INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
McDonald, Lynn Tracy

THE EFFECT OF DIMENSION CONTENT ON OBSERVATION OF JOB
PERFORMANCE AND RATINGS

The Ohio State University

Ph.D. 1983

University Microfilms International

300 N. Zeeb Road, Ann Arbor, MI 48106
THE EFFECT OF DIMENSION CONTENT ON
OBSERVATION OF JOB PERFORMANCE AND RATINGS

DISSER TA TION

Presented in Partial Fulfillment of the Requirements for
for the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Lynn Tracy McDonald, B.A., M.A.

* * * * *

The Ohio State University
1983

Reading Committee: Approved by
Robert S. Billings
Milton D. Hakel
Richard J. Klimoski

Advisor
Department of Psychology
ACKNOWLEDGEMENTS

I would like to express gratitude to my committee members Robert S. Billings, Milton D. Hakel, and Richard J. Klimoski. Sincere gratitude is expressed to my advisor, Milton D. Hakel, who provided a source of support and advice throughout the project. Special thanks are also due to Robert S. Billings for his valuable suggestions during the early phases of the research.

There are many individuals without whom this project would have not have been possible. I thank Bob Owens for technical support, Mike Coover for his help in data handling, Elliot Schwartz and Ginny Wemmerus for statistical advice, Sue Schechtman and Anita Bryan for serving as experimenters, and Auralee Childs for her role in creating the videotapes. Special thanks are expressed to Alisa Lyman without whose clerical support, many deadlines would have never been met.

Finally, utmost appreciation is expressed to my husband, Alain for his continuous support and encouragement throughout the dissertation project. For now, it's off to Europe.
VITA

January 26, 1956 . . . . . . . . . . Born - Houston, Texas

1977 . . . . . . . . . . . . . . . . . . B.A., Psychology
University of Texas at Austin

1977 - 1979 . . . . . . . . . . . . Research Technician,
Department of Neuropsychology,
Baylor College of Medicine,
Houston, Texas

1979 - 1980 . . . . . . . . . . . . Graduate Teaching Associate,
Department of English,
The Ohio State University

1980 - 1981 . . . . . . . . . . . . Co-director of
The Ohio State University Poll

1981 . . . . . . . . . . . . . . . . . . M.A., Psychology,
The Ohio State University

1981 - 1982 . . . . . . . . . . . . Research Associate, Organizational
Research & Development, Inc.,
Columbus, Ohio

1982 - 1983 . . . . . . . . . . . . Graduate Teaching Associate,
Department of Psychology,
The Ohio State University

AREAS OF STUDY

Major Area: Industrial/Organizational Psychology
Advisor: Milton D. Hakel, Ph.D.

Minor Area: International Business
Advisor: Lee C. Nehrt, Ph.D.

Research Interests: Cognitive processes in performance assessment and
decision-making; job stress and health; cross-cultural issues in
industrial/organizational psychology and management.
Vita (continued)

PAPERS


McDonald, T., & Hakel, M. D. *Effects of applicant race, sex, suitability, and answers on interviewer's questioning strategy and ratings.* Manuscript submitted for publication.

DISTINCTIONS

Winner of graduate student paper competition, Twenty-fifth Annual Conference of the Midwest Academy of Management, March, 1982.

Phi Beta Kappa, 1977
Psi Chi, President, 1977; Treasurer, 1976
Alpha Lambda Delta, 1974
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>I. STATEMENT OF PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Overview</td>
<td>1</td>
</tr>
<tr>
<td>Observational Goals and Schema Activation</td>
<td>5</td>
</tr>
<tr>
<td>Schema Theory</td>
<td>6</td>
</tr>
<tr>
<td>Cohen and Ebbesen's Research</td>
<td>8</td>
</tr>
<tr>
<td>Unitization of the Behavior Stream</td>
<td>11</td>
</tr>
<tr>
<td>Theoretical Limitations of Research Examining</td>
<td>13</td>
</tr>
<tr>
<td>the Perception of Ongoing Behavior</td>
<td></td>
</tr>
<tr>
<td>Extending Schema Theory to Performance</td>
<td>18</td>
</tr>
<tr>
<td>Appraisal Contexts</td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>26</td>
</tr>
<tr>
<td>II. METHOD</td>
<td>32</td>
</tr>
<tr>
<td>Overview</td>
<td>32</td>
</tr>
<tr>
<td>Subjects</td>
<td>33</td>
</tr>
<tr>
<td>Creating the Videotape</td>
<td>33</td>
</tr>
<tr>
<td>Experimental Group</td>
<td>37</td>
</tr>
<tr>
<td>Comparison Group</td>
<td>38</td>
</tr>
<tr>
<td>Incorrect Information Group</td>
<td>39</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>40</td>
</tr>
<tr>
<td>Pilot Results</td>
<td>40</td>
</tr>
<tr>
<td>Manipulation Check</td>
<td>43</td>
</tr>
<tr>
<td>Units of Perception</td>
<td>43</td>
</tr>
<tr>
<td>Unitization Patterns of Delivery and Relating</td>
<td></td>
</tr>
<tr>
<td>Instructed Subjects</td>
<td>45</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mean Ratings of Instructor’s Delivery and Relating Skills Displayed in Tape 1 and Tape 2</td>
<td>42</td>
</tr>
<tr>
<td>2.</td>
<td>Analysis of Variance: Dimension Ratings Made by Experimental Group Subjects</td>
<td>44</td>
</tr>
<tr>
<td>3.</td>
<td>Analysis of Variance: Number of Units Used by Experimental Group Subjects</td>
<td>46</td>
</tr>
<tr>
<td>4.</td>
<td>Observed Hit Rates, Chance Hit Rates, Confidence Intervals for Observed Hit Rates, and Number of Subjects Exhibiting Hit Rates of 0.0</td>
<td>48</td>
</tr>
<tr>
<td>5.</td>
<td>Means and Standard Deviations Across Three-Second Intervals and the Number of Breakpoints per Condition</td>
<td>52</td>
</tr>
<tr>
<td>6.</td>
<td>Analysis of Variance: Dimension Ratings Made by Comparison Group Subjects</td>
<td>61</td>
</tr>
<tr>
<td>7.</td>
<td>Dimension Rating Means and Variances for Experimental and Comparison Group Subjects</td>
<td>63</td>
</tr>
<tr>
<td>8.</td>
<td>Dimension Rating Means and Variances for Subjects Who Observed and Rated Different Dimensions</td>
<td>65</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>Depiction of Schematic Processing in Appraisal Contexts of Differential Dimension Assessment</td>
<td>21</td>
</tr>
</tbody>
</table>

viii
Overview

A major conclusion of Landy and Farr's (1980) review of performance rating was that great progress will not be made in this area until more research is conducted examining the cognitive processes of the rater. One variable which has potential implications for the cognitive processing of job performance is the content of the dimension upon which performance is to be assessed. Employee performance may be rated on job dimensions as diverse as ability to learn new tasks, job knowledge and skills, ability to cooperate and get along with others, and attitude toward work. For each of these dimensions, different aspects of employee performance are relevant for making a rating. The content of the dimension upon which performance is to be rated is an essential component of the rating process. Yet the effect of dimension content on appraisal processes and outcomes is an area which has received scant attention in the research literature.

When two raters observe identical sequences of job performance but are to make ratings on different dimensions, their cognitive processing of job performance may differ as a consequence of the dimension to be rated. Compare, for example, two raters, one who observes a teacher giving a lecture in order to rate the delivery skills of the teacher,
and the other who observes the same behavior in order to rate the
degree to which the teacher relates the lecture to the students' own
experiences. One way in which the two raters may differ is in terms of
the aspects of the teacher's performance to which they attend. For
example, the first rater may attend to behaviors such as eye contact,
rate of speech, loudness of voice and the use of lecture notes. The
second rater may attend more to the verbal content of the lecture and
to the examples the teacher provides to illustrate various concepts or
principles.

If the dimension on which performance is to be rated affects the
aspects of performance to which a rater attends, then a second type of
comparison may be made. Compare a rater who, prior to observing job
performance, is told what dimension of job performance is to be rated
to a rater who does not receive such prior information. Giving prior
information regarding the nature of the performance dimension may cause
the rater to attend to aspects of performance relevant to the partic-
ular dimension. The rater given no prior information, because of lack
of information, may not attend to those aspects of performance most
relevant to making a rating on the dimension in question. If this
is the case, it would be expected that ratings made by raters given
prior information would be of a higher quality compared to ratings from
raters not given prior information.

Finally, a third comparison can be made. Compare a rater who,
prior to observation of performance, receives correct information con-
cerning dimension to be rated to a rater who receives incorrect infor-
mation. If the receipt of correct information causes the first rater
to attend to dimension relevant aspects of performance, then high quality ratings should result. The second rater, given incorrect information, may attend to aspects of performance irrelevant to the dimension he/she will finally be asked to rate. As a result, these ratings may be of a low quality.

A useful heuristic for conceptualizing the situations just described is schema theory (Bartlett, 1932; Norman, 1976; Neisser, 1976). Recent research (Cohen and Ebbesen, 1979) suggests that when raters observe performance to make ratings on various performance dimensions, they activate appropriate schemata. A schema is a hypothetical cognitive structure which represents an individual's knowledge about the world and of the relationship among elements in the world (Cohen, 1981). When an individual encounters a stimulus configuration, a schema is matched against the configuration and is imposed upon its elements. Individuals attend to only that information which is relevant to the currently activated schema (Neisser, 1976; Newton, Enquist, and Bois, 1977). Further, that which is not attended to will not be stored and consequently will not be remembered (Neisser, 1976). Thus, when a rater observes performance in order to assess a teacher's delivery skills, a schema may be activated which causes the rater to process information quite differently from the rater whose goal is to assess the degree to which a teacher relates lecture material to the students' own experiences.

The current study examines the effect of performance dimension to be assessed on the cognitive processing of job performance. Schema theory is used as a heuristic to help conceptualize this relationship.
There are three questions this study attempts to answer. First, do raters who observe performance in order to rate it on different dimensions activate dimension-relevant schemata and thus attend to different aspects of performance? Second, are ratings of a higher quality, because of schematic processing, when made by raters receiving prior information compared to those made by raters receiving no prior information? Finally, do raters receiving incorrect information concerning dimension to be rated exhibit ratings of a lower quality compared to raters receiving correct information?

Schematic processing has several implications. First, raters who observe performance to rate it on one dimension should not be asked to rate performance on another dimension. Second, raters should be informed concerning what dimensions they will be asked to rate prior to observing performance. Third, research in this area has implications for appraisal contexts where consensus among raters is important (e.g., three raters, all of whom have observed performance to rate it on a different dimension, must decide whether to promote the ratee). Fourth, research examining the effect of dimension content has implications for appraisal contexts of varying purpose. For example, it seems likely that a rater observing performance for the purpose of providing feedback to an employee would activate a schema containing performance dimensions unique from the dimension contained in the schema activated by a rater observing performance for the purpose of making a promotion decision. As a result, the two raters would be expected to cognitively process an identical performance sequence in a dissimilar manner. Finally, if raters activate different schemata and/or make less
and/or make less accurate ratings as a result of dimension content, there are implications for the training of raters.

This chapter consists of a review of theory and research relevant to performance dimension content and schema activation. To gain more insight into the cognitive processes just discussed, the framework for observation of behavior proposed by Cohen and Ebbesen (1979) will be described, including a discussion of schema theory. Additionally, research relevant to the perception of behavior conducted by Newtson (1979) will be presented. In a following section, schema theory will be extended to the performance appraisal context with an emphasis on the impact of performance dimension being assessed. Finally, the proposed study will be briefly described.

Observational Goals and Schema Activation

Cohen and Ebbesen (1979), borrowing from the ideas of other authors (e.g., Norman, 1976; Neisser, 1976, Bartlett, 1932), have described a framework which can perhaps lead to greater understanding of how differences in dimension to be assessed may affect cognitive processing of ratee performance. A crucial component of their framework is the concept of schema. Cohen and Ebbesen postulate that observational goals (e.g., knowledge of dimension to be assessed) determine the selection of goal relevant schemata which, in turn, influences an observer's processing of an individual's behavior. To fully appreciate Cohen and Ebbesen's research, schema theory will first be described.
Schema Theory

The notion of schema originated in the research conducted by Sir Frederick Bartlett (1932) in the late 1920's and early 1930's. In his studies of human memory Bartlett found that when subjects were asked to reproduce stories or drawings after varying intervals of time the reproductions were often distorted and a constructive process was particularly evident. Bartlett proposed that individuals remember by organizing stimuli within the framework of their experiences and that memory should be viewed as a process of reconstruction rather than a process of recollection. The term "schema" was used by Bartlett to describe this "active organization of past recollections" imposed on stored material.

Current theorists (e.g., Taylor and Crocker, 1981; Moates and Schumacher, 1980) conceive of schemata as hierarchically organized pyramidal structures. The top of the pyramid contains abstract information and categories of more specific information are nested within the abstract information. The lowest level of the pyramid contains specific examples or instances of the particular schema (e.g., people or events). Different schemata are seen as being interconnected, particularly at the lower levels where there is greater specificity.

Neisser (1976) describes a schema as similar to a computer format which specifies that certain information, if of the proper type, is to be interpreted and other information is to be ignored. A schema is a "plan" for finding out about objects, events, and individuals and for obtaining more information to fill in the format. A major function of schemata is the encoding and representational function. Individuals
match schemata to stimuli and impose schema structure on the elements of the stimuli (Taylor and Crocker, 1981). Information as diverse as concrete objects (Moates and Schumacher, 1980), abstract patterns, (Posner and Keele, 1970), and social information has been described in terms of schema structure. Additionally, the relationship among components of schemata has been depicted as spatial (Posner and Keele, 1977), temporal (Shank and Abelson, 1977), and associative (Markus, 1977).

Schemata develop early in life. They are crude at first and become refined over time (Bartlett, 1932; Miller and Johnson-Laird, 1976; Piaget, 1954). Early schemata have their roots in biological functions. Later, schemata for concepts are developed on the basis of functional or perceptual similarities of objects, events, or individuals. As a person has more experiences, new schemata develop and existing schemata become more extensive and refined.

There are two important implications of schema theory which will become important when schema utilization in contexts of differing dimensional assessment is discussed. First, schemata determine to which information in our environment we will attend. Individuals perceive everything that is going on around them but attend only to that information relevant to the structure of the currently activated schema (Neisser, 1976; Newton, Enquist, and Bois, 1977). Second, the selective attention aspect of schemata has implications for memory. That which is not originally attended to will not be remembered (Neisser, 1976).
In this discussion of the characteristics of schemata, two important questions become evident. First, how does a particular schema or set of schemata become activated? Related to this, can more than one schema be utilized at a time? Most researchers (e.g., Hastie, 1981; Norman, 1976) make the assumption that schema utilization is necessary for any perceptual act. The individual is viewed as an active, goal-seeking, purposive creature who attempts to match schemata to events. This assumption, however, does not provide any information concerning how one schema or set of schemata is selected over another. Some insight into this question is provided by Bobrow and Norman (1975) who suggest three variables that govern schema selection: 1) purpose for engaging in a particular activity (e.g., driving home), 2) motivation, and 3) expectations regarding a particular situation. Most certainly, there are other variables which can govern schema selection.

An issue related to schema selection is the question of whether more than one schema can be activated at a time. Moates and Schumacher (1980) state that schemata are often linked to other schemata. Additionally, due to their hierarchical structure, schemata may be embedded within more general schemata (Collins and Loftus, 1975). Because of this interconnectedness among schemata, when one schema is activated, it is quite possible that other schemata or parts of schemata are activated.

Cohen and Ebbesen's Research

Cohen and Ebbesen conducted a study testing schema theory in a behavior observation context. The researchers hypothesized that differing goals for the observation of behavior cause individuals to
activate different schemata relevant to the particular observational goal. They reasoned that if individuals activate different schemata, then they should attend to different aspects of an actor's behavior. By attending to different behavior features individuals should also remember different aspects of the actor's behavior. Moreover, as a result of schematic processing, individuals who observe behavior with an information seeking goal will remember detailed aspects of behavior and individuals who observe behavior with the goal of personality analysis will be influenced by their implicit personality theories.

To test their predictions, Cohen and Ebbesen (1979) instructed thirty-two female subjects either to 1) form an impression of or, 2) learn the task performed by an actress prior to observing the actress in a series of videotaped sequences. Additionally, subjects were instructed to "unitize", or indicate the meaningful behavioral units of information contained in the tapes by depressing a button when they thought one behavior unit had ended and another had begun. After viewing the videotaped sequences of the actress engaging in both task and expressive behavior, subjects responded to questions which were either consistent or inconsistent with their observational instructions. That is, in both the impression forming and task learning groups, half the subjects responded to a memory test concerning the actress' behavior (consistent with task learning condition) while the other half completed personality ratings (consistent with impression forming condition). After a break, subjects were re instructed with the instructions they had not received previously, viewed the same
videotaped sequences, and completed the dependent measure which they did not complete after the first viewing of the videotape.

The results indicated that the subjects in the two instructional groups unitized the actress' behavior differently. Subjects given task learning instructions used a significantly greater number of units than subjects given impression formation instructions. Further analyses indicated that task learning subjects were not simply subdividing the same information into smaller units, but were attending to behavioral units with differing boundaries. Moreover, impression forming subjects were more influenced by their implicit personality theories (as measured by the similarity trait ratings made by experimental subjects and those made by another group of subjects who rated the hypothetical co-occurrence of the traits listed on the personality rating form) than were task learning subjects. This result, however, while approaching significance was not statistically significant. Finally, subjects who observed the videotapes under task learning instructions scored significantly higher on the memory test than subjects in the impression formation group.

Cohen and Ebbesen interpret their results concerning behavior unitization as supporting the idea that the two observational goals activated sets of schemata containing different features. They further contend that implicit theories of intertrait relationships seemed to act as one type of high level schema for the impression forming subjects. Finally, the authors hold that task subjects appeared to use a set of schemata that enabled them to more accurately remember the details of the actress' behavior. This study illustrates how the
cognitive set of observers affects the way in which they process information.

**Unitization of the Behavior Stream**

Before describing the extension of schema theory to performance appraisal contexts, the unitization measure used by Cohen and Ebbesen deserves discussion since it is so important to the interpretation of their results and for the current study. This dependent measure was developed by Newtson (1973; 1977) who has conducted a series of studies to gain insight concerning how individuals perceive ongoing behavior. His typical paradigm is to provide the subject with a button to press (connected to a continuous event recorder) while viewing a videotape of an actor. The subject is instructed to press the button (or "mark") when one meaningful action unit ends and another begins.

Newtson (1976) has demonstrated a variety of properties of this measure. First, the distribution of button presses over videotaped sequences appears to be reliable. Some points elicit nearly unanimous button presses by subjects while no subjects press the button at other points. Newtson has termed those points which elicit nearly unanimous button presses "breakpoints". The points at which nearly no one presses the button are termed "non-breakpoints".

Additional evidence of reliability has been demonstrated in a study, using a five-week test-retest format, in which subjects were instructed to mark off either large or small units. For both groups of subjects, the correlation between number of units marked off at the first viewing and the number marked off at the second was .87. The
same type of results were obtained for both structured and unstructured sequences of action.

A second property of the unitization measure is that level of perceptual analysis can be varied at will. When instructed to mark off either the largest or smallest actions that seem natural and meaningful, subjects in the small unit group mark off significantly more units than subjects in the large unit group. Additionally, the markings made by the two groups have a hierarchical relationship. Newtonson interprets this as indicating that the measure has sensitivity to the structure of the videotaped sequence.

Much of Newtonson's research was undertaken to determine the nature of breakpoints. When Newtonson inspected a series of breakpoints, they appeared almost comic-strip like. They seemed to summarize the action sequences contained in the videotapes while non-breakpoints seemed much more ambiguous. Because of this quality, Newtonson believes that breakpoints are the best summary points and contain more information about an action sequence than other points.

In a study concerning the information contained in a breakpoint, Newtonson first obtained consensus from a group of subjects on the location of breakpoints in a videotaped sequence. Another group of subjects then looked at slides of either breakpoints or non-breakpoints in which the order was either correct or incorrect. Subjects were asked to a) write a description of the action which was occurring, b) rate its intelligibility, and c) judge whether the slides were in correct or incorrect order. The results indicated that breakpoints yielded more description than non-breakpoints and breakpoints in correct order.
yielded more accurate descriptions than non-breakpoints in correct order. Further, breakpoints in correct order were rated as intelligible as the actual videotaped sequence. Finally, order was judged correctly 79% of the time for sequences of breakpoints and only 41% of the time for non-breakpoints. Newtson believes that these results indicate that breakpoints are the point of perceptual organization itself.

Another issue investigated by Newtson is reactivity to the measure. He postulated that evidence for processing at breakpoints, such as superior recognition could be an artifact of the measure itself. In this study, a pretest was conducted to identify breakpoints and non-breakpoints in a videotaped sequence. A group of subjects watched a videotape while marking action units while another group merely watched the videotape. Two weeks later subjects were asked to judge whether they had seen a set of breakpoints and non-breakpoints from the videotape sequence. There were no differences in recognition accuracy between the two groups of subjects. Thus, when recognition is the dependent variable it appears that the unitization measure is non-reactive.

Based on the results of these studies, Newtson, Enquist, and Bois (1977) argue that breakpoints are both the perceptual boundaries between action units and are the points in the behavior stream at which action units are defined. Individuals perceive action by discriminating and/or selecting successive definition points in the behavior stream. These definition points are also action boundary units because
when defining information for one action occurs, a point of reference is provided for its discrimination from the preceding action.

Newtson, Enquist, and Bois suggest two bases for the discrimination of action units in the behavior stream. First, actions may be defined by the achievement of distinctive states which are meaningful because they define the occurrence of a recognizable goal state. If this is true, each individual breakpoint defines an action. Alternately, actions may be perceptually defined as state to state changes exhibited by the stimulus person. If this is true, the distinctiveness of breakpoints is due to a change having occurred rather than a state having been attained.

To test these two alternatives, Newtson et al. (1977) coded feature changes of an actor, at one second intervals, using the Eshkol-Wachman movement notation. Pre-tests were conducted to determine the location of breakpoints. If the meaningful state hypothesis is true, then greater than average feature changes should occur between breakpoints and non-breakpoints than between pairs of stimulus points chosen at random. If the meaningful change hypothesis is true, then the difference between successive breakpoints should be greater than the degree of change within an action unit (controlling for elapsed time). The meaningful change hypothesis was strongly supported and additional analyses provided even more support. The authors contend that these analyses provide evidence that objective patterns of change in the behavior stream are the basis of action units.

Newtson et al. (1977) conclude that an action is defined by a change in a stimulus array. Behavior perception is a feature
monitoring process whereby the perceiver monitors a particular set of features, segmenting the behavior into parts when one or more of the monitored features change their state. Breakpoints are the points of feature change. One final contention of Newtson's which is of importance to this discussion is that expectancies or sets can affect behavior perception by altering the set of features monitored. Included here would be the job dimension on which behavior is to be assessed.

The results of Cohen and Ebbesen's (1979) study can be interpreted using Newtson's framework. The observational goals given to subjects activated different schemata for which certain behavior features were relevant. As a result, subjects in the two conditions were monitoring different sets of features. If breakpoints represent feature changes and if different features change at different rates, then differences in unitization due to observational goal would be the expected result. This is precisely the result Cohen and Ebbesen obtained.

Theoretical Limitations of Research Examining the Perception of Ongoing Behavior

Newtson's framework has several shortcomings. First, he does not have enough evidence to justify the contention that breakpoints are the points in the behavior stream at which actions are defined. The strongest statement that can be made concerning the nature of breakpoints is that they appear to carry more information than other points in the behavior stream. Second, Newtson assumes that behavior consists of a continuous series of units. This assumption is evident when he instructs his subjects to indicate when one unit of behavior has ended and another has begun. In reality, however, there may be instances in
which one behavior unit ends and another one does not begin or when multiple units of behavior occur simultaneously.

Cohen (1981) suggests a third weakness of Newtson's contentions. Newtson does not recognize that there are certain features to which an observer might attend that are static (e.g., manner of dress). More specifically, if observers perceive by monitoring changes in features, how are features which do not change processed? This shortcoming may not be as critical as it sounds if the goal of the research is to determine whether individuals are monitoring different features of performance. It seems reasonable to believe that when individuals observe performance, they monitor a set of features rather than monitoring only one feature. Even if some features monitored are static, most likely, dynamic features are also monitored. Since different dynamic features change at different rates, differences should occur in the unitization measure. This would indicate that subjects are attending to different behavior features even though some of the features they are monitoring may be static.

Another issue regarding the unitization measure is whether it is truly a reflection of the perceiver's cognitive processes. Cohen (1981) contends that although there is not an unqualified answer to this question, the unitization measure at least reflects the location of information in the behavioral stream that is encoded by the observer. She supports this with the finding from the Cohen and Ebbesen (1978) study that subjects who differed in their unitization of behavior also differed in what information they stored and retrieved.
Additionally, Newtson (1973) found a relation between size of unit subjects used and characteristics of the action sequence (e.g., smaller units were used when viewing unpredictable behavior). These findings imply that there is a relation between unitization and encoding.

Finally, regarding whether the measure truly reflects cognitive processes, Newtson contends that if actions were not perceived as cognitively discrete events, then individuals would experience behavior as gradual transitions between different actions. As a result, individuals would not be able to reliably discriminate boundaries between actions. This runs contrary to the reliability results obtained by Newtson.

A further problem related to both Newtson's and Cohen and Ebbesen's research is that the researchers have never clearly articulated exactly what aspects of performance observers attend to under the different conditions. Newtson conceives of behavior perception as a feature monitoring process, and although he has demonstrated that subjects can reliably partition units, he has not specified in a predictive sense which features of behavior observers monitor. All of his content analyses of breakpoints have been post-hoc in nature.

Cohen and Ebbesen's study suffers from similar problems. Although the researchers predicted differences in unitization patterns based on differential schema activation due to purpose, they did not clearly articulate the content of impression forming and task learning schemata. The meaning of different unitization patterns cannot be clearly discerned unless the researcher has an idea of the behavioral features being monitored.
One difference which should be noted between Cohen and Ebbesen's (1979) and Newton's (1977) research is that Cohen and Ebbesen have investigated the monitoring (via schema utilization) of both task-relevant and personality information whereas Newton has investigated only the monitoring of task-relevant information. Given the results of Cohen and Ebbesen's research, however, it seems that the processes postulated by Newton generalize to situations where behavior other than task behavior is monitored.

Despite its shortcomings, Newton's unitization measure is useful as a dependent variable. Though breakpoints may not be the points of perceptual organization itself, they do appear to reflect feature changes. Thus, if the goal of the research is to determine whether subjects are attending to different behavioral features, then the unitization measure is a reasonable choice as a dependent variable. When using the measure however, the researcher must seek to determine what performance features subjects will attend to, and based on this, make predictions regarding unitization patterns.

In the next section, the extension of schema theory to rating contexts where raters are assessing different dimensions will be discussed.

Extending Schema Theory to the Performance Appraisal Context

Extending schema theory to performance appraisal contexts may be helpful, in a heuristic sense, in understanding the possible effect of dimension being assessed on the cognitive processes of the rater. There are several reasons which support making this extension. First, schema theory provides enhanced understanding of contexts in which
behavior is to be observed and recalled and where cognitive set has an affect on these processes. Appraisal contexts where different dimensions are being assessed is clearly one context where all these components may be present. Second, schema theory has received strong support in the areas of cognitive psychology (e.g., Shank and Abelson, 1977; Tversky and Kahneman, 1977; Kintsch, 1977; Rumelhart, 1975; Neisser, 1976) and social psychology (e.g., Cohen and Ebbesen, 1979; Cantor and Mischel, 1977; Markus, 1977; Hastie, 1981). The usefulness of the theory in these areas of psychology suggests that it can be further extended to even more diverse situations such as performance appraisal. Extending schema theory to this context would place performance appraisal research within the larger study of information processing. This merging of different areas of psychology in order to achieve greater understanding is consistent with the goals of any science.

Third, as research in the performance appraisal area has moved away from being predominantly concerned with the effect of rating format on rating outcomes, psychologists in the area (e.g., Landy and Farr, 1980; Feldman, 1981) have begun to recognize that the cognitive processes of raters may be the key to understanding the diversity of results in the performance appraisal literature. Using schema theory to help understand performance appraisal situations in which differing dimensions of performance are being assessed is a step in this direction.

Fourth, studying the performance appraisal context under a schema theory framework will not only provide insight into understanding phenomena in this context but will help refine schema theory by further
specifying the units and boundaries of the theory. Finally, the schema theory extension has real world implications. If it is found that the performance dimension being assessed affects the cognitive processes of the rater, then organizational practices can be modified to take this into account.

Figure 1 is a depiction of how schematic processes are hypothesized to operate in contexts of differing dimension assessment. The components of the model are numbered from one to eight. The process begins (Component 1) when the rater has a particular dimension in mind on which to assess the ratee's performance. Knowledge of dimension to be rated is typically gained from such sources as supervisors, job manuals, knowledge of standard operating procedures, etc. The rater then searches his/her long term memory for the appropriate schema (Component 2). This represents a matching process whereby the rater attempts to match his/her conception of the performance dimension with the schema relevant to that dimension. Once the schema is identified, the rater activates it or readies it for input. During observation (Component 4), the rater attends to those aspects of performance relevant to the activated schema and observed performance is then stored within the schema structure (Component 5).

At sometime in the future, when the rater appraises performance, long term memory is searched for schemata relevant to the ratee and the performance dimension (Component 6). When located, the schema is used to recall stored performance information (Component 7). Because of schematic processing, the rater will not recall any information that was not attended to.
Figure 1. Depiction of schematic processing in appraisal contexts of differential dimension assessment.
Finally, the ratee's performance is appraised based upon recall (Component 8). In most appraisal situations, ratee performance is observed and this information is stored a multitude of times before the rater actually rates performance. Thus, performance ratings represent the accumulation of information over many schema utilization episodes.

Consider, for example, a rater who is observing a ratee to assess the degree to which the ratee gets along with coworkers. The rater may activate a "cooperation" schema which might cause him/her to attend to those situations in which the ratee is interacting with others and ignore other information that could potentially be gained through observation. Behavior which is attended to will be stored which in turn will affect memory so that performance not relevant to the cooperation schema will not be remembered.

There are several issues which must be addressed regarding schema utilization in performance appraisal contexts. First, from where do the schemata for various performance dimensions come? As mentioned previously, schemata develop through experience. Experience regarding a particular dimension may be gained in many ways. One way a schema for a particular dimension may be developed is simply through many episodes of rating individuals on the dimension. During the initial observation and rating instances, the schema may be less developed and refined. Or, observations and ratings may be based on other, conceptually similar schemata. For example, imagine a rater, assessing the dimension "ability to cooperate and get along with co-workers" for the first time. The rater having no experience assessing this dimension may likely evoke a conceptually similar schema such as one relating to
prototype situations of getting along with others in a general sense (i.e., not specifically in the work setting). Through experience of assessing this dimension and observing individuals' cooperation in the work setting, the rater's schema should become more refined and differentiated.

A second type of schema development occurs when schemata are "imposed" on raters. One way this may occur is through training. Organizations commonly initiate rater training programs where raters learn the definitions of various dimensions and the types of behaviors related to the particular dimensions. At the end of training, raters should have developed refined schemata for the various dimensions. A similar way in which schemata may be imposed on raters is through reading rating manuals developed by the organization to assist raters. Additionally, the rating forms used by the organization may contribute to schema development by providing detailed definitions and behavioral examples for the various dimensions.

Related to the issue of schema development is the nature of schema content. More specifically, when a rater activates a schema to assess performance or a particular dimension, does the rater look for signs or samples of behavior on the dimension in question? The answer to this depends on the nature of the dimension being assessed. For example, when assessing a dimension such as "job skill" it seems likely that the rater would activate a schema causing him/her to attend to samples of behavior demonstrating job skill. On the other hand, when a dimension like "job involvement" is assessed, the evoked schema would probably cause the rater to attend to behaviors which are signs of job
involvement, since job involvement must be inferred rather than observed.

A third issue regarding schema utilization in performance appraisal contexts concerns the role of motivation. More specifically, does motivation affect schema selection and utilization? Since searching long term memory takes effort, it would seem that schema selection and utilization also take effort and if a more motivated rater puts forth more effort, than he/she may be more effective as a rater. For example, a motivated rater may search more thoroughly to find an appropriate schema than would a less motivated rater. Motivation probably also plays a role once a particular schema has been selected and is activated. Since schemata are conceived of as having hierarchically related nodes, a motivated rater may go further down the hierarchy when observing performance, thereby attending to more relevant and detailed aspects of performance. The end result would be that more motivated raters make more accurate ratings.

A fourth issue concerns movement away from a particular schema once it has been activated. Consider a rater observing performance to assess a particular dimension. Once the schema relating to the particular dimension is activated, can it become deactivated and can another schema take its place while the rater is observing performance? First, as was previously stated, more than one schema can be activated at a time (which is the case when a rater observes performance to assess more than one dimension). In cases when more than one schema is activated, it may well be that the schemata vary in terms of their salience during performance observation. While observing performance, different
behaviors may be relevant to the different dimensions. While a behavior or set of behaviors is occurring that is relevant to a particular dimension, the schema representing that dimension may be more fully activated than the schemata representing the other dimensions. Throughout observation, the set of schemata may be differentially activated, and the degree to which any particular schema is activated depends upon the differential relevance of the performance being observed to the various schemata.

It is also possible that schemata irrelevant to the dimensions for which performance is being observed may become activated. One case in which this might occur is when something very unexpected happens while the rater is observing performance (e.g., the ratee gets in a fight with a coworker) causing the dimension relevant schemata to become deactivated and a new schema or set of schemata to become activated. Another case in which irrelevant schemata may become activated is when the rater is not motivated and allows impertinent schemata (e.g., vacations) to intrude.

A final issue concerning schema selection and activation in performance rating contexts relates to cases in which the rater observes performance without the knowledge of the particular dimensions on which performance is to be rated. In such cases, what type of schemata will be activated? In situations in which the rater knows that performance is to be rated, but does not know on which dimensions, the rater will probably activate schemata relevant to the dimensions of performance he/she considers most important. These dimensions may or may not be the same ones which the rater will actually be asked to assess.
A similar situation occurs when the rater observes performance without knowing that he/she will be asked to make ratings at some time in the future. When the rater is asked to rate performance, he/she will remember behavioral incidents which were relevant to the schemata which were activated when the rater had opportunities to observe performance. As in the case above, these schemata may or may not be relevant to the particular dimensions on which the rater is asked to assess performance. For example, schema irrelevance may occur when the rater activated schemata during observation which caused him/her to attend to the performance of the work group as a whole rather than to individual workers. A consequence of situations such as the ones described here is that ratings may be less valid than in situations in which the rater receives prior information concerning the dimension on which performance is to be rated. The degree to which ratings may suffer depends on the relevance of the schemata activated during observation to the dimensions on which performance is to be rated.

Conclusion

The theory and research reviewed in this chapter have many implications for appraisal situations in which raters are to assess different performance dimensions. It is proposed here that the dimension to be assessed determines the selection of dimension relevant schemata which, in turn, influences the rater's processing the the ratee's performance. First, two raters assessing different dimensions should attend to different features of performance. Second, as a result of schematic processing, ratings made by raters given prior information regarding the dimension to be assessed should be of higher quality compared to
those of raters not given prior information. Finally, raters given incorrect information concerning dimension to be rated should exhibit lower quality ratings compared to raters given correct information. The current study examines postulations such as these.

In the study reported in the next chapters, undergraduate subjects viewed a videotape of a teacher giving a lecture. Using a counterbalanced, repeated measures design, subjects viewed the videotape in order to 1) assess the teacher's delivery skills and 2) assess the degree to which the teacher related the lecture material to the students' own experiences. In a second condition, subjects viewed the videotape without receiving prior information concerning the dimensions they would be asked to assess. Subjects in the two groups unitized (Newtonson, 1977) the teacher's performance while viewing the videotape. After viewing the videotape, subjects rated the teacher's performance on the two dimensions. In a third condition, subjects were given pre-observational instructions to either 1) rate delivery, but after viewing the tape were instructed to rate relating or 2) rate relating, but after viewing the tape were instructed to rate delivery. This group did not unitize performance.

Two pilot studies were conducted in developing the videotape to insure 1) that subjects shared a common conception of the two dimensions and 2) that subjects could perceive and differentiate between the two dimensions while viewing the videotape.

There are three hypotheses. First, it is hypothesized that the three different instructional conditions (i.e., delivery, relevance, no prior information) will result in different unitization patterns.
Moreover, subjects in the delivery and relating conditions will attend to aspects of performance relevant to the dimension they are rating. Additionally, subjects in the delivery condition will use more button presses in unitizing than subjects in the relating condition. The second hypothesis is that the ratings made by subjects in the prior information conditions will be of a higher quality, compared to ratings made by subjects in the no prior information condition. Finally, subjects receiving incorrect prior information will produce lower quality ratings compared to subjects receiving correct prior information.

Giving subjects correct prior information concerning the dimension they will be asked to assess is presumed to cause them to activate appropriate schemata and thus attend to and remember dimension relevant aspects of performance. As a result, they should produce high quality ratings. Subjects not receiving prior information and subjects receiving incorrect information are not expected to evoke appropriate schemata. Because of schematic processing, these subjects will not attend to dimension relevant aspects of performance and will produce ratings of a lower quality.

An issue of importance to the current study concerns how to estimate rating quality. Several conceptual and operational definitions of rating quality have been offered (Saal, Downey, and Lahey, 1980). According to one definition quality ratings result when raters discriminate among conceptually distinct dimensions of performance. That is, ratings are not subject to halo error. Halo is usually estimated by computing intercorrelations between ratings on several performance
dimensions and its interpretation is strengthened by knowledge of true scores. Using halo as an estimate of rating quality is not appropriate in the current study because subjects made ratings on only two dimensions and true scores for the dimensions are not known.

A second definition of quality is ratings that are not consistently higher or lower than some external criterion. In other words, ratings are not lenient or severe. Again, since true scores on the two dimensions are not known, this estimate of quality is not appropriate for the current study. Third, quality ratings are free from central tendency error. That is, ratings discriminate among different ratees in terms of level of performance. This estimate of rating quality is not appropriate here since subjects rated only one ratee.

A final definition of quality is ratings which show high levels of inter-rater reliability. This index of quality is often accepted as evidence of convergent validity and is associated with small rating variances. Because true scores on the two dimensions are not known in the current study, interrater reliability will be used to estimate quality and will be operationalized as variance in ratings on the two dimensions.

The decision to have subjects view a teacher giving a lecture was made because this is an occupation with which they should all be familiar. Further, students are regularly asked to rate instructor performance at the end of a course. Thus, even though undergraduates were used as subjects, they are in essence, real-life raters of instructor performance.
There are several reasons why the "delivery" and "relating" dimensions were chosen. First, it was desired to select dimensions which have shown some validity in teacher rating and that were relevant to the lecturing aspect of teacher performance. Second, it was considered important to locate dimensions for which raters would share a common conceptualization. Research conducted by Zedeck and his colleagues (Zedeck, Jacobs, and Kafry, 1976; Harari and Zedeck, 1973) partially satisfied these two criteria. These researchers have developed a set of Behaviorally Anchored Rating Scales which exhibit validity in rating teacher performance. The set provided a starting point for selecting dimensions for which subjects should share similar conceptualizations (since BARS require a great deal of consensus in their development).

A third criterion was to locate dimensions of behavior which could be perceptably exhibited in a relatively short videotape. Fourth, it was considered important to have subjects assess dimensions which could naturally change at different rates (so that the unitization measure would be sensitive to differences in attention). A final criterion was that subjects in the current sample agree as to the behaviors composing the different dimensions, thereby indicating that subjects in the population share similar schemata. The delivery and relating dimension satisfy the above criteria.

This study attempts to deal with some of the limitations of previous appraisal research. First, it is an investigation of an important variable which has received little attention in the research literature. Second, this study is guided by theory and is process-oriented rather than outcome-oriented. Finally, predictions are made
in this study regarding what type of behavior features subjects will monitor. Cohen and Ebbesen (1979) and Newtson (1977) made no predictions regarding what features subjects would monitor.

There are, however, deficiencies in the approach used here. First, as discussed in a previous section, there is the question of whether the unitization measure accurately represents cognitive processes. Second, subjects will view a videotape as opposed to watching actual performance. Related to this, subjects will view only one performance episode compared to real world raters who have the opportunity to view a multitude of performances episodes before making a rating. However, it is felt that at this stage of investigation, the control offered by the laboratory outweighs the shortcomings in terms of external validity. A final deficiency of the approach is that subjects will view performance to assess only one dimension at a time. In organizations, performance is usually observed to make ratings on several different dimensions. Again, it is felt that at this state of investigation, greater understanding of the hypothesized processes will be achieved by starting with the most basic operationalizations possible.
CHAPTER II

Method

Overview

There were several phases involved in the experimental proceedings. At the onset, subjects were told that they would be viewing a videotape of a psychology instructor giving a lecture. Using a counterbalanced, repeated measures design, subjects in the experimental group, first viewed the videotape in order to rate either 1) the instructor's delivery skills or 2) the degree to which the instructor relates the lecture material to the students' own experiences. When the tape was completed, subjects rated the instructor on the dimension for which they had viewed the tape. After the first viewing of the tape, subjects viewed the tape a second time in order to rate the instructor on the dimension they had not previously assessed. In a comparison group, subjects received no prior information concerning the dimension they would be asked to rate. Subjects in this group viewed the videotape and rated the instructor on the two dimensions. While viewing the videotape, subjects in both groups unitized the instructor's performance. A third group of subjects were given pre-observational instructions to either 1) rate delivery but after observation were asked to rate relating or 2) rate relating but after
observation were asked to rate delivery. This group of subjects did not unitize performance.

Subjects

Subjects were 113 male and 131 female students enrolled in Psychology 100, a course in which participation in experiments is a requirement. Undergraduate students were used because of their availability as subjects. Further, it seems reasonable to use these subjects because the task which they performed is a task at which they should all be familiar since students regularly observe and rate instructor performance. Of the subjects, 91 were involved in pilot testing and 143 participated in the experiment.

Creating the Videotape

Several steps were involved in creating the videotape to maximize the probability that subjects would perceive most instructor behaviors as belonging to either one or the other of the two dimensions being assessed. As a first step, four dimensions were selected from the Behaviorally Anchored Rating Scales developed by Zedeck et al. (1976). These dimensions were delivery, interpersonal relations with students, relating skills, and organization of lecture. The four were selected as dimensions of teaching behavior which could be exhibited in a short videotape.

After selecting the dimensions, three judges (two Industrial/Organizational Psychology graduate students and a senior undergraduate who had taken many courses in Industrial/Organizational Psychology) generated as many behavioral examples as possible for each dimension. After generating the behavioral examples, one of the judges edited the
behaviors so that they conformed to a common format and eliminated redundant behaviors.

A second step was taken to assess the degree to which subjects from the population to be used in the experiment could agree in categorizing each of the behaviors as representing one of the four different dimension. Twenty-seven subjects were given definitions of the four dimensions and were instructed to categorize each of the forty-four behavioral examples (arranged in random order) as reflecting one of the four dimensions. Subjects were told that they did not have to classify any behaviors which they felt were unclassifiable. (Appendix A contains the dimension definitions and the behavioral examples.)

It was desired to have the instructor exhibit only two dimensions of performance rather than all four since, in a relatively short videotape, it would be difficult to display behavior relevant to all four dimensions. Further, since little research has been conducted concerning the cognitive processing of job behavior, it was felt that more understanding of such processes could be gained with simpler operationalizations. Subjects exhibited high levels of agreement for most behaviors in assigning them to the four dimensions. The dimension interpersonal relations with students was excluded because it would involve interactions with students which could potentially confound the unitization measure. It was desired to have the instructor display dimensions of performance that subjects would perceive as independent of each other. It might be difficult to make a short tape in which subjects perceive delivery and organization of the lecture as independent. Because delivery skills (compared to organization) are more
amenable to being displayed in a short videotape, "delivery" and "relating skills" were chosen for inclusion in the videotape. Behaviors on the two dimensions for which there was 78% or greater agreement in categorization were considered for possible inclusion in the videotape.

A 6'27" videotape was made of a psychology instructor giving a lecture. The instructor demonstrated performance relevant to both the delivery and relating skills dimensions by exhibiting behaviors which had been categorized by subjects as representing the respective dimensions. Since Newton's (1976) work has demonstrated that breakpoints (i.e., button presses) occur when subjects perceive behavior changes, the videotape was constructed so that behavior on the two dimensions occurred at different rates. That is, changes in delivery style did not occur simultaneously with a shift in the lecture from theory to relating the theory to the students' experience. It was intended that breakpoints on the two dimensions would not occur simultaneously. This was done because even if subjects in the different conditions were attending to different features of behavior, the unitization measure would not reflect this if the features changed concurrently. Additionally, behavior on the delivery skills dimension changed at a faster rate (i.e., more behavior changes occurred) than did behavior on the relating dimension.

In constructing the videotape, all efforts were made so that the instructor would be perceived as exhibiting "above average" skills in relating but "below average" skills in delivery (e.g., the instructor
related many aspects of the lecture to the students' experiences, but
displayed little eye contact, spoke in a monotone voice, and clasped
and unclasped her hands). This was done to enhance the interpretation
of the results concerning whether subjects could accurately perceive
the two separate dimensions of performance. For example, if the
instructor performed "above average" on both dimensions and subjects
gave above average ratings on the two dimensions, it would be unclear
whether the subjects were inaccurately perceiving one global (above
average) dimension of performance or accurately perceiving the two
dimensions in question.

Finally, a second performance sequence was videotaped in which
efforts were made so that the instructor would be perceived as ex­
hibiting "below average" relating skills and "above average" delivery
skills, in contrast to the previous tape. This was done to provide
additional data concerning whether the subjects in the population can
accurately perceive the two dimensions (i.e., that subjects can per­
ceive both "above" and "below" average performance on both dimensions
is stronger evidence than evidence based on the first tape only).
The second videotape was used only for manipulation check purposes.

After the two videotapes were made, data were gathered from two
groups of subjects to determine the degree to which subjects could
accurately perceive the dimensions in question. Using a counter­
balanced, repeated measures design, one group of thirty-two subjects
received dimension definitions (see Appendix B), viewed the first tape
and then rated either delivery skills or relating skills. Subjects
then received definitions of the dimension they had not previously
assessed, viewed the tape a second time and rated the instructor's performance on the second dimension. On both occasions, subjects were asked to supply behavior examples from the videotape to support their ratings. A second group of thirty-two subjects proceeded in the same manner except they rated the teacher's performance on the two dimensions in the second videotape. Analysis of subjects' ratings indicated that they could accurately perceive the two dimensions and the experiment began.

**Experimental Group**

Upon arrival to the laboratory, subjects were seated and were told that they would be watching a videotape of a psychology instructor giving a lecture (see Appendix C for Experimenter Script). Using a counterbalanced, repeated-measures design, fifty subjects were told that they were to view the videotape in order to rate either 1) the instructor's delivery or 2) the instructor's relating skills. Prior to viewing the videotape, subjects read definitions of the dimension (see Appendix D) they would be rating.

Subjects were then told that they would be unitizing the instructor's performance as they observed it (Appendix C). They then viewed a short practice tape of another instructor giving a lecture after being told to press down the button in front of them when they perceived one behavior to have ended and another to have begun. The buttons which subjects pressed were connected to a computer that records time elapsed between button presses in one-second units. Subjects were instructed to place the apparatus containing the button under the table so as not to influence each other in their button pressing.
After the practice tape was over, subjects viewed the experimental tape and unitized. When the tape was completed, subjects were given rating sheets (see Appendix E) on which to make their ratings. They were also asked to provide behavior examples from the tape to support their ratings. Subjects were then informed that they would also be assessing the dimension they had not previously assessed and proceeded in the same manner (i.e., read dimension definition, watched and unitized tape, made rating, supplied behavior examples) as before. Order of dimension assessed was randomly determined at the beginning of each session. Subjects were run in groups of two to eight. After the final ratings had been made, subjects were debriefed and thanked for their participation (see Appendix F for debriefing sheets).

**Comparison Group**

Comparison group subjects proceeded in a similar manner except they were given no prior instructions regarding the performance dimensions they would be rating (see Appendix G for Experimenter Script). The fifty-four subjects in this group were told that they would be watching a tape of a psychology instructor giving a lecture. They were told that the would be unitizing the instructor's performance as they observed it (Appendix G). Subjects first viewed and unitized the practice tape after which they viewed and unitized the experimental tape. When the experimental tape was completed, subjects read a dimension definition of either delivery or relating skills, rated that dimension, and supplied behavior examples to support their ratings. Subsequently, they read a definition of the dimension they had not previously rated, made a second rating, and supplied behavior examples.
The order in which subjects made the ratings was counterbalanced and rating order was randomly determined at the beginning of each session. Subjects were run in groups of two to eight. After completing their final ratings, subjects were debriefed and thanked for participation (see Appendix F for debriefing forms).

Incorrect Information Group

Using a counter-balanced design, the fifty-two subjects in this group were told that they were to view a videotape of an instructor giving a lecture in order to rate either 1) the instructor's delivery or 2) the instructor's relating skills. Prior to observation, subjects read definitions of the dimensions they were asked to rate (See Appendix D for dimension definitions). Subjects in this group had no reason to suspect they were being given incorrect information. They then viewed the videotape. When the tape was completed, delivery instructed subjects were told that they would instead be rating relating and relating instructed subjects were told that they would instead be rating delivery. Subjects then read definitions of the dimensions they would actually be rating and made their ratings (see Appendix E for rating forms). Subjects were run in groups of twelve to fifteen. After ratings had been made, subjects were debriefed and thanked for their participation.
CHAPTER III

Results

The results will be discussed in four sections: pilot results, manipulation checks, units of perception, and dimension ratings.

Pilot Results

Percentages of agreement were computed for each of the forty-four behavioral examples in order to assign each behavior to one of four dimensions: delivery, interpersonal relations with students, degree to which the instructor relates the lecture material to the students' own experience, and organization of lecture. It was decided that 80% was an acceptable level of agreement for assigning a behavior to a dimension. High levels of agreement were observed for most of the behaviors. However, for several behaviors, between 78% and 79% (twenty-one out of twenty-seven subjects) agreement among subjects was observed so the 80% criterion was relaxed to 78%. Behaviors for which 78% or more of the thirty-two subjects categorized as relevant to delivery and relating skills were considered for inclusion in the videotape. (The other two dimensions were not used in constructing the videotape.) Appendix H contains the percent of agreement for each behavior. Also continued in Appendix H is the mean rating and standard deviation assigned by subjects for each of the forty-four behaviors.
To determine whether subjects could accurately perceive the two dimensions exhibited in the two (experimental and "manipulation check") videotapes, the ratings made by the sixty-four subjects who received dimension definitions, observed performance, and then made ratings were analyzed. The two-between, one-within analysis of variance resulted in a Tape x Dimension Rated interaction, $F(1,60) = 112.19$, $p < .0001$, indicating that subjects rated the delivery and relating dimensions significantly differently depending on which tape they observed. Table 1 contains the means and standard deviations for this interaction. In tape 1, delivery was rated below the midpoint and relating was rated above the midpoint. The opposite results were obtained for tape 2.

A post-hoc test using Dunn's procedure indicated that the two delivery means were significantly different from each other as were the two relating means, $p < .01$ for both comparisons. It was concluded that subjects could accurately perceive the two dimensions exhibited in the tapes. Subjects' justifications of their ratings further supported this. In general, subjects justified their ratings of delivery with those behavior examples which were intended to reflect the delivery dimension of teaching performance (e.g., degree of eye contact, use of lecture notes, vocal quality). Similarly, subjects justified their relating skills ratings by citing behavioral examples which were intended to reflect relating skills (e.g., using examples the student could understand, tying the lecture to the students' experiences).
### Table 1

**Mean Ratings of Instructor's Delivery and Relating Skills**

Displayed in Tape 1 and Tape 2<sup>a</sup>

<table>
<thead>
<tr>
<th></th>
<th>Tape 1 (N=32)</th>
<th></th>
<th>Tape 2 (N=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Delivery</td>
<td>2.31&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.05</td>
<td>5.00</td>
</tr>
<tr>
<td>Relating</td>
<td>4.594</td>
<td>1.4</td>
<td>2.72</td>
</tr>
</tbody>
</table>

<sup>a</sup> In tape 1, efforts were made so that subjects perceived "above average" skills in relating and "below average" skills in delivery. Tape 2 was constructed so that subjects would have the opposite perceptions.

<sup>b</sup> 1 = Extremely ineffective

5 = Extremely effective
Manipulation Check

To determine whether experimental group subjects were accurately perceiving behavior on the delivery and relating dimensions, a one-between, one-within analysis of variance was computed using the experimental subjects' ratings on the two dimensions. Table 2 shows the results of this analysis. Like subjects in the pilot group, experimental group subjects rated the delivery and relating dimensions significantly differently. Delivery received a mean rating of 1.84 while relating received a mean rating of 4.44. No significant effects emerged due to the order in which subjects received dimension definitions and assessed performance, (i.e., observe and rate delivery, then observe and rate relating versus vice versa). It was concluded that experimental group subjects were accurately perceiving the two dimensions exhibited in the tape. Again subjects correctly justified their ratings with the behavior examples intended to reflect the two dimensions.

Units of Perception

It was hypothesized that subjects in the three unitizing conditions (delivery, relating, and no prior information) will attend to different aspects of the instructor's performance. The results relevant to this hypothesis will be presented in three parts: 1) comparisons between unitization patterns of delivery and relating instructed subjects, 2) comparisons between unitization patterns of subjects receiving no prior information and delivery and relating instructed subjects, and 3) analyses of the behavior content associated with the unitization patterns exhibited by the three groups.
Table 2

Analysis of Variance: Dimension Ratings Made by Experimental Group Subjects (N=50)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>1</td>
<td>3.24</td>
<td>2.23</td>
<td>.14</td>
</tr>
<tr>
<td>Error</td>
<td>48</td>
<td>1.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension Rated</td>
<td>1</td>
<td>169</td>
<td>150.33</td>
<td>.00001</td>
</tr>
<tr>
<td>Dimension Rated x Order</td>
<td>1</td>
<td>.044</td>
<td>.04</td>
<td>.85</td>
</tr>
<tr>
<td>Error</td>
<td>48</td>
<td>1.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Unitization Patterns of Delivery and Relating Instructed Subjects

Several analyses were performed to determine whether subjects instructed to rate delivery were attending to different aspects of the instructor's performance compared to subjects instructed to rate relating. First, if the two groups of subjects were attending to different features, subjects in the delivery condition should exhibit a greater number of units in segmenting performance since the videotape was constructed so that behavior on the delivery dimension changed at a faster rate than behavior on the relating dimension. To test this hypothesis, a one-between, one-within analysis of variance was performed on the number of button presses subjects generated while viewing the 6'27" videotape.

Table 3 contains the results of this analysis. A significant main effect due to dimension observed was obtained. Subjects used a greater number of units when observing performance to rate delivery, \( M = 17.72 \) compared to observing performance to rate relating skills, \( M = 8.16 \). No significant effects due to instruction order emerged.

The analysis of the number of button presses subjects generated is not sufficient evidence that subjects in the two conditions were attending to different performance features. It is possible that subjects in the delivery condition were dividing, into smaller subdivisions, the same units they used in the relating condition (i.e., units used in the delivery condition are subsets of units used in the relating condition). If subjects were attending to the same features of performance, the units produced in the relating condition should coincide with those produced in the delivery condition.
Table 3

Analysis of Variance: Number of Units Used by Experimental Group Subjects (N=50)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>1</td>
<td>153.76</td>
<td>1.36</td>
<td>.25</td>
</tr>
<tr>
<td>Error</td>
<td>48</td>
<td>112.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension Rated</td>
<td>1</td>
<td>2284.84</td>
<td>45.35</td>
<td>.00001</td>
</tr>
<tr>
<td>Dimension Rated x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>1</td>
<td>4.0</td>
<td>0.08</td>
<td>.78</td>
</tr>
<tr>
<td>Error</td>
<td>48</td>
<td>50.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since subjects viewed the videotape under both sets of rating instructions, it was possible to determine the proportion of unit boundaries which coincided for each subject. Using Cohen and Ebbesen's (1979) method, a "hit rate" was computed for each subject. Hit rate was defined as the proportion of unit boundaries (i.e., button-presses) produced by a subject in the relating condition that coincided with boundaries produced by the same subject in the delivery condition. The following formula was used: number of unit boundaries that coincided across the two conditions divided by the total number of boundaries used in the relating condition. Since most subjects produced a smaller number of unit boundaries in the relating condition (except for four subjects), a hit rate of 0.0 would indicate that no unit boundaries lined up and 1.0 would indicate that all relating unit boundaries coincided with units defined in the delivery condition.

Five estimates of hit rate were calculated, each based on intervals of a different size. For each estimate, the 387 second videotape was divided into continuous intervals of either 1-second, 2-seconds, 3-seconds, 4-seconds, or 5-seconds duration. A hit occurred whenever a subject produced a button press in the relating condition that occurred in the same (1-, 2-, 3-, 4-, or 5-second) interval as a button press produced in the delivery condition. If more than one button press occurred in a given interval under either of the conditions, only one button press was counted.

Table 4 contains the mean hit rate for each of the five interval sizes. Hit rate varied from .05, for 1-second intervals, to .23, for 5-second intervals. Thus, of the boundaries produced in the relating
Table 4

Observed Hit Rates, Chance Hit Rates, Confidence Intervals for Observed Hit Rates, and Number of Subjects Exhibiting Hit Rates of 0.0$^a$

<table>
<thead>
<tr>
<th>Interval Size</th>
<th>Observed Hit Rate</th>
<th>Chance Hit Rate</th>
<th>Confidence Intervals for Observed Hit Rates</th>
<th># of Subjects Exhibiting Rates of 0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-second</td>
<td>.05</td>
<td>.0026</td>
<td>.014 - .132</td>
<td>33 (66%)</td>
</tr>
<tr>
<td>2-second</td>
<td>.11</td>
<td>.0028</td>
<td>.05 - .21</td>
<td>24 (48%)</td>
</tr>
<tr>
<td>3-second</td>
<td>.13</td>
<td>.03</td>
<td>.08 - .24</td>
<td>17 (34%)</td>
</tr>
<tr>
<td>4-second</td>
<td>.17</td>
<td>.034</td>
<td>.09 - .28</td>
<td>14 (28%)</td>
</tr>
<tr>
<td>5-second</td>
<td>.23</td>
<td>.05</td>
<td>.14 - .34</td>
<td>14 (28%)</td>
</tr>
</tbody>
</table>

$^a$ N=50
condition, 5% using 1-second intervals and 23% using 5-second intervals coincided with boundaries produced in the delivery condition. Table 4 also contains the hit rates expected by chance alone, 95 percent confidence intervals around each of the chance hit rates, and the number of subjects exhibiting hit rates of 0.0 for the five interval sizes.

The observed hit rates, except the one based on 1-second intervals, are significantly greater than the hit rates expected by chance alone. However, when the confidence intervals are examined, even the upper bound of the largest interval (based on 5-second intervals) indicates that at the least, almost two-thirds of subjects' unit boundaries do not coincide. Table 4 also shows the large proportions of subjects exhibiting hit rates of 0.0.

An important issue in interpreting the hit rate data concerns which interval size is appropriate to estimate "true" hit rate. A 1-second interval may be too stringent. Even if subjects were attending to the same performance features in the two conditions, they may not be so consistent and quick in their perceptions and reaction times that they could press the button at the same exact second across two viewings of the videotape. Thus, a 1-second interval may underestimate the true hit rate.

On the other hand, a 5-second interval is probably too lenient. In constructing the videotape, efforts were made so that behaviors on the two dimensions do not change concurrently. However, because of the number of behavior changes occurring in the tape, it is highly likely that changes relevant to both of the dimensions occur within five
seconds of each other. Thus, even if subjects were attending to different performance features, a hit would be counted if changes relevant to both sets of features occurred within the same five second interval and an overestimation of true hit rate would result.

Because of the above issues, either the 2-, 3-, or 4-second interval size is probably most appropriate for estimating hit rate with the sizes varying in terms of conservativeness in estimation. Even using the confidence interval upper bound of the more lenient 4-second interval, almost three-fourths of subjects' unit boundaries do not coincide across the two conditions. Since such low levels of agreement on unit boundaries were observed, it is concluded that subjects exhibited qualitatively different unitization patterns. The hypothesis that subjects assessing different dimensions attend to different aspects of performance is supported.

Unitization Patterns of Experimental Group Subjects and Subjects Receiving No Prior Information

It was hypothesized that subjects receiving prior information concerning dimension to be rated would exhibit unitization patterns different from subjects receiving no information. Using Newton's (1973) procedure, breakpoints were determined separately for each of the three unitization groups (delivery, relating, no prior information). If subjects in each group were attending to different aspects of performance, then a different set of breakpoints should be identified for the three groups.

The following strategy (Newton, 1973) was used for each of the three sets of data. The videotape was divided into 129 continuous 3-
second intervals. The number of subjects producing a button press within each interval was determined and a mean and standard deviation was computed across all 129 intervals. A breakpoint was defined as any interval containing a number of button presses greater than or equal to the mean plus one standard deviation for button presses for the particular group. This criterion was used because the joint probability of two frequencies one standard deviation above the mean of a normal distribution is .028. That is, the chance probability of the same 3-second interval being identified as a breakpoint under two different instruction conditions is .028. In determining breakpoints for the experimental groups, data from both rounds of each subject's unitizations were used.

Table 5 contains the means and standard deviations for the number of button presses produced during 3-second intervals for each of the groups. For example, button presses were exhibited by 6.74 subjects per 3-second interval in the delivery condition. Also shown is the number of breakpoints in each of the conditions resulting from using Newton's procedure. Any 3-second interval containing twelve or more button presses in the delivery condition (M + 1 sd = 12.14), six or more button pressing in the relating condition (M + 1 sd = 5.92), and nine or more button presses in the comparison group (M + 1 sd = 8.63) was considered a breakpoint.

The next analysis addressed whether experimental group subjects exhibited unitization patterns different from subjects receiving no prior information. Expected frequencies were computed from the joint probability of a breakpoint being selected in common by chance alone in
Table 5

Means and Standard Deviations Across Three-second Intervals and Number of Breakpoints Per Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>M</th>
<th>SD</th>
<th>Number of Breakpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery</td>
<td>6.74</td>
<td>5.4</td>
<td>27</td>
</tr>
<tr>
<td>Relating</td>
<td>3.15</td>
<td>2.77</td>
<td>25</td>
</tr>
<tr>
<td>Comparison</td>
<td>4.83</td>
<td>3.8</td>
<td>23</td>
</tr>
</tbody>
</table>

\(a\) \(N=50\)

\(b\) \(N=54\)
1) the delivery group and comparison condition and 2) in the relating group and comparison condition. Chi-square analyses were used to determine whether the observed matches were greater than chance.

Of the forty-eight breakpoints defined in the relating and comparison groups combined (25 in the relating group, 23 in the comparison group), only five were common. By chance alone, 4.46 common breakpoints would be expected. The chi-square statistic ($X^2 = 0.065$) was not significant. Eleven breakpoints were common among the fifty breakpoints defined in the delivery and comparison groups combined. By chance, 4.8 common breakpoints would be expected and the resulting chi-square ($X^2 = 8.31$) was significant at the .01 level.

To gain more insight into these results, contingency coefficients were computed to determine the relationship between breakpoints identified in the comparison condition and breakpoints identified in 1) the delivery condition and in 2) the relating condition. A contingency coefficient of .25 ($p < .01$) indicated that there was a relationship between breakpoints identified in the delivery and comparison groups. For relating and comparison group breakpoints, the contingency coefficient of .02 was not significantly different from zero. This indicates that there is no relationship between the two groups' breakpoints.

---

1 The contingency coefficient is a measure of strength of association between two sets of variables. It is based on the Chi-square distribution and ranges from zero to almost one. A zero indicates no relationship and higher values indicate relationships of increasing strength. In the analyses performed here, the upper limit of the coefficient is .71.
Considering these results, the hypothesis that subjects in the comparison group will exhibit unitization patterns different from subjects in the experimental groups received mixed support. For the relating group, the non-significance of the chi-square test and contingency coefficient suggests that subjects in this group were attending to different aspects of performance compared to subjects receiving no prior information. However, for the delivery group, the chi-square statistic and the contingency coefficient were significant. This indicates that delivery instructed subjects were attending to some of the same features of performance as comparison group subjects. Since the contingency coefficient did not reach its upper limit, subjects in the two groups were also attending to some unique performance features across the two conditions.

A final analysis examined breakpoints common to the delivery and relating conditions. Since unitization data in the two conditions are not independent, only data generated in the subjects' first rounds of unitization were used. Using the same procedures outlined before, twenty-one breakpoints were defined in the delivery condition and twenty-five breakpoints were defined in the relating condition. Only two common breakpoints existed and the chi-square test ($X^2 = 1.08$) was non-significant. This result supports the previous conclusion that subjects in the delivery and relating conditions attended to different aspects of performance.

Behavior Content Associated With Unitization Patterns

To further interpret the unitization results, it is important to determine whether subjects were perceiving feature changes exhibited
in the videotape as they were intended to perceive them. It was expected that relating instructed subjects would perceive feature changes when the instructor made transitions from discussing theory to relating the theory to the students' experiences or when the instructor used examples in explaining the theory. Delivery instructed subjects were expected to monitor changes taking place in the behavioral aspects (e.g., eye contact, use of notes, gestures) of the presentation.

For comparison purposes, non-breakpoints were defined for the three groups (delivery, relating, comparison) as any 3-second interval containing a number of button presses less than or equal to the mean minus one standard deviation for button presses produced during 3-second intervals for the particular group. In contrast to breakpoints, non-breakpoints may be viewed as points where subjects do not perceive behavior changes. It was expected that non-breakpoints would contain behaviors irrelevant to the dimension subjects were assessing. Sixteen non-breakpoints were defined for the delivery group, forty-six for the relating group, and twenty-seven for the comparison group. Eight of the relating non-breakpoints were delivery breakpoints and two of the delivery non-breakpoints were relating breakpoints.

To determine what subjects were and were not attending to, the breakpoints and non-breakpoints in each of the three unitization groups was assessed. For each of the three sets of 3-second breakpoint intervals, the videotape was viewed twice. On the first viewing, each time a breakpoint interval occurred, the verbal content of the interval was written down. During this viewing, the video portion of the interval was not observed. On the second viewing, the behavior content of the
interval was recorded. The videotape was viewed six times, twice for each of the three conditions. Additionally, the entire process was repeated to insure completeness of notes taken during observation. The same procedure was followed to assess the content of non-breakpoints. Thus, the videotape was observed a total of twenty-four times, twelve times to assess the content of breakpoints and twelve times to assess the content of non-breakpoints.

Appendix J contains the verbal and behavioral content of breakpoints and non-breakpoints for each of the three conditions. For the relating condition the verbal content analysis indicated breakpoints occurred when the instructor made a transition from discussing theory to a student-oriented example demonstrating the theory (e.g., "Returning to the test-taking example, your inputs would be . . . "), made the opposite transition (e.g., "Let's get back to the theory and look at situations of equity and inequity . . . "), or used specific examples to explain concepts of the theory (e.g., "an outcome would be the grade you received"). In contrast, non-breakpoints occurred after a transition had already been made and the instructor was in the process of explaining the theory (e.g., "The second important term is outcome.") or the the example (e.g., "You both study very hard and both of you have very high inputs."). Non-breakpoints also occurred when the instructor introduced and summarized the lecture and when the instructor lost her place or appeared to lose her train of thought (e.g., "Uh . . . where was I? Just a minute . . . ").

An examination of the behavioral content revealed that in only five of the twenty-five breakpoints did physical behavior changes
occur. Typically, when a breakpoint occurred, the instructor was either looking straight ahead or was looking down at her notes and she remained in that state, displaying no other physical changes, throughout the duration of the 3-second interval. In contrast, physical behavior changes occurred in twenty-eight of the forty-six non-breakpoints. Based on the verbal and behavior content of breakpoints, it is concluded that subjects were attending to aspects of the instructor's performance relevant to the relating dimension. The content analysis of non-breakpoints suggests that subjects were not attending to performance features irrelevant to relating.

The verbal content of delivery group breakpoints was next examined. In seven of the twenty-seven breakpoints, the instructor lost her place (e.g., "Uh, where was I? Just a minute... What is the effect...") or appeared to lose her train of thought (e.g., "Uh-uh, effect on behavior of receiving positive and negative test-taking rewards"). No other consistent pattern of verbal content was discerned. Additionally, no consistent pattern of verbal content was discerned for non-breakpoints.

An examination of the behavioral content of delivery breakpoints showed that behavior changes occurred in every single one of the twenty-seven 3-second intervals. Breakpoints occurred when the instructor, for example, made changes in eye contact from looking straight ahead to looking at her notes and vice-versa, looked at her watch, shuffled her notes, leaned on the podium, and rubbed her hands together. In contrast, no behavior changes occurred in any of the sixteen non-breakpoints. Based on the verbal and behavioral content
analyses of breakpoints and non-breakpoints, it is concluded that subjects were attending to aspects of the instructor's performance relevant to the delivery dimension.

Finally, the content of comparison group breakpoints was examined. The examination of the verbal content revealed that six of the twenty-three breakpoints occurred when the instructor introduced the theory (e.g., "The major question of equity theory is . . ."), made a transition (e.g., "The last thing I want to talk about today is . . ."), or summarized (e.g., "I want to sum up the lecture and talk about . . ."). Three breakpoints occurred when the instructor lost her place or appeared to lose her train of thought. No other consistent pattern was discerned. For non-breakpoints, no consistent pattern of verbal content was evident.

Finally, the behavioral content of comparison group breakpoints was examined. In nineteen of the twenty-three breakpoints, behavior changes occurred. The changes occurring during the breakpoints were of the same type as those which occurred in delivery group breakpoints. In only five of the twenty-seven non-breakpoints did behavior changes occur. The examination of comparison group breakpoints and non-breakpoints suggests that subjects may have been attending to the structure or organization of the lecture and, like subjects in the delivery condition, to behavioral changes. This latter observation is consistent with the previous conclusion that delivery and comparison group subjects were attending to some of the same aspects of performance.
Summary

The results of the analyses of units of perception support three major conclusions. First, subjects instructed to assess delivery attended to different aspects of performance compared with subjects instructed to assess relating. Second, delivery instructed subjects attended to a different set of performance features compared to subjects receiving no prior information. Finally, delivery and comparison group subjects attended to some common performance features. The content analyses of breakpoints support these conclusions.

Dimension Ratings

There were two hypotheses concerning ratings made by subjects on the delivery and relating dimensions. First, it was hypothesized that subjects receiving no prior information concerning dimension to be rated would exhibit ratings of a lower quality compared to subjects receiving prior information. Second, it was hypothesized that subjects instructed to observe performance to rate one dimension but then asked to rate a different dimension would exhibit ratings of a lower quality compared to subjects receiving correct information concerning the dimension to be rated. As was explained in the first chapter, since true scores on the dimensions are not known, rating variances were used to estimate inter-rater reliability (one characteristic of rating quality).

The Effect of Prior Information

To test the hypothesis that receiving no prior information concerning dimension to be rated would adversely affect rating quality, several analyses were performed. First, a one-between, one-within
analysis of variance was calculated on comparison group subjects' ratings. Table 6 contains the results of this analysis. Like subjects in the experimental group, comparison group subjects rated the delivery and relating conditions significantly differently. Delivery received a mean rating of 1.65 while relating received a mean rating of 3.76. Table 6 also shows that there was a significant Dimension Rated x Order interaction. When delivery was rated first, it received a slightly higher rating (1.7 versus 1.6). Conversely, when relating was rated first, it received a lower rating (3.5 versus 4.0). A post-hoc test using the LSD procedure indicated that the two relating means were significantly different from each other (LSD = 2.22, p < .05) but the two delivery means were not.

This order effect can be tentatively explained. If comparison group subjects were attending to some of the same aspects of performance as delivery group subjects (as suggested earlier) and if below average performance was exhibited on the delivery dimension, perhaps subjects rating relating on the first round let the below average nature of delivery style impinge on their ratings of relating. This result is consistent with the literature on halo error (e.g., Cooper, 1981) which describes raters as allowing general impressions of the ratee to affect their ratings. However, the finding that relating received a higher rating when it was rated on the second round is inconsistent with the literature since it would be expected that delivery style would affect ratings of relating regardless of rating order. When asked to make their first round of ratings, subjects did not know on what dimension they would be rating performance on the
Table 6

Analysis of Variance: Dimension Ratings Made by Comparison Group Subjects (N=54)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>1</td>
<td>.59</td>
<td>.4</td>
<td>.53</td>
</tr>
<tr>
<td>Error</td>
<td>52</td>
<td>1.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension Rated</td>
<td>1</td>
<td>120.33</td>
<td>175.44</td>
<td>.0001</td>
</tr>
<tr>
<td>Dimension Rated x Order</td>
<td>1</td>
<td>3.0</td>
<td>4.37</td>
<td>.04</td>
</tr>
<tr>
<td>Error</td>
<td>52</td>
<td>.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
second round. Subjects who rated relating on the second round could perhaps better cognitively differentiate between delivery and relating since they had read a definition of and rated delivery prior to rating relating. Although there was an order effect, in general, comparison group subjects rated delivery and relating significantly differently and in the predicted manner. It was concluded that they accurately perceived the two dimensions exhibited in the videotape.

To further investigate rating quality, rating variances were examined for comparison, delivery instructed, and relating instructed subjects. Table 7 contains the rating means and variance for the two groups of subjects. F-tests indicated that the comparison and experimental groups' rating variances on the delivery dimension were not significantly different (F = .60) from each other nor were the two groups' rating variances on the relating dimension (F = .97). If rating variance provides an estimate of rating quality, then the two groups do not differ in this respect.

Since these results were unexpected, the mean ratings on the two dimensions produced by both groups were examined. T-tests were computed and it was found that ratings made by the two groups on the delivery dimension did not differ significantly (t = 1.14) nor did ratings on the relating dimension (t = .21).

Effect of Receiving Incorrect Information Concerning Dimension To Be Rated

The next series of analyses were conducted to test the hypothesis that subjects who were instructed to observe performance on one dimension but asked to rate it on a different dimension will exhibit lower
Table 7

Dimension Rating Means and Variances for Experimental and Comparison Group Subjects

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Experimental(^a)</th>
<th>Comparison(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery</td>
<td>M 1.84 Var .88</td>
<td>M 1.65 Var .53</td>
</tr>
<tr>
<td>Relating</td>
<td>M 4.44 Var 1.66</td>
<td>M 3.76 Var 1.61</td>
</tr>
</tbody>
</table>

\(^a\) N=50
\(^b\) N=54
quality ratings compared to subjects instructed to observe and rate the same dimension. Table 8 contains the means and variances for ratings of delivery and relating made by subjects who received incorrect information concerning dimension to be rated. Again, rating variances for this group were compared to experimental subjects' rating variances.

The difference between the two variances on the delivery dimension was non-significant ($F = .95$). On the relating dimension, the difference in variances approached significance ($F = 1.57$, critical $F$ value $= 1.6$). If rating variance provides an index of quality, then the two groups do not differ in this respect.

$T$-tests were performed to further examine the ratings. The first $t$-test indicated that the mean rating on delivery made by subjects who observed to rate relating was significantly different from the experimental group's mean rating of delivery ($t = 1.7, p < .05$). Subjects who observed to rate relating rated delivery lower than subjects who observed to rate delivery. A second $t$-test indicated that subjects observing to rate delivery exhibited significantly lower ratings on relating ($t = 3.35, p < .01$) compared to subjects instructed to rate relating. Although true scores on the two dimensions are not known, ratings produced by subjects given incorrect prior information are significantly different from ratings produced by subjects given correct prior information. These mean differences imply that there are quality differences since the mean produced by one group must lie closer to true scores than the mean produced by the other group. It is most probable that ratings made by subjects given incorrect information concerning dimension to be rated are of a lower quality since having
Table 8

Dimension Rating Means and Variances for Subjects Who Observed and Rated Different Dimensionsa

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery</td>
<td>1.48</td>
<td>.84</td>
</tr>
<tr>
<td>Relating</td>
<td>3.10</td>
<td>2.60</td>
</tr>
</tbody>
</table>

a N=52
subjects observe to assess one dimension and then rate a different one would not be expected to improve ratings.

**Are Rating Variances a Good Estimate of Rating Quality?**

The failure to obtain significant differences among the rating variances for the different groups was an unexpected result. The sample sizes were fifty (experimental group), fifty-four (comparison group), and fifty-two (incorrect information group). The lack of significant differences is probably not due to the power of the test because even if the sample size had been infinite, the F tests (except for the group which observed to rate delivery but actually rated relating) would not have reached significance. In order for the F tests to have been significant the variances in the comparison and incorrect information groups would have had to be 1.41 for delivery and 2.7 for relating. For the comparison group, the observed variance were .53 for delivery and 1.61 for relating. For the incorrect information group, the variances for the respective dimensions were .84 and 2.6.

It is possible that variance is not a sensitive measure of rating quality. To gain insight into this possibility, a group of forty-six subjects were asked to make ratings under conditions where large variances would be expected. More specifically, subjects in this group were asked to rate "any teacher they wanted to" on the delivery and relating dimensions. The variance on the delivery dimension was 2.20 and on the relating dimension was 2.6. An F-test indicated that the delivery rating variance for this group of subjects was significantly greater ($F = 2.6, p < .01$) than the variance of delivery ratings made
by experimental group subjects. However, the variances for the two groups of subjects for the relating dimension were not significantly different ($F = 1.57$). Thus, even in a situation where large rating variance should occur, only for the delivery dimension was the observed variance greater than that of the experimental group. At the very least, is is questionable whether variance is a sensitive measure of quality when ratings are made on the relating dimension.

Summary

The results of the analyses support two major conclusions. First, no qualitative differences were detected between ratings made by subjects given no prior information concerning ratings to be made and those made by subjects receiving correct prior information. Second, ratings made by subjects receiving incorrect prior information did differ in terms of quality compared to subjects receiving correct prior information. Additional analyses showed that rating variance may not be a sensitive measure of rating quality.
Chapter IV

Discussion

The current study examined the effect of dimension content on attentional processes of raters and on accuracy of ratings. The discussion is divided into four parts. First, an interpretation of the results will be presented. Second, the limitations of the study will be stated. Third, future research needs will be addressed. Finally, the study will be summarized and implications will be suggested.

Interpretation of Results

Unitization Results

It is proposed that when given a dimension on which to rate performance, a rater will activate a dimension relevant schema. The activated schema determines to which features of performance the rater will attend. Performance observation is conceived of as a feature monitoring process and button presses serve to mark the points at which raters perceive changes in one or more of the monitored features. Thus, raters observing to rate dimensions of different content should exhibit different unitization patterns.

In the current study, subjects who viewed the videotape in order to rate the instructor's delivery exhibited qualitatively different unitization patterns compared to subjects who viewed the videotape in order to rate relating skills. The videotape was constructed so that
behavior on the delivery dimension changed at a faster rate than did behavior on the relating dimension and subjects used significantly more button presses in the delivery condition. The hit rate measure indicated that only a small proportion of unit boundaries produced by subjects under the two conditions coincided. In fact, less agreement was observed in this study compared to Cohen and Ebbesen's (1979) study where 25% agreement was found using 1-second intervals and 50% agreement using 5-second intervals. Thus, subjects were attending to different features of performance rather than segmenting the same performance into smaller units.

The breakpoint data provide further evidence for attentional differences. When breakpoints were defined for the delivery and relating conditions, the number of such points common to the two conditions did not exceed chance. Moreover, the analysis of breakpoint content indicated that subjects were attending to qualitatively different aspects of performance. When instructed to assess relating, subjects attended to transitions in the lecture from theory to relating the theory to the students' experience and to the instructor's use of examples in explaining concepts. Subjects instructed to rate delivery attended to behavior changes in the instructor's presentation style. The results concerning the unitization patterns exhibited by experimental group subjects support the notion that raters observing performance to assess different dimensions activate appropriate and qualitatively different schemata.

When given no prior information concerning dimension on which performance is to be rated, it would be expected that raters will evoke
schemata different from raters given prior information. Subjects instructed to assess relating skills exhibited unitization patterns qualitatively different from those exhibited by subjects not receiving pre-observational information. First, the number of common breakpoints did not exceed chance and no relationship was demonstrated between breakpoints defined in the two groups. Second, further examination of both groups' breakpoints revealed few content similarities. Although subjects in the comparison group showed a tendency to attend to the organization of the lecture, this is qualitatively different from attending to the way in which the instructor relates the lecture to the students' experiences.

In contrast to subjects in the relating group, subjects instructed to assess delivery exhibited unitization patterns similar to subjects in the comparison group. First, delivery instructed subjects showed significant agreement, as to the location of breakpoints, with comparison group subjects. Second, there were similarities in breakpoint content for the two groups. Finally, the contingency coefficient indicated that there was a relationship between delivery and comparison group breakpoints.

When Newton's (1976) original research is considered, these results are not too surprising. Like subjects in the no prior information condition, Newton's subjects were instructed simply to press the button when they perceived one meaningful behavior to have ended and another to have begun. Upon examination of breakpoints, Newton found that subjects were monitoring overt features of performance (i.e., breakpoints occurred when the actor exhibited physical changes).
The results concerning the similarity in unitization patterns can be tentatively explained. Subjects instructed to assess delivery were attending to physical changes in the instructor's lecture (e.g., eye contact, use of notes, posture). If subjects receiving no prior information also attended to physical changes in the instructor's performance, they should exhibit unitization patterns similar to subjects assessing delivery. This is precisely the result that was obtained.

The results concerning comparison group unitization patterns support the notion that raters receiving no prior information activate qualitatively different schemata compared to raters assessing relating. However, it appears that the schemata activated by these raters are similar to those activated by delivery instructed subjects.

**Dimension Ratings**

It was proposed that subjects receiving correct prior observation concerning dimension to be rated would activate appropriate schemata and thus attend to dimension relevant aspects of performance. As a result of schematic processing, these subjects were expected to produce high quality ratings compared to subjects receiving no prior information and subjects receiving incorrect prior information. Since true scores were not known, rating variance was used to estimate quality and it was expected that subjects receiving correct prior information would agree with each other more in the ratings they made than would subjects not receiving such information. Small rating variances are viewed as evidence of convergent validity.

For the three groups, no significant differences among rating variances were observed. The power dictated by the sample sizes does
not seem to be an issue because even if data were collected from an infinite number of subjects, the observed differences between variances would not been that significant (except for the group which observed to rate delivery, but were then asked to rate relating). Even for the group who made ratings in a context where large variances would be expected ("rate any teacher"), only variance for the delivery dimension was significantly different from experimental group subjects' rating variance. Further, the group which received incorrect prior information produced significantly lower ratings than experimental group subjects. However, these mean differences were not associated with variance differences. Thus, it is possible that variance is not a sensitive measure of rating quality, at least not on the relating dimension.

When the results of the study are considered together, the finding that ratings made by subjects receiving no prior information did not differ from ratings made by subjects receiving correct prior information can be tentatively explained. As was previously demonstrated, raters who do not receive prior information attend to many of the same aspects of performance as do subjects observing to rate delivery. Since these two groups of raters activate similar schemata for observing performance, ratings of similar quality would be expected.

There are two possible explanations for the lack of significant findings concerning ratings on the relating dimension. First, analysis of the verbal content of breakpoints for the group receiving no prior information indicated that these subjects may have been attending to the structure and organization of the lecture. Although this type of
observational schema is dissimilar from a schema which causes raters to attend to the way in which the instructor relates the lecture to the students' experiences (relating schema), perhaps it is not dissimilar enough to affect ratings. As was previously stated, a second possible explanation of the result is that rating variance is simply not sensitive enough as a measure of quality on the relating dimension.

The results concerning ratings made by subjects instructed to observe one dimension, but then asked to rate another were interesting. No variance differences emerged, although the variance of ratings on the relating dimension approached being significantly different from the experimental group's rating variance on the same dimension. Further, subjects receiving incorrect prior information produced significantly lower ratings compared to subjects receiving correct prior information. While the difference between mean ratings made by the two groups on the relating dimension was substantial (4.44 versus 3.1), the difference between means on the delivery dimension was not (1.84 versus 1.48).

If subjects receiving incorrect prior information activated schemata similar to experimental group subjects, a possible explanation for these results exists. A schema relevant to the relating dimension would cause subjects to attend to what is being said by the instructor. However, while attending to what is being said, subjects also probably notice how the lecture is being presented (i.e., delivery) and thus can produce ratings of delivery not substantially different from ratings made by experimental group subjects. In contrast, a schema relevant to delivery is independent of the content of the lecture. If subjects
instructed to assess delivery attend very little to what is being said, then their ratings of relating should be more influenced by quality of delivery than subjects instructed to assess relating. This is the type of result that was obtained.

It is concluded that ratings made by subjects receiving incorrect information (especially ratings made on the relating dimension) differ qualitatively from ratings made by experimental group subjects. These ratings are most probably of a lower quality since giving subjects incorrect information concerning dimension to be rated would not be expected to enhance quality.

Limitations

The limitations of the study concern the issue of external validity and thus generalizability of results. The task performed by subjects in the study is different, in many respects, from the task performed by raters in organizations. In organizations, raters have many opportunities to observe an employee's performance relevant to a large number of dimensions. Further, in organizations, there is typically a larger interval of time during which performance relevant information can decay. Finally, the unitization measure introduces a degree of artificiality in the rating situation that is not there when raters in organizations observe behavior. All of these differences serve to limit the generalizability of results. However, as was stated in the first chapter, it is felt that at this stage of investigation, the control offered by the laboratory outweighs the shortcomings in terms of external validity. It should be noted that if measurements like the ones made in this study were made in an organization stronger
results would be expected because real-world raters probably observe performance many times while activating schemata incongruent with the dimensions they will be rating. Also, since memory decays over time raters may store only information which is most relevant to the activated schemata and/or use default values to fill in missing information.

Future Research Needs

There are several directions for future research in this area. First, as in other areas of performance appraisal, efforts are needed in developing a sensitive measure of rating quality. Second, research must be conducted to determine the nature of the schemata raters typically evoke in situations when they are observing performance without a particular dimension in mind. It is possible that in such situations raters are evoking schemata inconsistent (e.g., race or sex schemata) with the dimension to be rated. As a result, ratings may be lower as they were in this study when raters received incorrect prior information. Related to this, research is needed to determine how great a difference between observational and rating schemata is necessary to affect ratings.

Third, since raters in organizations typically rate employees on several dimensions, an important direction for future research is to determine the maximum number of dimensions raters can accurately attend to and rate. If current organization practices are such that raters are being asked to assess too many dimensions, then appropriate actions can be taken. Fourth, future research should be conducted to ascertain the similarity of various dimension relevant schemata. If such
similarities can be determined, then raters can be asked to make
ratings on additional dimensions that are relevant to the schemata
evoked during performance observation. Or, perhaps, raters should be
asked to make ratings only on dimensions which evoke similar schemata.
That is, different raters would be responsible for rating a different
set of dimensions each of which is relevant to a particular schema.

A fifth research direction concerns the possible effect of appraisal
purpose (e.g., to provide feedback, to make a promotion decision)
on schematic processing. Zedeck and Cascio (1982) demonstrated that
purpose of appraisal affected the way subjects weighted, combined, and
integrated identical performance dimensions. Those results suggest
that raters observing performance for different purposes may activate
purpose relevant schemata and thus attend to different features of
performance. Future research should address this possibility.

A sixth direction for future research is to examine the decay of
schema relevant information over time. It is possible that schema
information that has decayed is substituted with "default" performance
information which may cause ratings of a lower quality to be made.
Finally, research similar to the current study should be conducted in
which raters observe live performance. It may be that videotaped
performance lacks the richness of actual performance and this may have
an effect on schematic processing.

Implications and Conclusions

The results of this study have important implications for organi-
zational practices. In the study, it was found that raters have the
capability to attend to dimension relevant aspects of performance. It
was also found that ratings seem to be very robust. Ratings did not suffer when subjects did not receive prior information concerning the dimension they would be asked to assess. Raters receiving incorrect information gave lower performance ratings than those receiving correct prior information. However, only on the relating dimension did the mean rating differ substantially from the mean rating produced by experimental group subjects.

These findings suggest that, in general, organizations do not have to be concerned about the type and timing of information that they give raters regarding the dimensions they will be rating. It may be that different observational schemata cause raters to monitor certain features of performance which are not totally independent of performance features relevant to other dimensions. This is consistent with the suggestion made earlier that subjects observing to rate relating also encode information relevant to delivery because by attending to what is being said, the way in which it is delivered is encoded.

The critical variables here then are both the schema that is activated and the multidimensionality of performance. A given unit of behavior probably contains information relevant to a number of dimensions. Regardless of the schema activated, if a rater monitors features of performance that contain information relevant to the dimensions to be assessed, then enough dimension relevant information may be encoded to make quality ratings. Perhaps, the only situation in which schematic processing may lead low quality ratings is when raters activate schemata which cause them to attend to features of performance which contain no information or very little information concerning the
dimension to be rated. This is consistent with the suggestion made earlier that subjects observing to rate delivery attend to features of performance which do not carry much information about relating and thus produce ratings of a low quality on the relating dimension.

In an organization, a rater probably observes a given employee's performance over a period of time with a variety of different schemata activated. Even if the rater does not observe with a particular dimension in mind, given the multidimensionality of performance and the diversity of activated schemata, information relevant to a number of dimensions is probably encoded. Thus, the rater may be able to produce quality ratings on a variety of dimensions.

An important implication of the current study concerns the unitization measure. Since the measure was shown to behavior in a systematic manner, new possibilities emerge for using it in various appraisal contexts. For example, it may be used in a diagnostic sense to determine whether raters are attending to relevant performance features. A second possible use of the measure is in assessing the success of rater training programs.

The results of this study demonstrate that the observational processes described by Cohen and Ebbesen (1979) generalize to contexts in which the nature of the observational goal and the type of behavior being observed are qualitatively different. Like Cohen and Ebbesen's subjects, subjects in the current study attended to different behavior features as a result of pre-observational instructions. Thus, content of dimension to be assessed may be viewed as a member of the broader class of observational goals.
The results of the study also provide stronger evidence for the validity of the unitization measure. First, in previous research using the measure (Newtson, 1973; Cohen and Ebbesen, 1979) no predictions were made regarding the features of behavior subjects would attend to or the meaning of different unitization patterns. In the current study, predictions were made and were supported by the results.

Second, Newtson (1973) proposed that sets or expectancies can affect which feature of performance are monitored. This proposition was supported here since dimension to be assessed affected what subjects attended to.

The current study demonstrates that raters observing performance to rate different dimensions attend to different aspects of performance. Ratings made by subjects given no prior information did not differ qualitatively from ratings made by subjects given prior information. However, when raters were given incorrect prior information, ratings of a lower quality resulted. These results have many implications for both future research and for organizational practices.
Appendix A

Instructions, Dimension Definitions, and
Behavioral Examples Used in Pilot Testing
INSTRUCTIONS

There are two parts to this experiment. For both parts, you will need to refer to the sheet entitled "Instructor Behaviors".

Complete Part I before you start Part II.

If you have any questions, raise your hand and the experimenter will come answer your questions.

Turn to the next page and read the instructions for Part I.
PART I—INSTRUCTIONS

The sheet entitled "Instructor Behaviors" lists forty-four behaviors a psychology instructor might exhibit. These behaviors differ in terms of the degree to which they characterize an effective versus an ineffective instructor.

Your are to read each behavior and rate its degree of effectiveness as a teaching behavior.

Record your ratings on the answer sheet attached to this page.

You are to use the following scale.

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Very Ineffective</td>
<td>Moderately Ineffective</td>
<td>Neither Effective nor Ineffective</td>
<td>Moderately Effective</td>
</tr>
</tbody>
</table>

For example, if you think a particular behavior is "Very Effective", you should put a "5" in the blank on the answer sheet which is next to the number of the particular behavior. If you think a particular behavior is "Very Ineffective", you should put a "1" in the blank. If you think a behavior is somewhere in between, you should put a "2", "3", or "4".

Please begin Part I using the attached answer sheet and the sheet entitled "Instructor Behaviors".
INSTRUCTOR BEHAVIORS

1. Instructor shuffles noisily through his notes while presenting the lecture.

2. When a student is asking the instructor a question, the instructor interrupts the student in the middle of the question.

3. Instructor writes illegibly on the blackboard.

4. When a student asks how a particular theory applies to an aspect of student life at the university, the instructor will devote class time talking about the application.

5. When students respond to the instructor's questions in class, the instructor always points out the flaws in the answer, no matter how minimal.

6. Instructor introduces current lecture by integrating it with previous lecture material.

7. The instructor talks about the historical perspective of a theory so that students can relate its current status to its origin.

8. Instructor speaks in a loud commanding voice.

9. The instructor uses concrete examples with which the students are familiar when explaining theories.

10. If instructor perceives that students are not understanding a particular topic, he tells them that he doesn't have time to discuss the topic further and must move on to another topic.

11. When arriving early, instructor ignores the presence of students already in class and those who are arriving.

12. The instructor humiliates or embarrasses students who disagree with him.

13. When discussing a particular topic, instructor relates the lecture back to issues or concepts previously discussed in the course.

14. Instructor assigns the class to read on one topic and then lectures on another topic.

15. Instructor finishes lecture twenty minutes before the bell rings and says "I didn't get to finish my lecture today, so I'll answer any questions you have until the bell rings.

16. Instructor gives a brief overview of what will be covered during the lecture before starting the lecture.
17. Instructor hands out outlines of material to be covered in a particular lecture, prior to beginning the lecture.

18. Instructor writes a brief outline on the board prior to discussing the lecture and indicates to students this is the schedule of topics for that particular day.

19. Instructor uses notes as a reference to make sure all the important points were covered in adequate detail.

20. When the class is discussing an issue, the instructor treats every student's point as a valid one.

21. When the class does not understand a certain concept, the instructor is able to sense it and correct the situation.

22. Instructor speaks slowly, with appropriate pauses and repeats phrases so that students can accurately complete notes.

23. Instructor presents historical material chronologically so that students have a clear idea of when events occurred in relation to each other.

24. Instructor uses vocabulary appropriate to the level of the students.

25. Instructor shuffles madly through his notes in an attempt to find the material to be presented in the current lecture.

26. Instructor attempts to relate each topic to everyday experiences that the students may have had.

27. Instructor searches for appropriate words, saying "uh" and "you know" a lot.

28. At the end of the lecture, the instructor summarizes the important concepts that were covered in the lecture.

29. Instructor speaks in a monotone voice when presenting the lecture.

30. The instructor often tells students that their questions are ignorant or unintelligent.

31. Instructor reads from his notes making no eye contact with students.

32. Instructor speaks very rapidly making it difficult for students to write notes.

33. In addition to explaining a particular theory, the instructor discusses its implications for everyday behavior.
34. Instructor constantly looks at his watch during the lecture, demonstrating that he would rather be somewhere else.

35. Instructor tells class that the reason he is behind in his lectures is because class members have asked so many repetitive questions.

36. Instructor attempts to stimulate interest by throwing questions out for the class to respond to.

37. Instructor draws diagrams or flow charts on the board so that students can better understand the topic being discussed.

38. Instructor nervously clasps and unclasps his hands throughout the lecture.

39. Instructor hands back exams at the beginning of class, telling class he is sure that everyone is anxious to get the results.

40. When walking into class the instructor does not smile or greet the class before beginning the lecture.

41. Instructor changes back and forth among topics without making a tangible point about any of the topics.

42. Instructor tells the students the topics and subtopics that will be covered before actually starting the lecture.

43. Instructor speaks so softly that many students cannot hear him.

44. Instructor walks to desk and immediately starts speaking about topic with no introduction of issues.
PART II-INSTRUCTIONS

For Part II, you will again use the sheet entitled "Instructor Behaviors". This time, you are to assign each behavior to one of four categories.

First, read the attached sheet entitled "Category Definitions". This sheet describes four categories of psychology teaching behavior. Read these descriptions carefully. Note that to the left of each category title is a category code.

After you have read the four category definitions, go back and read each instructor behavior. After reading each instructor behavior, decide which behavior category it belong to.

Use the attached answer sheet for recording your decisions. In the blank next to each number, write the category code for the category that you believe the instructor behavior belongs in.

Assign each behavior to only one category. Try to assign each behavior of the forty-four behaviors to a category. If you absolutely cannot decide which category a particular instructor behavior belongs in, put a "?" in the blank next to the behavior number.

Please begin part II using the attached answer sheet and the sheet entitled "Instructor Behaviors".
DIMENSION DEFINITIONS

CODE

0 ORGANIZATION OF LECTURES

This category refers to the degree to which the instructor provides structure and organization to his/her lectures. Some ways in which the instructor might accomplish this goal is by telling the students the topic and subtopics to be covered before actually starting the lecture, distributing outlines of material to be covered in the lecture, relating the lecture back to previously covered topics, and exhibiting preparedness. Anything the instructor does that affects the organization of the lecture belongs under this category. This behavior category is more related to the way the instructor actually structures the lecture rather than the content of the lecture.

D DELIVERY SKILLS

This category refers to the degree to which the instructor is a skillful speaker in front of the class and is what most people think of as public speaking ability. Examples of some behaviors contained in this category are loudness of voice when lecturing, eye contact, and use of lecture notes. This category refers only to the manner in which the lecture material is presented, not to the content or organization of the material.

I INTERPERSONAL RELATIONS WITH STUDENTS

This category refers to the degree to which the instructor establishes and maintains rapport with the class and demonstrates sensitivity to students' questions and in interactions with the students. For example, an instructor who is effective in maintaining interpersonal relations with the students would answer students' questions in a friendly, non-demeaning manner whereas an instructor who is ineffective might humiliate or embarrass students in his/her interactions with them.

R RELATING LECTURE TO STUDENTS' EXPERIENCES

This category refers to the degree to which the instructor relates the subject matter of the lectures to the students' own experiences and knowledge to make the subject matter more important and meaningful to the students. This might be accomplished, for example, by using everyday experiences to illustrate theories or concepts or by applying psychological principles to explain real-world situations.
Appendix B

Instructions and Dimension Definitions for
Pilot Ratings of Videotape
INTRODUCTION

In a minute you will see the videotape of an instructor giving a psychology lecture. Your task is to watch the instructor in order to rate the degree to which she RELATES THE LECTURE TO THE STUDENT'S OWN EXPERIENCE.

Read the following definition of relating skills very carefully and keep the definition in mind as you watch the videotape.

RELATING SKILLS refer to the degree to which the instructor relates the subject matter of the lecture to the students' own experiences and knowledge to make the subject matter more important and meaningful to the students. This might be accomplished, for example, by using everyday experiences to illustrate theories or concepts or by applying psychological principles to explain real-world situations.

The experimenter will turn on the videotape in a minute. Watch the instructor's behavior so that you will be able to rate her RELATING SKILLS after the tape is complete. As you watch, try to pay attention to those behaviors that you think are relevant to making your rating of relating skills.

Remember, you are observing the instructor to make a rating only of her relating skills. So, do not take into account any behaviors that are not relevant to relating skills, when you observe her performance and make your rating. Thus, even if the instructor is very good or very bad in other aspects of her performance, this should not influence your rating of relating skills. Your rating should reflect effectiveness of only those teaching behaviors relevant to the definition of relating skills.

Read the definition of relating skills one more time before the tape is started.
INSTRUCTIONS

In a minute you will see the videotape of an instructor giving a psychology lecture. Your task is to watch the instructor in order to rate her DELIVERY SKILLS.

Read the following definition of delivery skills very carefully and keep the definition in mind as you watch the videotape.

DELIVERY SKILLS refer to the degree to which the instructor is a skillful speaker in front of the class. This category of teaching behavior is what most people think of as public speaking skills. Examples of some behaviors relevant to delivery skills are loudness of voice when lecturing, eye contact, and use of lecture notes. Delivery skills refer only to the manner in which the lecture is presented, not to the content and organization of the material.

The experimenter will turn on the videotape in a minute. Watch the instructor's behavior so that you will be able to rate her DELIVERY SKILLS after the tape is complete. As you watch, try to pay attention to those behaviors that you think are relevant to making your rating of delivery skills.

Remember, you are observing the instructor to make a rating only of her delivery skills. So do not take into account any behaviors that are not relevant to delivery skills, when you observe her performance and make your rating. Thus, even if the instructor is very good or very bad in other aspects of her performance, this should not influence your rating of delivery skills.

Read the definition of delivery skills one more time before the tape is started.
Rating: RELATING SKILLS

NOTE: Remember that your ratings should reflect only performance relevant to RELATING SKILLS and SHOULD NOT be influenced by any behavior you observe that are irrelevant to RELATING SKILLS.

<table>
<thead>
<tr>
<th>Extremely Ineffective</th>
<th>Extremely Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

1 = Extremely Ineffective
2 = Very Ineffective
3 = Ineffective
4 = Neither Effective nor Ineffective
5 = Effective
6 = Very Effective
7 = Extremely Effective

List the specific behaviors that you observed that support your rating. Be as specific as possible.
Rating: DELIVERY SKILLS

NOTE: Remember that your ratings should reflect only performance relevant to DELIVERY SKILLS and SHOULD NOT be influenced by any behavior you observe that are irrelevant to DELIVERY SKILLS.

Extremely Ineffective

1 2 3 4 5 6 7

1 = Extremely Ineffective
2 = Very Ineffective
3 = Ineffective
4 = Neither Effective nor Ineffective
5 = Effective
6 = Very Effective
7 = Extremely Effective

List the specific behaviors that you observed that support your rating. Be as specific as possible.
Appendix C

Experimenter Script—Experimental Group
As indicated on the sign-up sheets, you will be rating teacher behavior. In a minute, you will be watching a videotape of a teacher giving a lecture to rate two different categories of teaching behavior. Please read the instructions I am handing out very carefully.

(Hand out instructions 1A)

We're also interested in how you get out the information that you're going to use to make your rating. What I'd like you to do is break-up the tape for me into the units of chunks of information that you find most helpful for rating the teacher's delivery skills (relating skills). In this way, you can show us where you get your information from in the behavioral sequence.

You will do this by pressing the button in front of you. Now, take the button in front you in your hand. To familiarize yourself with that task, I will turn on the videotape machine and you will watch a short practice tape. As you watch the tape, press the button down and then let it up every time you perceive one meaningful behavior to have ended and another to have begun. Please hold the button under the table so that it is hidden from view.

(Turn on practice tape)

(When practice is over, hand out instruction 1B)

Now you will observe the actual experimental tape. Remember, use the button by pressing it down and letting it up, to break up the tape into the units or chunks of information that you find most helpful for rating the teacher's delivery skills (relating skills). This will help us determine where you get your information from for making your rating.

(Turn on videotape)

(When tape is complete, hand out rating sheets)

Now, we will be doing something a little different. This time you will be observing the same performance for a different purpose.

(Hand out instructions 2)

Again, we're interested in how you get out the information that you're going to use to make the rating. Once again, I'd like you to break-up the tape into the units or chunks of information that you find helpful for rating the teacher's delivery skills (relating skills). In this way, you can show us where you get your information from in the behavioral sequence. Remember, hold your button under the table so that it is hidden from view.

(When tape is complete, hand out rating sheets)

(Hand out debriefing forms)
Appendix D

Instructions and Dimension Definitions for Experimental Group
INSTRUCTIONS—DELLA

In a minute you will see a videotape of an instructor giving a psychology lecture. Your task is to watch the instructor in order to rate her DELIVERY SKILLS.

Read the following definition of delivery skills very carefully and keep this definition in mind as you watch the videotape.

DELIVERY SKILLS refer to the degree to which the instructor is a skillful speaker in front of the class. This category of teaching behavior is what most people think of as public speaking skills. Examples of some behaviors relevant to delivery skills are loudness of voice when lecturing, eye contact, and use of lecture notes. Delivery skills refer only to the manner in which lecture is presented, not to the content or organization of the material.
INSTRUCTIONS-DELB

The experimenter will turn on the videotape in a minute. Watch the instructor's behavior so that you will be able to rate her DELIVERY SKILLS after the tape is complete. As you watch, try to pay attention to those behaviors that you think are relevant to making your rating of delivery skills.

Remember, you are observing the instructor to make a rating only of her delivery skills. So do not take into account any behaviors that are not relevant to delivery skills, when you observe her performance and make your rating. Thus, even if the instructor is very good or very bad in other aspects of her performance, this should not influence your rating of delivery skills.

Read the definition of delivery skills one more time before the tape is started.

DELIVERY SKILLS refer to the degree to which the instructor is a skillful speaker in front of the class. This category of teaching behavior is what most people think of as public speaking skills. Examples of some behaviors relevant to delivery skills are loudness of voice when lecturing, eye contact, and use of lecture notes. Delivery skills refer only to the manner in which lecture is presented, not to the content or organization of the material.
INSTRUCTIONS-REL2

Again, you will see the videotape of the instructor giving a psychology lecture. This time your task is to watch the instructor in order to rate the degree to which she RELATES THE LECTURE TO THE STUDENTS' OWN EXPERIENCES.

Read the following definition of relating skills very carefully and keep the definition in mind as you watch the videotape.

RELATING SKILLS refer to the degree to which the instructor relates the subject matter of the lecture to the students' own experiences and knowledge to make the subject matter more important and meaningful to the students. This might be accomplished, for example, by using everyday experiences to illustrate theories or concepts or by applying psychological principles to explain real-world situations.

The instructor will turn on the videotape in a minute. Watch the instructor's behavior so that you will be able to rate her RELATING SKILLS after the tape is complete. As you watch, try to pay attention to those behaviors that you think are relevant to making your rating of relating skills.

Remember you are observing the instructor to make a rating only of her relating skills. So, do not take into account any behaviors that are not relevant to relating skills, when you observer her performance and make your rating. Thus, even if the instructor is very good or very bad in other aspects of her performance, this should not influence your rating of relating skills. Your rating should reflect effectiveness of only those teaching behaviors relevant to the definition of relating skills.

Read the definition of relating skills one more time before the tape is started. (See above for definition.)
INSTRUCTIONS-RELIA

In a minute you will see the videotape of an instructor giving a psychology lecture. Your task is to watch the instructor in order to rate the degree to which she RELATES THE LECTURE TO THE STUDENT'S OWN EXPERIENCE.

Read the following definition of relating skills very carefully and keep the definition in mind as you watch the videotape.

RELATING SKILLS refer to the degree to which the instructor relates the subject matter of the lectures to the students' own experiences and knowledge to make the subject matter more important and meaningful to the students. This might be accomplished, for example, by using everyday experiences to illustrate theories or concepts or by applying psychological principles to explain real-world situations.
The experimenter will turn on the videotape in a minute. Watch the instructor's behavior so that you will be able to rate her RELATING SKILLS after the tape is complete. As you watch, try to pay attention to those behaviors that you think are relevant to making your rating of relating skills.

Remember, you are observing the instructor to make a rating only of her relating skills. So, do not take into account any behaviors that are not relevant to relating skills, when you observe her performance and make your rating. Thus, even if the instructor is very good or very bad in other aspects of her performance, this should not influence your rating of relating skills. Your rating should reflect effectiveness of only those teaching behaviors relevant to the definition of relating skills.

Reread the definition of relating skills one more time before the tape is started.

RELATING SKILLS refer to the degree to which the instructor relates the subject matter of the lecture to the students' own experiences and knowledge to make the subject matter more important and meaningful to the students. This might be accomplished, for example, by using everyday experiences to illustrate theories or concepts or by applying psychological principles to explain real-world situations.
INSTRUCTIONS–DEL2

Again, you will see a videotape of an instructor giving a psychology lecture. Your task is to watch the instructor in order to rate her DELIVERY SKILLS.

Read the following definition of delivery skills very carefully and keep this definition in mind as you watch the videotape.

DELIVERY SKILLS refer to the degree to which the instructor is a skillful speaker in front of the class. This category of teaching behavior is what most people think of as public speaking skills. Examples of some behaviors relevant to delivery skills are loudness of voice when lecturing, eye contact, and use of lecture notes. Delivery skills refer only to the manner in which lecture is presented, not to the content or organization of the material.

The experimenter will turn on the videotape in a minute. Watch the instructor’s behavior so that you can rate her DELIVERY SKILLS after the tape is complete. As you watch, try to pay attention to those behaviors that you think are relevant to making your rating of delivery skills.

Remember, you are observing the instructor to make a rating only of her delivery skills. So, do not take into account any behaviors that are not relevant to delivery skills, when you observe her performance and make your rating. Thus, even if the instructor is very good or very bad in other aspects of her performance, this should not influence your rating of delivery skills.

Read the definition of delivery skills one more time before the tape is started. (See previous page for definition.)
Appendix E

Rating Forms for Experimental Group
Rating: RELATING SKILLS

NOTE: Remember that your ratings should reflect only performance relevant to RELATING SKILLS and SHOULD NOT be influenced by any behavior you observe that are irrelevant to RELATING SKILLS.

Extremely Ineffective 1 2 3 4 5 6 7
Extremely Effective

1 = Extremely Ineffective
2 = Very Ineffective
3 = Ineffective
4 = Neither Effective nor Ineffective
5 = Effective
6 = Very Effective
7 = Extremely Effective

List the specific behaviors that you observed that support your rating. Be as specific as possible.
Rating: DELIVERY SKILLS

NOTE: Remember that your ratings should reflect only performance relevant to DELIVERY SKILLS and SHOULD NOT be influenced by any behavior you observe that are irrelevant to DELIVERY SKILLS.

Extremely Ineffective

1 2 3 4 5 6 7

1 = Extremely Ineffective
2 = Very Ineffective
3 = Ineffective
4 = Neither Effective nor Ineffective
5 = Effective
6 = Very Effective
7 = Extremely Effective

List the specific behaviors that you observed that support your rating. Be as specific as possible.
Appendix F

Debriefing Sheets for Experimental and Comparison Groups
Experimental Group

This experiment is concerned with how people cognitively process job performance. We are interested in whether you paid attention to different aspects of the teacher's performance in the two different conditions (Delivery versus Relevance). We are also interested in whether watching the teacher's performance for the two different purposes affected your ratings of the teacher's behavior. Research like this will help us understand how individuals cognitively process job performance and the effect that cognitive processes may have on performance ratings.

Thank you very much for your time and cooperation.
Comparison Group

This experiment is concerned with how people cognitively process job performance. We are interested in what aspects of the teacher's performance you paid attention to as you were watching the videotape. We are also interested in whether what you paid attention to affected your ratings of the teacher's performance. Your responses will be compared to a group of subjects who watched the same tape, but who received information about behavior categories to be rated prior to watching the videotape. Research like this will help us understand how individuals cognitively process job performance and the effect that cognitive processes may have on performance ratings.

Thank you very much for your time and cooperation.
Appendix G

Experimenter Script - Comparison Group
Experimenter Script - Comparison Group

In a minute, I am going to show you a videotape of a teacher giving a psychology lecture. We are interested in how you observe the teacher's performance. What I'd like you to do is to break-up the tape into the units or chunks of information that you find most meaningful to you. In this way, you can show us where you get information that you consider meaningful in the teaching behavior sequence.

You will do this by pressing the button in front of you. To familiarize yourself with the task, I will turn on the videotape machine and you will watch a short practice tape. As you watch the tape, press the button and then let it up every time you perceive one meaningful behavior to have ended and another to have begun. Please hold the button under the table so that it is hidden from view.

(Turn on tape)

When practice is over . . .

Now you will observe the actual experimental tape. Remember, press the button and let it up to break the tape into the units or chunks of information that you find most meaningful. This will help us determine where you get your information from in the tape.

(Turn on the tape)

(When tape is over, hand out rating forms)

(Hand out debriefing forms)
Appendix H

Dimension Definitions and Rating Forms for
Comparison Group
Now that you have viewed the videotape, we would like you to rate the teacher's performance.

Please make your ratings on the following two pages. Turn the page and make the first rating. When you have done so, turn to the last page and make the second rating.

Please complete your first rating before turning to the last page.
Rate the teacher's DELIVERY SKILLS.

DELIVERY SKILLS refer to the degree to which the teacher is a skillful speaker in front of the class. This category of teaching behavior is what most people think of as public speaking skills. Examples of some behaviors relevant to delivery skills are loudness of voice when lecturing, eye contact, and use of lecture notes.

Use the following scale.

Extremely Ineffective | Extremely Effective
1 — 7

1 = Extremely Ineffective
2 = Very Ineffective
3 = Ineffective
4 = Neither Effective nor Ineffective
5 = Effective
6 = Very Effective
7 = Extremely Effective

Now list the specific behaviors that you observed that support your rating. Be as specific as possible.
Rate the teacher's RELATING SKILLS.

RELATING SKILLS refer to the degree to which the teacher relates the subject matter of the lecture to the students' own experiences and knowledge to make the subject matter more important and meaningful to the student. This might be accomplished, for example, by using everyday experiences to illustrate theories or concepts or by applying psychological principles to explain real-world situations.

Use the following scale.

<table>
<thead>
<tr>
<th>Extremely Ineffective</th>
<th>Extremely Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

1 = Extremely Ineffective
2 = Very Ineffective
3 = Ineffective
4 = Neither Effective nor Ineffective
5 = Effective
6 = Very Effective
7 = Extremely Effective

Now list the specific behaviors that you observed that support your rating. Be as specific as possible.
Appendix I

Categorization Percentages, Means and Standard Deviations for Behavioral Examples
### Categorization Percentages, Means, and Standard Deviations for Behavioral Examples

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.93</td>
<td>.03</td>
<td>.07</td>
<td>-</td>
<td>-</td>
<td>1.22</td>
<td>.55</td>
</tr>
<tr>
<td>2</td>
<td>.03</td>
<td>.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.47</td>
<td>.76</td>
</tr>
<tr>
<td>3</td>
<td>.63</td>
<td>.07</td>
<td>.09</td>
<td>.21</td>
<td>-</td>
<td>1.56</td>
<td>.95</td>
</tr>
<tr>
<td>4</td>
<td>.09</td>
<td>.38</td>
<td>.53</td>
<td>-</td>
<td>-</td>
<td>3.63</td>
<td>.16</td>
</tr>
<tr>
<td>5</td>
<td>.03</td>
<td>.94</td>
<td>-</td>
<td>.03</td>
<td>-</td>
<td>2.22</td>
<td>.29</td>
</tr>
<tr>
<td>6</td>
<td>.09</td>
<td>-</td>
<td>.28</td>
<td>.59</td>
<td>.03</td>
<td>4.16</td>
<td>.95</td>
</tr>
<tr>
<td>7</td>
<td>.13</td>
<td>.03</td>
<td>.50</td>
<td>.34</td>
<td>-</td>
<td>3.97</td>
<td>1.09</td>
</tr>
<tr>
<td>8</td>
<td>.94</td>
<td>-</td>
<td>.03</td>
<td>.03</td>
<td>-</td>
<td>3.66</td>
<td>1.04</td>
</tr>
<tr>
<td>9</td>
<td>.03</td>
<td>.07</td>
<td>.78</td>
<td>.13</td>
<td>-</td>
<td>4.75</td>
<td>.80</td>
</tr>
<tr>
<td>10</td>
<td>.03</td>
<td>.84</td>
<td>.07</td>
<td>.07</td>
<td>-</td>
<td>1.22</td>
<td>.49</td>
</tr>
<tr>
<td>11</td>
<td>.07</td>
<td>.89</td>
<td>.03</td>
<td>.03</td>
<td>-</td>
<td>2.06</td>
<td>.88</td>
</tr>
<tr>
<td>12</td>
<td>.03</td>
<td>.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.5</td>
<td>1.11</td>
</tr>
<tr>
<td>13</td>
<td>.09</td>
<td>.03</td>
<td>.25</td>
<td>.53</td>
<td>.07</td>
<td>4.09</td>
<td>.96</td>
</tr>
<tr>
<td>14</td>
<td>.07</td>
<td>-</td>
<td>.07</td>
<td>.81</td>
<td>.07</td>
<td>1.31</td>
<td>.54</td>
</tr>
<tr>
<td>15</td>
<td>.13</td>
<td>.28</td>
<td>-</td>
<td>.56</td>
<td>.03</td>
<td>2.00</td>
<td>1.08</td>
</tr>
<tr>
<td>16</td>
<td>.13</td>
<td>.28</td>
<td>-</td>
<td>.88</td>
<td>-</td>
<td>3.84</td>
<td>.99</td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>-</td>
<td>.07</td>
<td>.94</td>
<td>-</td>
<td>4.25</td>
<td>1.02</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>.07</td>
<td>.81</td>
<td>-</td>
<td>4.31</td>
<td>.82</td>
</tr>
<tr>
<td>19</td>
<td>.28</td>
<td>-</td>
<td>-</td>
<td>.73</td>
<td>-</td>
<td>4.66</td>
<td>.55</td>
</tr>
<tr>
<td>20</td>
<td>.09</td>
<td>.87</td>
<td>.03</td>
<td>-</td>
<td>-</td>
<td>4.13</td>
<td>.93</td>
</tr>
<tr>
<td>21</td>
<td>.16</td>
<td>.81</td>
<td>-</td>
<td>.03</td>
<td>-</td>
<td>4.84</td>
<td>.37</td>
</tr>
<tr>
<td>22</td>
<td>.87</td>
<td>.09</td>
<td>-</td>
<td>.03</td>
<td>-</td>
<td>4.75</td>
<td>.76</td>
</tr>
<tr>
<td>23</td>
<td>.09</td>
<td>.03</td>
<td>.34</td>
<td>.56</td>
<td>-</td>
<td>4.59</td>
<td>.61</td>
</tr>
<tr>
<td>24</td>
<td>.53</td>
<td>.21</td>
<td>.19</td>
<td>.07</td>
<td>-</td>
<td>4.41</td>
<td>.87</td>
</tr>
<tr>
<td>25</td>
<td>.25</td>
<td>-</td>
<td>.03</td>
<td>.71</td>
<td>-</td>
<td>1.72</td>
<td>.89</td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>.07</td>
<td>.91</td>
<td>.03</td>
<td>-</td>
<td>3.94</td>
<td>.08</td>
</tr>
<tr>
<td>27</td>
<td>.88</td>
<td>-</td>
<td>.03</td>
<td>.07</td>
<td>-</td>
<td>1.88</td>
<td>.94</td>
</tr>
<tr>
<td>28</td>
<td>.13</td>
<td>-</td>
<td>.07</td>
<td>.81</td>
<td>-</td>
<td>4.38</td>
<td>.87</td>
</tr>
<tr>
<td>29</td>
<td>.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.03</td>
<td>1.91</td>
<td>1.17</td>
</tr>
<tr>
<td>30</td>
<td>.03</td>
<td>.94</td>
<td>.03</td>
<td>-</td>
<td>-</td>
<td>1.03</td>
<td>.18</td>
</tr>
<tr>
<td>31</td>
<td>.88</td>
<td>.13</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.38</td>
<td>.71</td>
</tr>
<tr>
<td>32</td>
<td>.91</td>
<td>.03</td>
<td>.07</td>
<td>-</td>
<td>-</td>
<td>1.22</td>
<td>.55</td>
</tr>
<tr>
<td>33</td>
<td>.09</td>
<td>.07</td>
<td>.78</td>
<td>.07</td>
<td>-</td>
<td>4.13</td>
<td>.66</td>
</tr>
<tr>
<td>34</td>
<td>.66</td>
<td>.25</td>
<td>-</td>
<td>.03</td>
<td>.07</td>
<td>1.66</td>
<td>.75</td>
</tr>
<tr>
<td>35</td>
<td>.09</td>
<td>.75</td>
<td>.07</td>
<td>.07</td>
<td>.03</td>
<td>1.66</td>
<td>.75</td>
</tr>
</tbody>
</table>

**Abbreviations:**
- Del. = Delivery
- Int. Rel. = Interpersonal Relations
- Relev. = Relating
- Organ. = Organization
- Not Classif. = Not Classified
Categorization Percentages, Means, and Standard Deviations for Behavioral Examples\textsuperscript{A} \textsuperscript{B} (Continued)

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Del.</th>
<th>Int. Rel.</th>
<th>Relev.</th>
<th>Organ.</th>
<th>Not Classif.</th>
<th>Mean\textsuperscript{B}</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>.38</td>
<td>.44</td>
<td>.13</td>
<td>.07</td>
<td>-</td>
<td>4.09</td>
<td>.73</td>
</tr>
<tr>
<td>37</td>
<td>.34</td>
<td>.03</td>
<td>.03</td>
<td>.63</td>
<td>-</td>
<td>4.09</td>
<td>.64</td>
</tr>
<tr>
<td>38</td>
<td>.91</td>
<td>.07</td>
<td>-</td>
<td>.03</td>
<td>-</td>
<td>2.34</td>
<td>.90</td>
</tr>
<tr>
<td>39</td>
<td>.09</td>
<td>.69</td>
<td>.03</td>
<td>.19</td>
<td>-</td>
<td>3.53</td>
<td>1.08</td>
</tr>
<tr>
<td>40</td>
<td>.09</td>
<td>.88</td>
<td>.03</td>
<td>-</td>
<td>-</td>
<td>2.00</td>
<td>.95</td>
</tr>
<tr>
<td>41</td>
<td>.31</td>
<td>-</td>
<td>.03</td>
<td>.56</td>
<td>-</td>
<td>1.22</td>
<td>.61</td>
</tr>
<tr>
<td>42</td>
<td>.07</td>
<td>-</td>
<td>.07</td>
<td>.88</td>
<td>-</td>
<td>4.00</td>
<td>.45</td>
</tr>
<tr>
<td>43</td>
<td>.88</td>
<td>.03</td>
<td>-</td>
<td>.07</td>
<td>-</td>
<td>1.34</td>
<td>.70</td>
</tr>
<tr>
<td>44</td>
<td>.40</td>
<td>.13</td>
<td>.03</td>
<td>.44</td>
<td>-</td>
<td>1.90</td>
<td>.93</td>
</tr>
</tbody>
</table>


\textsuperscript{A} Percentages may not add up to 100 due to rounding

\textsuperscript{B} 1 = Extremely Ineffective, 5 = Extremely Ineffective
Appendix J

Verbal and Behavioral Content of Breakpoints and Non-Breakpoints
Breakpoints - Relating Group - Behavioral Content

BP 7  Looking straight ahead
BP 15 Looking straight ahead
BP 25 Looking at watch
BP 27 Eyes straight ahead, looks down at notes
BP 35 Looking down at notes
BP 40 Looking straight ahead
BP 42-43 Looking straight ahead
BP 45-46 Looking down at notes
BP 47 Looking straight ahead
BP 53-54 Looking down at notes
BP 56 Looking down at notes
BP 58 Looking down at notes
BP 76 Looking down, then up, taps finger
BP 82 Looking straight ahead
BP 85 Looking down, then up
BP 96 Looking down at notes
BP 98 Looking down at notes
BP 100-101 Looking straight ahead
BP 111 Looking up, then down
Breakpoints - Relating Group - Verbal Content

BP 7  "... Relate the theory to test taking behavior ..."
BP 15 "... test taking can help explain the theory."
BP 25 "... positive reward would be making an A on the test."
BP 27 "... making an E on the test."
BP 35 "talk about ... Adam's approach. Getting back to the theory ..."
BP 40 "... Input can be knowledge or skills."
BP 42-43 "Returning to the test-taking example, your inputs would be how ..."
BP 45-46 "... how smart you are, how hard you studied for the test."
BP 47 "Let's get back to the theory ..."
BP 53-54 "... what happens to you in the situation. If we get back to our test-taking example ..."
BP 56 "... outcome would be the grade you received."
BP 58 "... A would be a positive outcome and an E would be a negative outcome."
BP 76 "There are two types of situations of equity ..."
BP 82 "Let's return to the test-taking example."
BP 85 "Suppose you and your roommate are taking a Psychology course."
BP 96 "In this situation, your inputs and outcomes are high."
BP 98 "You both made a high grade."
BP 100-101 "Let's go back to the theory and look at the situation of equity and inequity."

BP 111 "You and your roommate both had equal inputs."
"Good morning. The lecture today is on equity theory. through out the discussion I will relate the theory to your own experiences and use the example . . . "

"Um . . . more specifically, I'll use the example . . . "

"An example that can really help explain the theory."

"Uh . . . where was I? Just a minute . . . "

"... rewards are equitable or inequitable . . . "

"... becomes what is the effect on behavior . . . "

"uh . . . uh . . . the effect on behavior of receiving . . . "

"Now these test-taking rewards can affect behavior . . . "

"... there are two terms with which you should be familiar."

"The second important term is outcome . . . um."

"... reward you receive from the situation is called an outcome."

"... test-taking behavior can influence how well you do and your type . . . "

"So the outcome . . . (loses place)"

"... the theory. The last thing I want to talk about . . . "

"He or she compares them with the outcomes of the other person."

"... situations that can result from the ration . . . "
"The first situation is equity. The second situation is inequity."

"... feelings of equity in the test-taking examples. Suppose you and your roommate ..."

"Suppose both of you have very high inputs and you work very hard."

"You and your roommate received very high grades ..."

"You both study very hard and both of you have very high inputs."

"You both work real hard studying for the test. You both studied ten hours."

"Your roommate got a good grade but you, you got a bad grade."

"The test-taking example would help you understand why ..."

"Finally, we examined situations ..."

"I want to sum up the lecture and talk about everything we talked about today so it will stay clear in your mind. We discussed the concepts of Adam's approach ..."

"Finally, we examined situations of ..."

"Tomorrow, we'll talk about choosing a comparison other. Thank-you."
Non-Breakpoints - Relating Group - Behavior Content

NBP 1-3  Looks down, hand on podium
NBP 5-6  Eyes straight ahead
NBP 14  Looking straight ahead
NBP 17  Eyes straight ahead, looks down at notes
NBP 19  Looking down at notes
NBP 29  Looking down at notes
NBP 30  Eyes straight ahead, looks down
NBP 32  Shuffles notes
NBP 37  Taps finger on podium, touches face
NBP 49  Rubs hands together, looks at watch
NBP 52  Looking down at notes
NBP 61-62  Looking down at notes
NBP 65  Shuffles notes
NBP 67-68  Looking down at notes
NBP 72  Looking down at notes
NBP 77-78  Looks down, then up; shuffles papers
NBP 81  Looks down, then up
NBP 83-84  Looking straight ahead
NBP 86  Holds notes perpendicular to podium, and taps them on podium
NBP 87-88  Looks down, then straight ahead
NBP 94  Looking straight ahead
NBP 103-104  Rubs hands together, looks at watch, touches face
NBP 105-106  Looking down at notes
NBP 113-114  Shuffles notes and straightens them
NBP 118-119  Looks down, then up
NBP 120     Looks up, then down
NBP 123-125  Eyes straight ahead, looks down at notes
NBP 126     Looks down at notes, looks at watch
NBP 128-129  Looks up, then down
Breakpoints - Delivery Group - Behavioral Content

BP 8  Eyes straight ahead, looks down at notes
BP 13 Eyes looking down at notes, looks up at class
BP 17-18 Eyes straight ahead, looks down at notes. Loses place -
        "Uh, where was I, just a minute . . ."
BP 25  Looks at watch
BP 30-31 "Uh, uh . . ." Eyes straight ahead, looks down, shuffles
        notes
BP 36-38 Eyes straight ahead, taps finger on podium, rubs face
BP 44  Eyes straight ahead, looks down at notes, "Uh, uh"
BP 49  Rubs hands together, looks at watch
BP 59  From upright position, leans on podium
BP 66  Leans down on podium
BP 69  Loses place, shuffles papers, "Where is that?"
BP 71  From upright position, leans on podium
BP 79  Looks at watch
BP 86  Holds lecture notes perpendicular to podium and taps them
        on podium
BP 91  Rubs hands together
BP 103-104 Rubs hands together, looks at watch, touches face
BP 108 "Uh, uh - taps finger on podium
BP 111 Eyes straight ahead, looks down at notes, taps finger on
        podium
BP 116 Eyes looking down at notes, looks up at class
BP 123-124 Eyes straight ahead, looks down at notes
BP 127  Looks at watch
Breakpoints - Delivery Group - Verbal Content

BP 8 "I'll relate the theory to the example . . . ."

BP 13 "And test-taking is . . . ."

BP 17-18 "Loses place "Uh, where was I, just a minute . . . ." "What is the effect?"

BP 25 "positive reward would be making an A on the test . . . ."

BP 30-31 "Uh . . . uh . . . effect on behavior of receiving positive and negative test-taking rewards."

BP 36-39 "Using Adam's approach, then are two terms with which you should be familiar - inputs and outcomes. An input is what you bring into the situation."

BP 44 "How smart you are . . . ."

BP 49 "The second important term is outcome . . . ."

BP 59 "Outcomes you could receive - an A on the test or an E on the test."

BP 66 "Where is that ?"

BP 69 " . . . fair or unfair"

BP 71 "compares his or her outcomes to the outcomes of others."

BP 79 Shuffling papers only

BP 86 "Suppose both you and your roommate took a test."

BP 91 "Uh - uh"

BP 103-104 "You both study very hard and both of you have very high inputs."

BP 108 "And you, you make a E on the test . . . ."

BP 111 "Both you and your roommate have equal inputs."
BP 116  "Uh . . . now"

BP 123-124  "I want to sum up the lecture and talk about everything we talked about today so it will stay clear in your mind."

BP 127  " . . . situation of equity and inequity."
Non-Breakpoints - Delivery Group - Verbal Content

NBP 12 "... taken many tests here at O.S.U. ...

NBP 16 "The major question of equity theory is ...

NBP 19 "... rewards are equitable or inequitable."

NBP 29 "... becomes what is the effect on behavior ...

NBP 34 "... what you bring into the situation ...

NBP 41 "... anything you bring into the situation ...

NBP 56 "An outcome would be the grade you received."

NBP 61-63 "... test-taking behavior can influence how well you do and your type of behavior."

NBP 68 "The last thing I want to talk about is ...

NBP 84 "... equity in the test-taking example. Suppose you and your roommate ...

NBP 95 "That would be a situation of equity."

NBP 101 "... and look at situations of equity and inequity."

NBP 106 "You studied ten hours, ...

NBP 118 "This test-taking example would ..."
<table>
<thead>
<tr>
<th>NBP</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>16</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>19</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>29</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>39</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>41</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>56</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>61-63</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>68</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>84</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>95</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>101</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>106</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>118</td>
<td>Looking down at notes</td>
</tr>
</tbody>
</table>
Breakpoints - Comparison Group - Behavior Content

BP 16  Looking straight ahead
BP 18  Loses place "Uh, where was I . . . just a minute"
BP 25  Looks at watch
BP 27  Eyes straight ahead, looks down at notes
BP 30  Eyes straight ahead, looks down
BP 34-35 Shuffles and straightens lecture notes, looks down
BP 37  Taps finger on podium, touches face
BP 47-49 Eyes straight ahead, looks at notes, rubs hands together,
        looks at watch
BP 59  From upright position, leans on podium
BP 65-66 From upright position, leans on podium
BP 70-71 Straightens notes, leans on podium
BP 75  Taps finger on podium
BP 100 Looking straight ahead
BP 103 Rubs hands together, looks at watch
BP 121 Looking straight ahead
BP 123 Eyes straight ahead, looks down at notes
BP 125 Looking down at notes
BP 127 Looks at watch
Breakpoints - Comparison Group - Verbal Content

BP 16 "The major question of equity theory is ..."
BP 18 "Uh ... what is the effect?"
BP 25 "Positive reward would be making an A on the test ..."
BP 27 "... making an E on the test ..."
BP 30 "Uh ... uh effect on behavior of receiving."
BP 34-35 "Ok, now I'd like to talk about Adam's Approach. . . ."
BP 37 "... there are two terms with which you should be familiar."
BP 47-49 "Let's get back to the theory. The second important term is outcome. Um ...
BP 59 "There are two outcomes you can receive."
BP 65-66 "So the outcome" Loses place "Where is that?"
BP 70-71 "... last thing I want to talk about today is how a person compares his or her outcomes to the outcomes of others."
BP 75 "... fairness of outcomes. When a person . . ."
BP 100 "... ratio of one person's outcomes is the same . . ."
BP 103 "You both study very hard . . ."
BP 121 "and you both studied hard for the test."
BP 123 "I want to sum up the lecture and talk about"
BP 125 "We discussed the concepts of Adam's approach"
BP 127 "... situation of equity and inequity . . ."
Throughout this discussion, I will relate to your own experiences and use the example . . .

. . . getting results of such tests. This is an example . . .

An example that can really help explain the theory.

. . . rewards are equitable or inequitable.

. . . test-taking example, the major question becomes what is the effect on behavior . . .

Returning to the test-taking example . . .

. . . rewards you receive for the situation is called an outcome.

. . . test-taking behavior can influence how well you do and your type of behavior.

. . . rate is the ration of one person's inputs or outcomes.

You both studied very hard, worked hard for the test . . .

You got to class, take the test . . .

You and your roommate received very high grades . . .

You both studied very hard and both made a high grade.

Um . . . you both studied very, very hard for the test.

You studied ten hours, say, you both take the test and your roommate made an A.
"This test-taking example would help you understand why you feel inequity."

"... comparison other. Thank-you."
Non-Breakpoints - Comparison Group - Behavioral Content

<table>
<thead>
<tr>
<th>NBP</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-3</td>
<td>Hand on podium, looking down at notes</td>
</tr>
<tr>
<td>10-11</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>14</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>19</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>28-29</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>40</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>42</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>55</td>
<td>Looking straight ahead</td>
</tr>
<tr>
<td>61-62</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>74</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>88-89</td>
<td>Looking down, then straight ahead</td>
</tr>
<tr>
<td>92</td>
<td>Taps finger on podium</td>
</tr>
<tr>
<td>97</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>102</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>106-107</td>
<td>Looking down at notes</td>
</tr>
<tr>
<td>114</td>
<td>Straightens notes</td>
</tr>
<tr>
<td>118-120</td>
<td>Looking down, then looks straight ahead</td>
</tr>
<tr>
<td>129</td>
<td>Looking down</td>
</tr>
</tbody>
</table>
LIST OF REFERENCES


