	Below Average	Average	Above Average
enthusiasm	XX	..	..
presentation clarity	..	XX	..
resourcefulness	..	..	XX
grading and marking	XX	..	..

How suitable is this instructor for PROMOTION to a  
higher rank? Circle a point on the scale below :



please turn over to the next page..

## 2. For Formative Condition

There are 27 profiles† of instructors presented here, one per page. You are asked to rate each one of these profiles on the scale at the bottom of the page. Try not to compare one profile with another - it is important that you rate each profile on its own merit. Do the ratings on *your subjective criteria*, and use the same criteria for all the profiles.

Each profile is comprised of observations made on 4 dimensions related to teaching. The observations are recorded at three levels: below average, average, and above average. It is suggested that you complete the rating of these profiles in one session. You may take as long as you wish.

*It is important that you keep in mind the function your rating will serve. Remember that the main purpose of your rating is to express a need for improvement or **provide feedback**. Evaluative ratings provide the instructors information on their effectiveness. The ratings will not be seen by the heads of the departments or any one else, and will not affect pay or tenure of the instructors. However, the general evaluative feedback you will provide may lead the instructors to improve their performance for the benefit of other students. As instructors need evaluative feedback to self-improve, you are asked to evaluate these instructors carefully.*

Please turn over the page and begin the **Rating Task A**.

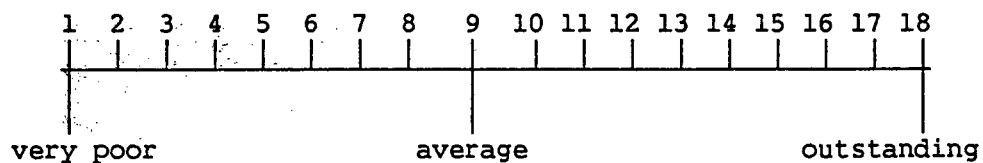
---

† Only one is included here.

Instructor F24

Information Dimension	<u>Observation Recorded as "XX"</u>		
	Below Average	Average	Above Average
grading and marking	XX	..	..
enthusiasm	..	..	XX
presentation clarity	XX	..	..
resourcefulness	..	XX	..

Evaluate this instructor's performance. In order to provide him/her some FEEDBACK, circle a point on the scale below:



please turn over to the next page..

### 3. Coding and Rotation of Profiles

	CODING				ROTATION			
	A	B	C	D				
1	0	0	0	0	A	B	C	D
2	1	2	1	1	B	C	D	A
3	2	1	2	2	C	D	A	B
4	0	0	2	2	D	A	B	C
5	1	1	0	0	A	B	C	D
6	2	1	1	1	B	C	D	A
7	0	0	1	1	C	D	A	B
8	1	2	2	2	D	A	B	C
9	2	1	0	0	A	B	C	D
10	0	1	0	0	B	C	D	A
11	1	0	1	1	C	D	A	B
12	2	2	2	2	D	A	B	C
13	0	1	2	2	A	B	C	D
14	1	0	0	0	B	C	D	A
15	2	2	1	1	C	D	A	B
16	0	1	1	1	D	A	B	C
17	1	0	2	2	A	B	C	D
18	2	2	0	0	B	C	D	A
19	0	2	0	0	C	D	A	B
20	1	1	1	2	D	A	B	C
21	2	0	2	2	A	B	C	D
22	0	2	2	2	B	C	D	A
23	1	1	0	0	C	D	A	B
24	2	0	1	1	D	A	B	C
25	0	2	1	1	A	B	C	D
26	1	1	2	2	B	C	D	A
27	2	0	0	0	C	D	A	B

Source for coding: Connor & Zelen (1959).

Values: 0 - below average, 1 - average, 2 - above average.

## D. COGNITIVE COMPLEXITY MEASURE

There are 10 different persons to rate on 10 different dimensions. The persons are listed on top and the dimensions are listed on the right of the grid below. In rating the persons, focus on a *particular individual* you may bring to mind in each case. For each person, choose the rating category (e.g. *decisive* or *indecisive*) that you feel is best for the individual you have in mind. Then use the scale corresponding the category (e.g. 1, 2, 3 for *decisive* or 4, 5, 6 for *indecisive*) to provide a rating for that person in the appropriate *cell* in the grid. When you finish, all cells in the grid will be filled up with a rating.

	Your best male teacher	Your best female teacher	A person easy to deal with	A male you admire	A female you admire	Your worst male teacher	Your worst female teacher	A person difficult to deal with	A male you dislike	A female you dislike	
											<div>1 Little 2 Moderately 3 Very</div> <div>decisive</div> <div>indecisive</div>
											<div>extrovert</div> <div>introvert</div>
											<div>considerate</div> <div>inconsiderate</div>
											<div>practical</div> <div>impractical</div>
											<div>independent</div> <div>dependent</div>
											<div>progressive</div> <div>conservative</div>
											<div>uncritical</div> <div>critical</div>
											<div>open-minded</div> <div>close-minded</div>
											<div>good-humored</div> <div>ill-humored</div>
											<div>systematic</div> <div>unsystematic</div>
											<div>1 2 3 4 5 6</div>

**E. GOOD INSTRUCTOR SCHEMA MEASURE**Directions:

Attached are *four sets* of questions concerning university instructors. Each question asks for your best estimate of a proportion. State your estimates as percentages in whole numbers between 1 and 99.

You are not expected to know the exact percentages for the questions. However, you are requested to complete every question based on your best estimate. Make sure you understand the *difference* between the four different sets of questions.

Example:

1. What percentage of students do their homework regularly? \_\_\_\_\_%
2. What percentage of GOOD students do their homework regularly? \_\_\_\_\_%
3. What percentage of ALL students who do their homework regularly are GOOD students? \_\_\_\_\_%
4. What percent of ALL students are GOOD students? \_\_\_\_\_%

Did you notice the difference? First question is about ALL students. The second question is about GOOD students only. The third question is about ALL students who really are GOOD students. The final question is not specific about doing homework but is asking for what percentage of students are GOOD students generally.

You are requested not to come back to the completed questions. Please turn over the page and begin with the first set of 10 questions.

First set of 10 questions

1. What percentage of instructors show enthusiasm? \_\_\_\_\_%
2. What percentage of instructors present the material with clarity? \_\_\_\_\_%
3. What percentage of instructors have outgoing personalities? \_\_\_\_\_%
4. What percentage of instructors are resourceful? \_\_\_\_\_%
5. What percentage of instructors plan and prepare thoroughly? \_\_\_\_\_%
6. What percentage of instructors actively participate in social work or community service? \_\_\_\_\_%
7. What percentage of instructors grade papers very well? \_\_\_\_\_%
8. What percentage of instructors show leadership? \_\_\_\_\_%
9. What percentage of instructors do a lot of professional travelling? \_\_\_\_\_%
10. What percentage of instructors are effective communicators? \_\_\_\_\_%

Please do not revise your estimates. Go onto the next page.



Second set of 10 questions

1. What percentage of GOOD instructors show enthusiasm? \_\_\_\_\_%
2. What percentage of GOOD instructors present the material with clarity?  
\_\_\_\_\_%
3. What percentage of GOOD instructors have outgoing personalities? \_\_\_\_\_%
4. What percentage of GOOD instructors are resourceful? \_\_\_\_\_%
5. What percentage of GOOD instructors plan and prepare thoroughly?  
\_\_\_\_\_%
6. What percentage of GOOD instructors actively participate in social work or  
community service? \_\_\_\_\_%
7. What percentage of GOOD instructors grade papers very well? \_\_\_\_\_%
8. What percentage of GOOD instructors show leadership? \_\_\_\_\_%
9. What percentage of GOOD instructors do a lot of professional travelling?  
\_\_\_\_\_%
10. What percentage of GOOD instructors are effective communicators?  
\_\_\_\_\_%

Please do not revise your estimates. Go onto the next page.

Third set of 10 questions

1. What percentage of ALL instructors who show enthusiasm ARE good instructors? \_\_\_\_\_%
2. What percentage of ALL instructors who present the material with clarity ARE good instructors? \_\_\_\_\_%
3. What percentage of ALL instructors who have outgoing personalities ARE good instructors? \_\_\_\_\_%
4. What percentage of ALL instructors who are resourceful ARE good instructors? \_\_\_\_\_%
5. What percentage of ALL instructors who plan and prepare thoroughly ARE good instructors? \_\_\_\_\_%
6. What percentage of ALL instructors who actively participate in social work or community service ARE good instructors? \_\_\_\_\_%
7. What percentage of ALL instructors who grade papers very well ARE good instructors? \_\_\_\_\_%
8. What percentage of ALL instructors who show leadership ARE good instructors? \_\_\_\_\_%
9. What percentage of ALL instructors who do a lot of professional travelling ARE good instructors? \_\_\_\_\_%
10. What percentage of ALL instructors who are effective communicators ARE good instructors? \_\_\_\_\_%

Please do not revise your estimates. Go onto the next page.

Fourth set - just ONE question

1. What percentage of ALL instructors are GOOD instructors? \_\_\_\_\_%

There are no more percentage questions. Please DO NOT REVISE your estimates.

**F. PERFORMANCE RATING TASK B**

Read the vignette of an instructor presented below. After reading the description, rate the instructor on the dimensions following the vignette.

Dr. T comes to class on time and is always very well prepared. Dr. T tries to present the subject matter clearly, but the students are often left confused. As a result, many students do not turn up for Dr. T's classes regularly. However, they all enjoy Dr. T's company and speeches at functions, parties, and other gatherings. Dr. T can accept criticism from students and also from colleagues. When asked, Dr. T takes up responsibilities on committees, and often does very well. Dr. T drives a Mustang, plays tennis, loves music, and seems to be a happy person most of the time.

Circle a number on the scales provided to the right of the dimensions. On these scales 1 = poor, 4 = average, and 7 = outstanding.

---

Planning, preparation	1	2	3	4	5	6	7
Enthusiasm	1	2	3	4	5	6	7
Lecture presentation	1	2	3	4	5	6	7
Sociability	1	2	3	4	5	6	7
Resourcefulness	1	2	3	4	5	6	7
Grading, marking	1	2	3	4	5	6	7
Leadership	1	2	3	4	5	6	7
Communication	1	2	3	4	5	6	7
Dependability	1	2	3	4	5	6	7
Research activity	1	2	3	4	5	6	7

---