Application of the Means-End Theoretic for Understanding the Cognitive Bases of Performance Appraisal

JAMES P. JOLLY

College of Business, Idaho State University

THOMAS J. REYNOLDS

School of Management, University of Texas at Dallas

AND

JOHN W. SLOCUM, JR.

Edwin L. Cox School of Business, Southern Methodist University

Applicability of means-end theory to the area of performance appraisal was tested in a sample of 22 nurse supervisors. Individual interviews were conducted which employed triadic sorting and laddering procedures in order to identify cognitive concepts most salient to performance rating. These concepts were classified according to level of abstraction—attribute, consequence, or value. This classification served as the basis for construction of a summary cognitive map of the performance appraisal process for this sample. Cognitive differentiation analysis revealed that, on average, value level concepts accounted for significant variance in performance judgments over and above that which was accounted for by attributes and consequences. Conclusions include an argument for further investigation of personal values and their effects on the performance appraisal process.

INTRODUCTION

The measurement of employee performance lies at the heart of most aspects of human resource management within organizations. Performance ratings often serve as criteria against which the success of motivation and training programs are measured. They are also the primary source of feedback given to employees for developmental purposes. The purpose of this research is to propose and empirically test a means-end
cognitive map to study the cognitive processes raters use when evaluating subordinates.

Problems with Performance Ratings

Despite their ubiquitous nature, performance ratings are subject to both intentional and unintentional systematic biases and to random errors (Wherry & Bartlett, 1982). As a result, organizational researchers have devoted considerable effort to improving the validity of performance ratings. These efforts can be divided into two primary categories of research: (1) format related and (2) rater training related.

The first category, intended to improve rating validity, consists of studies designed to develop and improve rating scale formats (cf. Blanz & Ghiselli, 1972; Smith & Kendall, 1963). Although these formats, such as behaviorally anchored rating scales and mixed standard scales, are intuitively appealing, empirical tests have generally been disappointing (cf. reviews by Bernardin & Beatty, 1984; Jacobs, Kafry, & Zedeck, 1980; Landy & Farr, 1983; Schwab, Heneman, & DeCotiis, 1975). The failure of rating scale research to significantly improve the validity of evaluations can be attributed to at least two problems. First, rating scales comprise only one small part of the total appraisal process (DeNisi, Cafferty, & Meglino, 1984). Second, error variance can arise from many sources (e.g., rater observation processes, purpose of the rating) besides the format of the rating scale (James, 1973).

Numerous other studies (e.g., Bernardin & Beatty, 1984; Bernardin & Buckley, 1981) have attempted to improve rating validity by training raters to avoid bias. While training has been successful at reducing certain psychometric errors such as halo, leniency, and central tendency, it has not been demonstrated to significantly improve rating process.

The failure of format and rater training research to improve the validity of performance rating has led numerous researchers (DeNisi et al., 1984; Feldman, 1981; Landy & Farr, 1980) to call for a new research direction aimed at understanding the cognitive processes that a rater goes through in constructing a rating. Landy and Farr (1980), in particular, have called for research focused on uncovering raters' cognitive structures.

Several studies (see Hobson & Gibson, 1983, for a review) have been conducted that addressed the information-processing behavior of raters through the use of policy capturing. As DeNisi et al. (1984, p. 361) state: "any improvement in (rating) accuracy depends upon a clear understanding of the appraisal process" (italics added). Uncovering raters' cognitive processes and structures in the rating task would be of considerable heuristic value for researchers interested in understanding (and in ultimately improving) the rating process. It also promises important ap-
plied benefits, such as the possibility of using knowledge of raters’ strategies in training them to improve those strategies.

Unfortunately, several serious problems with the policy-capturing approach have been detailed (Hobson & Gibson, 1983) which constrain its usefulness and applicability. For example, most policy-capturing studies have used hypothetical ratee profiles, leading to supposition that they reflect a “paper people” effect (Gorman, Clover, & Doherty, 1978). Researchers have also artificially created orthogonal ratee performance dimensions—a situation that clearly does not mirror the real world. Most importantly, DeNisi et al.’s (1984) cognitively based model of the rating process highlights a serious weakness of the method. They argue that performance rating consists of three major stages for the rater. The first is observation of the ratee. The second stage consists of encoding, storage, and retrieval of the observed information. The third and final stage is information integration and rating. Policy capturing assesses—at best—only the third stage; it examines the way raters integrate information to provide a rating. Thus, it appears to suffer from the same problem discussed earlier with regard to format research; it is looking at only one piece of the total appraisal process.

An alternative approach directed specifically at the explication of the components of the raters’ cognitive structure needs to be found. Recently, there has been a strong emphasis on relating consumers’ motives to perceptual and preference processes through the use of means-end theories (Gutman, 1982; Howard, 1977). Gutman’s model has served as the basis for methodologies that have proven useful in uncovering cognitive structures with respect to choosing between product classes (Reynolds, Gutman, & Fiedler, 1984; Reynolds & Jamieson, 1984). The similarity of the general perceptual processes involved in an individual’s relationship to a given stimulus class (i.e., a product class or a group of employees) may offer a parsimonious vehicle for the study of performance rating.

Gutman’s Means-End Theory

Gutman’s model rests on the fundamental assumption that personal values exert a powerful influence on people’s behavior (Gutman, 1982; Rokeach, 1968; Yankelovich, 1981). Values are the ultimate source of choice criteria people use in selecting among alternative products, behaviors, or other members of a particular stimulus class (Howard, 1977; Rosenberg, 1956; Locke, 1976). The model assumes that “people cope with the tremendous diversity of (stimuli) that are potential satisfiers of their values by grouping them into sets or classes so as to reduce the complexity of choice” (Gutman, 1982, p. 60). This is analogous to cate-
gorization processes in performance appraisal discussed by Feldman (1981) and by DeNisi et al. (1984): “the rater does not store the raw information but interprets it, and then stores this interpreted representation” (p. 376). This representation is accomplished by the fundamental process of assigning the stimulus to a category (Rosch, 1978).

Some type of categorization process is necessary because it allows people to overcome storage capacity limitations of their cognitive structures (Smith, Adams, & Schorr, 1978). This categorization process can be a source of inaccuracy in making performance evaluations. Gutman’s model offers a possible explanation as to why raters assign certain stimuli to the categories they do. The central focus of the model is that individuals choose actions that produce desired consequences and minimize undesired consequences. Gutman suggests, following Rokeach (1968), that values provide consequences with positive or negative valences. Thus, values can be viewed as the key factor underlying a person’s categorization processes.

According to Feldman (1981), categories are defined as “fuzzy sets.” The possession of attributes is probabilistically related to category membership. Categorization, then, is based on the extent to which the features of a given person overlap with those of a category prototype (Tversky, 1977). The prototype is an abstract analog or image that summarizes resemblances among category members. As Cantor and Mitchel (1979, p. 30) note, “It is probably the degree of ‘family resemblance’, not the continuous surpassing of a few critical properties tests, that determines category membership.” Feldman argues that the ratee is assigned to a category because he/she possesses certain salient attributes (perhaps only one or a few) which distinguish the rater’s category prototype. Gutman’s model offers an explanation as to why certain attributes are salient for the rater. He asserts that these attributes are simply intermediate, and not causal factors.

Gutman’s model, as adapted to the performance rating situation, is exhibited in Fig. 1. Essentially, it can be thought of as a three-step chain:

Attributes ——— Consequences ——— Values.

Attributes are the most concrete components of the taxonomy. They are descriptive characteristics of a stimulus. For example, “high quality” and “imported” would be considered attributes of certain beers. Similarly, “motivated,” “knowledgeable,” and “dependable” are possible attributes of workers. At the next highest level of abstraction are consequences. These represent the personal meaning possessed by a certain stimulus. For example, an “imported” beer might possess the consequence “impresses others.” A “dependable” worker might lead to the consequence “minimizes staffing problems.” Finally, at the highest level
of abstraction are personal values. Values, then, represent the end states that are important to an individual, which are closely tied to one’s idea of self. Continuing the previous examples, an “imported” beer might “impress others” and in turn achieve the value “self-esteem” from feeling good about oneself. That is, the more personal interpretation of the lower level consequence makes the consumer feel better about his/her self-worth. Similarly, a “dependable” worker might “minimize staffing problems” and therefore lead to a supervisor being “recognized or respected for a job well done,” which may also have implications to one’s self-worth.

The Gutman model, as adapted to performance appraisal in Fig. 1, hypothesizes that an individual creates arrays (categories) of stimuli (ratees) with certain shared attributes. These will be instrumental in helping the rater achieve desired consequences. These, in turn, move the individual toward valued end states. Values are seen as the ultimate causal factor in an individual’s categorization processes. These result in a series of categories which make up an individual’s “schema” that integrates existing information into a more cohesive unit.

Gutman’s model can serve as a source for mapping the cognitive struc-
ture of an individual. One of the most general perspectives on modeling cognitive structures is to view them as an associative network (Wickelgren, 1981). In the case of performance ratings, the specific attributes, consequences, and values a person uses to think about ratees are considered as nodes in a network. The linkages between the three levels are the associations or arcs of the network.

**General Means-End Research Methodology**

This methodology is designed to identify both the content (nodes) and organization (arcs) of an individual's cognitive structure. Through this methodology, the researcher can determine the concepts (cognitive elements) an individual uses to think about a given class of stimuli, and identify the way in which these separate concepts are linked together in memory.

Triadic sorting (Kelly, 1955) is a method designed to identify the lower level distinctions of cognitive structure. Distinctions are the cognitive elements that serve as the basis for an individual's formation of categories. They are dichotomies that represent the end points of dimensions along which objects may be compared and used to differentiate between members of a stimulus class. Triadic sorting uses the following steps to elicit these distinctions. First, members of the stimulus class are divided into sets of three, called triads. Subjects are presented with one triad at a time and asked to "Tell me some overall way of thinking about these three (stimuli) in which two of the three could be considered alike and yet different from the third." The responses to these questions across triads yield the bipolar distinctions, or base level nodes.

The salient distinctions gained through triadic sorting are usually concrete in nature and are representative of a relatively low level of abstraction. Therefore, a technique is needed by which we can reach higher levels. Laddering (Gutman & Reynolds, 1979; Reynolds & Gutman, 1984) accomplishes this by forcing the respondent up the "ladder of abstraction" to see how individuals link salient attributes to higher level consequences and, ultimately, to personal values.

The laddering procedure consists of presenting the subject one at a time with the distinctions elicited from the triadic sort and then asking which pole or end of the distinction is preferred. The respondent is then asked why that pole is preferred. The more personal "why is that important to you" question is repeated using the prior response until the individual can no longer answer. On most occasions, respondents reach a level where they are talking about values similar to those discussed previously (Gutman, 1982).

The responses obtained via the laddering procedure are content analyzed and classified according to their level of abstraction (attribute, con-
sequence, or value). For example, in the present study, a respondent's first ladder begins with the attribute "self-motivated" and leads to the consequence "achieves high-quality patient care," which in turn leads to the values "feeling of personal accomplishment" and "feel good about myself (self-esteem)." The results of the content analysis are used to form an implication or dominance matrix which is the basis for representing a cognitive structure as a tree diagram. Value-level concepts are positioned at the top of this map because they are the abstract, higher level "end" concepts which are linked through a series of associations to the more concrete distinctions.

Laddering Assessment Research

This method of eliciting higher level distinctions has been used to study the bases of brand choice in the consumer behavior area. Reynolds and Jamieson (1984) investigated the relationship of the various levels of abstraction with perceptual and preference distances in a study of 20 students using 16 retail (department) stores as stimuli. They first obtained cognitive concepts across levels of abstraction by using the triadic sorting and laddering procedures. They related these to judgments of preference and perceptual distance.

The analysis they used to relate these variables was cognitive differentiation analysis (CDA), which is fully described by Reynolds and Sutrick (1986). Briefly, this procedure, using only ordinal assumptions about the data, separately assesses the discrimination power of each respondent's descriptors (attributes, consequences, values) with respect to perceptual and preference differences. CDA provides an index for each descriptor that ranges from $-1$ to $+1$. Similar to Kendall's Tau (1955), it reflects the degree to which each attribute, consequence, and value level distinction accounts for the discrimination judgments (perceptual distances or preference distances) between stimuli.

CDA has several advantages over the more commonly used interpretive methods applied to multidimensional scaling. First, CDA focuses on the direct interpretation of pairwise judgments, and not on derived relationships in Euclidean space. As a result of the ability to operate directly on the pairwise judgments, CDA can be used when small data sets (less than 12) are involved in a multidimensional scaling study. Second, CDA permits only ordinal assumptions to be made with respect to the independent ratings, which are then related to the pairwise judgments. Lastly, CDA has the ability to operate upon multiple vectors in which the contribution of each vector can be assessed statistically.

In the Reynolds and Jamieson (1984) study, aggregate CDA summary measures from these analyses showed stronger associations between attribute level distinctions and perceptual distances than for higher level
consequence and value) distinctions. But for preference discriminations
the results were reversed; preference, on an aggregate basis, was more
highly associated with the highest level of abstraction (values). Thus, on
average, people appear to use attribute level terms when judging differ-
ces between stimuli, while for preference judgments values appear to
be the key determinant.

A study by Reynolds, Gutman, and Fiedler (1984) replicated these
findings with a sample of 35 individuals. Convenience restaurants served
as stimuli. Using similar methodology and analysis techniques, results
indicated that, moving up the hierarchical content structure from at-
tributes to consequences to values, the association between preference
and the laddered elements became stronger. An opposite pattern was
found, as before, between perceived similarity and the laddered ele-
ments. It seems that people tend to view similarity in terms of concrete
attributes, while their preference judgments appear to be based on
values.

RESEARCH DIRECTION AND HYPOTHESES

It is clear that present approaches to performance appraisal research
(e.g., improving rating scale formats and designing traditional rater
training programs) have not succeeded in improving our understanding of
the performance rating process. We agree with DeNisi et al. (1984) and
Landy and Farr (1980) that before rating validity can be significantly im-
proved we must first understand the processes through which ratings are
made and the cognitive structures that impact these processes. The
policy-capturing approach, to date, has not proven adequate to under-
stand the performance appraisal process. Gutman’s means-end chain
model has been introduced as an alternative way to approach the problem
of determining a rater’s cognitive structures. The present research is the
first step in extending Gutman’s theoretical paradigm and its accompa-
nying methodology to the performance appraisal domain. It is through
understanding of attributes, consequences, and values that we hope to
improve the knowledge of the performance appraisal process.

The first goal of the present research is to test whether perceptual
judgments will be more strongly related to attribute level distinctions
while preference judgments will be more strongly related to higher level
distinctions. The hypotheses designed to test these predictions are the
following.

Hypothesis 1a. Mean CDA indices with perceptual distances will be
greater for attribute level distinctions than those for consequence level
distinctions, which, in turn, will be greater than those for values.

Hypothesis 1b. Mean CDA indices with preference judgments will be
greater for higher level distinctions (consequences and values) than those for attribute level distinctions, with values being the highest.

Acceptance of these hypotheses would mean that previous findings in consumer behavior can be replicated in the performance appraisal area. Furthermore, a regression extension of CDA (Reynolds, Weeks, & Perkins, in press) can be used to assess the relative contribution of the higher levels of abstraction, thereby adding more statistical power to the analytical methods.

**Hypothesis 2a.** Mean $R^2$ terms from CDA regressions with perceptual distance judgments as the dependent variable will be greater for attribute level distinctions than those for consequences, which, in turn, will be greater than those for values.

**Hypothesis 2b.** Mean $R^2$ terms from CDA regressions with preference judgments as the dependent variable will be greater for higher level (consequence and value) distinctions than those for attribute level distinctions, with values being the highest.

The most important question to be addressed by the present research concerns the contribution of value level distinctions to improving our overall understanding of the performance appraisal process. Previous research in the performance appraisal area has explored attribute level data and some consequence level data, but has left values untouched. The question is, does the addition of value level data significantly increase our understanding of performance appraisal judgments? If the answer to this question is affirmative, this would strongly support the applicability of means-end theory and its extension to value orientations to the performance rating area. In order to answer this question, the following hypothesis was tested.

**Hypothesis 3.** CDA regression with a forced order of entry in which attributes are entered first, consequences second, and values third, will result in significant increases in explained variance ($R^2$) at each stage for a significant number of subjects.

If this hypothesis is supported, then we would have evidence that values add to our understanding of the rating process. Bernardin and Beatty (1984) and DeNisi et al. (1984) have mentioned the importance of consequences of the rating process for raters, but no one has gone the step further to values. This may be an important step in understanding the performance appraisal process.

**METHOD**

**Sample**

Twenty-two female nursing managers, each of whom evaluated at least
nine nurses on a regular basis, served as the sample. Their average age was 33.6 years, they had held their current supervisory positions for an average of 2.5 years, and had supervised the nurses used as stimuli for an average of 1.6 years. Thirteen of the respondents held bachelors degrees, 4 were graduates of 3-year diploma programs, 3 held associate degrees, 1 held a masters degree, and 3 did not report their level of education.

Procedure

Data were collected from each manager in an individual interview lasting from 1½ to 2 hours. Each interview consisted of the following parts.

a. Preinterview screening. If a manager met the span of control requirements (at least nine subordinates) and ratee position requirements (had actually used the organization's performance appraisal system to distribute rewards to the ratee), the names of nine ratees were recorded for use in the interview. A minimum span of control of nine was imposed because of a concern for the number of pairs of judgments for stability with CDA, as well as for consistency across subjects. A further screening requirement was that each subject had to have been involved in the performance appraisal process of her current subordinates at some previous point in time. This was to assure that each nurse had actually conducted an appraisal and was familiar with each subordinate whom she was asked to evaluate.

b. Warm-up. The first part of the actual interview session was an informal discussion of ratings in general and of rating situations the subject typically experienced. The object was to get the subject thinking seriously about performance ratings and to get her involved in the interview process. Previous uses of this procedure indicated that this is a very involving experience for most subjects.

c. Similarity judgments. Subjects were presented index cards, each containing 1 of 44 pairs of ratee names (the 36 possible pairs of 9 ratees plus 8 repeated pairs for use in a check of internal reliability). They were asked to rate the similarity of the members of each pair on a 9-point scale ranging from 1 = very similar to 9 = very dissimilar. This represents the measure of perceptual distance.

d. Triadic sorting. Next, 10-12 triads of ratee names that previously had been randomly chosen from all possible three-person combinations

1 This sample was subjected to a procedure modified by changes suggested in a pilot study of 18 respondents. Pilot study data indicated that descriptors were not being defined clearly enough during the interview. For example, "flexible" can change in meaning from one usage to the next for the same individual, or can mean different things to different individuals. The refined procedure involved the explicit defining (through behavioral examples) of each descriptor so that later uses (ratings) of that descriptor would be consistent.
of the nine ratees were presented, one at a time, to the subject. Each triad was on a separate index card. The interviewer said, "Please tell me a way in which two of these employees are the same and yet different from the third." The process continued until the subject had gone through all triads, and the interviewer had recorded each of the distinctions.

e. Distinction ranking. Once the distinctions were elicited and recorded, the respondents were asked to rank them in terms of their importance in determining the performance rating they would give an individual. For each respondent, the distinction ranked first in importance was used to begin the following step to elicit higher level distinctions.

f. Laddering. For a given distinction, the respondent was asked which pole of the distinction was most preferred. Then the interviewer asked why that pole was preferred. As described earlier, this why question was repeated until the respondent could no longer answer. All responses were recorded for later content analysis to determine their level of abstraction. This entire procedure was repeated using the distinction ranked second in importance, or subsequently ranked distinctions as necessary, until two complete ladders (each containing an attribute, a consequence, and a value) were obtained.

g. Preference judgments. Subjects were again presented with the 44 ratee pairs previously discussed. First, they were asked which member of the pair was the better overall performer. If there were no ties, the respondent was asked to indicate how much the performance of the preferred member of the pair exceeded that of the other member. A 9-point scale ranging from 1 = just slightly better to 9 = no comparison—one totally better was used.

h. Distinction based ratings. The distinctions from the ladders obtained in step f were used as the basis for ratings of each respondent's nine ratees on a 7-point scale ranging from 1 = (negative pole of distinction) to 7 = (positive pole of distinction). Thus, there were six ratings (two attribute level, two consequence level, and two value level) for each ratee. These ratings served as independent variables in CDA analyses with preference and perceptual distance judgments as the dependent variables.

i. Importance ratings. Subjects were asked to weight the importance they thought they would attach to each distinction when rating a subordinate's performance. A 7-point scale ranging from 1 = not at all important to 7 = extremely important was used.

j. Overall performance rating. Next, subjects were asked to provide an evaluation of each individual's overall performance on a scale ranging from 1 = poor performer to 7 = outstanding performer. This measure was included as a method check of convergent validity.
k. Demographic information. A short form was given to respondents to collect information on job experience, length of contact with the ratees, and related data.

ANALYSIS

Content Analysis of Elicited Concepts

The first step in analyzing the large number of responses to the triadic sorting and laddering tasks was to conduct a thorough content analysis of all elicited concepts. Categories were developed to capture the thoughts expressed by subjects in response to the sorting and laddering tasks. Table 1 displays the 21 categories that were developed in this manner.

<table>
<thead>
<tr>
<th>Title</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td></td>
</tr>
<tr>
<td>1. Caring/does extras</td>
<td>Compassionate, patient oriented, makes patient comfortable, teaches</td>
</tr>
<tr>
<td>2. Organized</td>
<td>Sets priorities, efficient, accomplishes goals in timely manner</td>
</tr>
<tr>
<td>3. Knowledgeable/skilled</td>
<td>Proficient at skills, good patient assessor, experience</td>
</tr>
<tr>
<td>4. Peer leader</td>
<td>Independent, assertive, leads by example, not afraid to ask questions</td>
</tr>
<tr>
<td>5. Motivated/self starter</td>
<td>Conscientious, dependable, shows up/on time</td>
</tr>
<tr>
<td>6. Calm/steady</td>
<td>Not easily flustered</td>
</tr>
<tr>
<td>7. Communicates well</td>
<td>(With doctors, patients, families)</td>
</tr>
<tr>
<td>Consequences</td>
<td></td>
</tr>
<tr>
<td>1. Good patient care</td>
<td>Optimal care, more than what is necessary, high-quality care</td>
</tr>
<tr>
<td>2. Unit runs smoother</td>
<td>Floor runs efficiently</td>
</tr>
<tr>
<td>3. Less patient anxiety</td>
<td>Patient more secure, at ease, more confident, better attitude, knows nurse cares</td>
</tr>
<tr>
<td>4. My job easier</td>
<td>Less intervention required, more time for other duties</td>
</tr>
<tr>
<td>5. Less complaints</td>
<td>Less hassles, complaints from patients, doctors, families</td>
</tr>
<tr>
<td>6. Costs stay down</td>
<td>Stay within budget</td>
</tr>
<tr>
<td>7. Less stress</td>
<td>Less personal stress, less frustration, I’m calmer</td>
</tr>
<tr>
<td>8. Better patient wellness</td>
<td>(Mental, physical, emotional, spiritual)</td>
</tr>
<tr>
<td>9. Good reflection</td>
<td>(On me; on unit; on hospital)</td>
</tr>
<tr>
<td>10. Increased morale</td>
<td>Increased teamwork/spirit, increased unit job satisfaction</td>
</tr>
<tr>
<td>Values</td>
<td></td>
</tr>
<tr>
<td>1. Personal accomplishment/fulfillment</td>
<td>Self-fulfillment from helping those in need, feel good about caring, nurturing, being needed</td>
</tr>
<tr>
<td>2. Self esteem/worth</td>
<td>Self-confidence</td>
</tr>
<tr>
<td>3. Pride as manager</td>
<td>Pride in being good manager/leader, recognition for good management job</td>
</tr>
<tr>
<td>4. Pride in profession</td>
<td>Pride in field of nursing, in professionalism of unit</td>
</tr>
</tbody>
</table>
Seven categories were classified as attributes, 10 were classified as consequences, and 4 were classified as values.

**Structural Analysis**

The content analysis identified the categories used in raters' cognitive structures. These 21 categories represent the nodes in the model. Structural analysis was then used to identify the linkages between the categories. These linkages are the arcs of the network model. The first step of this structural analysis was to construct a square implication matrix in which the rows and columns were denoted by the categories developed in the content analysis. Each cell of the matrix actually contained two entries. The first represented the number of times the row category immediately preceded the column category. Thus, the first cell entry corresponded to the number of times a particular (row) category directly elicited the other (column) category during the laddering procedure. The second entry in a cell represented the number of times a row category preceded a column category either directly or indirectly. This procedure produced an asymmetrical implication matrix that was then treated as a dominance matrix.

**Hierarchical Value Map (HVM)**

The dominance matrix, created by selecting a minimum number of relations to warrant a connection, was then used to construct a directed graph that provided the basis for diagrammatically representing the cognitive structure, the hierarchical value map. (For a more detailed example of this procedure, see Olson & Reynolds, 1983). The summary HVM that represents the most common concepts and linkages (nodes and arcs) across subjects is presented in Fig. 2. The individual linkages of each of the ladders for each of the 22 subjects were compared to the linkages contained in this map. As indicated in Fig. 2, 77% of all possible linkages between pairs across attributes, consequences, and values are represented on this map.

Figure 2 illustrates that there are a large number of possible pathways or orientations represented. However, when each of the 44 ladders (2 for each of 22 subjects) is compared to the HVM and to the other subjects' ladders, it becomes clear that there are a few dominant means-end orientations represented in the data. Table 2 summarizes the four most popular orientations.

Ladders I and II in Table 2 are the most common. These reflect a concern on the part of the nurse managers for providing good patient care, which in turn leads to a feeling of personal accomplishment and, finally, an increased sense of self-esteem and worth. Several other nurse managers provided ladders that also lead to a sense of personal accomplish-
In helping other people. These ladders, however, began with different attribute level distinctions, such as "knowledgeable," "organized," or "communicates well." A second major orientation is shown in ladders III and IV in Table 2. These ladders focus on achieving a smooth-running, efficient unit. This in turn reduces personal stress for the manager and increases her sense of pride in her management skills. Ladders III and IV are prototypical of the cognitive structure of a manager who has a value to perceive herself as a good manager.

In sum, the interview procedure and the summary HVM do appear capable of reflecting the cognitive structures of managers with respect to the performance appraisal process. Not only are most of the elicited ladders (77%) accounted for by the HVM, but also several dominant cognitive groupings are captured.
### TABLE 2
FOUR COMMON MEANS-END ORIENTATIONS

<table>
<thead>
<tr>
<th>Ladder</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td>Self-esteem/worth:</td>
</tr>
<tr>
<td></td>
<td>personal accomplishment ⇒</td>
</tr>
<tr>
<td></td>
<td>better patient wellness ⇒</td>
</tr>
<tr>
<td></td>
<td>less patient anxiety ⇒</td>
</tr>
<tr>
<td></td>
<td>caring/does extras ⇒</td>
</tr>
<tr>
<td>II.</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>Self-esteem/worth:</td>
</tr>
<tr>
<td></td>
<td>personal accomplishment ⇒</td>
</tr>
<tr>
<td></td>
<td>better patient wellness ⇒</td>
</tr>
<tr>
<td></td>
<td>good patient care ⇒</td>
</tr>
<tr>
<td></td>
<td>motivated/self-starter ⇒</td>
</tr>
<tr>
<td>III.</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>Pride as manager:</td>
</tr>
<tr>
<td></td>
<td>less stress ⇒</td>
</tr>
<tr>
<td></td>
<td>unit runs smoother ⇒</td>
</tr>
<tr>
<td></td>
<td>high morale ⇒</td>
</tr>
<tr>
<td></td>
<td>peer leader ⇒</td>
</tr>
<tr>
<td>IV.</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>Pride as manager:</td>
</tr>
<tr>
<td></td>
<td>less stress ⇒</td>
</tr>
<tr>
<td></td>
<td>my job easier ⇒</td>
</tr>
<tr>
<td></td>
<td>unit runs smoother ⇒</td>
</tr>
<tr>
<td></td>
<td>organized/sets priorities ⇒</td>
</tr>
</tbody>
</table>

Cognitive Differentiation Analysis\(^2\) (CDA)

The analyses employed are within-person analyses. The relative within-person predictability of preference judgments by attributes, consequences, and values was investigated. Since the only direct statistical testing that can be employed is within-person, conclusions must be drawn from looking at general trends of these results across persons. CDA assesses the relationship of a descriptor (independent variable) vector to a symmetric matrix of pairwise distance judgments. In the present study, the six descriptors for each subject (two attributes, two consequences, two values) are related separately to two distance matrices—one containing perceptual distances and the other containing preference judgments.

Each nurse manager rated nine ratees in both of the paired comparison tasks. Thus, each distance matrix contained \(N(N-1)/2\), or 36 judgments.

\(^2\) For complete descriptions of univariate CDA and the CDA regression extension see Reynolds, Weeks, and Perkins (in press).
The six descriptor ratings for each subject formed separate vectors with nine elements each. CDA was used to assess the relationship of each of the 36 pairwise distance judgments to the respective ordinal relationships of each of the elements in each descriptor vector. This is accomplished by expanding both the 36 distance judgments and the descriptor ratings into pairs of pairs relations.

The expansion into a pairs of pairs model produces a transformation of the elements in each vector based on a simple ordinal comparison. For each comparison, if element A > element B, then a 1 is recorded for that pair of pairs relation; if A = B it receives a value of 0; and if element A < element B, it is scored as a −1. The resulting 1s, −1s, and 0s for the respective pairs of pairs relationships are then used in the computation of a CDA index that represents a measure of association. As noted by Reynolds and Sutrick (1986), this index, Tau-R, closely resembles Goodman–Kruskal gamma (Goodman & Kruskal, 1954). It is computed by the following formula:

\[ \text{C-D Tau-R} = \frac{C - D}{C + D} \]

where \( C \) is the number of concordant relations \((-1, -1)\) and \((1, 1)\), and \( D \) is the number of discordant relations \((-1, 1)\) and \((1, -1)\).

**CDA Regression Extension**

CDA can also be used to address situations in which multiple, equally weighted descriptor vectors are made into a composite vector and considered as a set. However, this multivariate generalization to the equal weighting case fails to adequately address the typical research question of finding some set of weights that optimally reproduces the dependent variable. A regression extension of CDA has been advanced to derive these weights (Reynolds et al., in press).

The regression extension of CDA performs a regression on the 1s, 0s, and −1s that are formed from the pairs of pairs model. The resulting regression equation produces beta weights for each descriptor vector and an \( R^2 \) value. However, the directly derived \( R^2 \) is not easily interpreted because it is based on all the pairs of pairs data. CDA regression produces a corrected \( R^2 \) that may be directly interpreted. The 1s, 0s, an −1s that have been expanded or "unfolded" from the original distance matrix are "refolded" according to a decision rule into 36 pairs corresponding to the original pairwise judgments. These "refolded" values are then related to the original judgments. This result is actually a squared Pearson product–moment correlation coefficient computed between the predicted distances and the true distances. This new measure, "\( R^2 \) re-
RESULTS

Internal Consistency

For each respondent, 8 of the 36 pairwise judgments were repeated in reverse order of stimulus presentation for both the perceptual distance and preference tasks. A Pearson correlation coefficient and a Kendall’s Tau were computed for each paired comparison task to assess the internal consistency of each manager’s ratings. Median Pearson correlations were .88 and .87 for the perceptual distance task and preference task, respectively. Median Taus were .73 and .76, respectively. For the perceptual distance task, 17 of the 22 Pearson correlations and 17 of the Taus were significant at less than the .05 level. For the preference task, 19 Pearson correlations and 16 Taus were significant at the .05 level. It appears, therefore, that the pairwise judgment tasks were understood by the nurse managers and captured stable cognitive concepts.

Independence of Pairwise Tasks

The two pairwise judgment tasks attempted to measure what should be relatively distinct cognitive processes. The perceptual distance task explored perceived similarities among stimuli; this was a much broader task than the measurement of performance specific dimensions—the goal of the preference task. To ascertain whether the two tasks were independent, a correlation was computed between the perceptual distance judgments and the preference judgments. The median correlation was $r = .23$. The relatively low median correlation indicates that the two tasks were independent for most subjects.

Convergent Validity Check

The overall performance ratings obtained for each employee were treated as descriptor variables in a univariate CDA analysis with preference as the dependent variable. A median Tau-$R$ of .79 was found. This indicates high convergence of the separate performance measures.

Subjective Weighting of Descriptors

Table 3 presents the results of a single-factor repeated measures analysis of variance on the importance ratings obtained from each manager for her six descriptor variables. These ratings represent the subjective
weights that these managers attach to the descriptors. Similar to results reported by Reynolds and Jamieson (1984) and by Reynolds et al. (1984), these results show that raters think they place much less importance on value level distinctions in the rating process than they actually do. These findings also parallel those in policy-capturing research. Policy-capturing researchers argue that people cannot accurately reproduce their own importance weights for performance dimensions.

**Univariate CDA**

Figure 3 shows the results of univariate CDA analyses with both perceptual distance and preference as dependent variables. Two points are evidenced from an examination of the mean Tau-Rs reported in Fig. 3. First, the descriptors are clearly much more predictive of preference than they are of perceptual distance. This is not surprising, since the descriptors are performance-related concepts arising from the roles of manager and subordinate.

Second, as predicted in Hypothesis 1b, the relationship with preference is progressively higher as we go from attributes to consequences to values. For both ladders, the mean Tau-Rs support this hypothesis. However, Hypothesis 1a, concerning the order of Tau-R values for perceptual distances, is not supported. In both ladders, value level distinctions have the highest mean Tau-R values.

**CDA Regression: Within ACV Levels**

Hypotheses 2a and 2b focus on the relative predictability across attribute (A), consequence (C), and value (V) levels. To test these hypotheses, separate regression analyses were run for each dependent variable with the two attributes as predictors. The same analysis was repeated twice—first with the two consequences as predictors and then with the two values as predictors. The data in Fig. 4 graphically show the results of these analyses.
For preference data, all three levels were quite predictive. $R^2$ values for preference ranged from .38 to .50. Further, consequences accounted for a higher percentage of variance than did attributes, and values accounted for even more variance than did consequences. Hypothesis 2b is therefore accepted; the predicted order of $R^2$ values was supported.

Hypothesis 2a, however, concerning perceptual distances was not supported. As predicted, attributes obtained a higher $R^2$ term than did consequences ($R^2 = .14$ for attributes and $R^2 = .12$ for consequences), although the difference is not statistically significant. More importantly, contrary to Hypothesis 2a, values ($R^2 = .24$) and not attributes ($R^2 = .14$) produced the highest $R^2$ term.

**CDA Regression: Incremental Contribution of Higher Levels**

The previous analysis provided evidence that, across subjects, values tend to be more predictive of preference than are attributes and consequences. One of the questions addressed by the present research is
whether values add predictability over and above what we are able to gain from traditional attribute and consequence concepts. Hypothesis 3 was proposed as an attempt to answer this question.3

The overall $R^2$ for attribute level descriptors was already known from the previous analysis. To test for incremental gain in predictability due to the addition of consequences, a second regression was performed that contained the two attributes and the two consequences. Overall $R^2$ terms were then compared to see if the addition of consequences significantly increased the amount of explained variance. Similarly, a third regression containing all three levels of descriptors was performed and then the in-

3 Note that this hypothesis is tested solely by the incremental change in $R^2$ from level to level. Possible interaction effects are not tested because of the nature of the means-end model itself. This conceptual model addresses only the relative contributions of the various levels of abstraction to the predictability of the dependent variable and does not concern itself with multiplicative effects.
TABLE 4

SUMMARY OF INCREMENTAL CONTRIBUTION OF CONSEQUENCES AND VALUES TO PREDICTABILITY OF PREFERENCE

<table>
<thead>
<tr>
<th>Regression step</th>
<th>Not significant</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Attributes only</td>
<td></td>
<td>.12/.3</td>
</tr>
<tr>
<td>2. Attributes + consequences</td>
<td>n.s.</td>
<td>sig.</td>
</tr>
<tr>
<td></td>
<td>.20/.2</td>
<td>.30/.1</td>
</tr>
<tr>
<td>3. Attributes + consequences + values</td>
<td>n.s.</td>
<td>sig.</td>
</tr>
<tr>
<td></td>
<td>.20/.1</td>
<td>.68/.1</td>
</tr>
</tbody>
</table>

Note. each cell contains mean $R^2/N$. Significance tested by (Nie et al., 1975) $F(M, N - k - 1) = \frac{(R^{2y,1} to K - R^{2y,12})/M}{(1 - R^{2y,1} to K)/(N - K - 1)}$ where $k =$ total number of independent variables and $M =$ number of variables in subset.

ccremental $R^2$ tested. The data in Table 4 display the results of these analyses. In 10 cases, consequences added a significant amount of variance. Values added a significant amount of the explained variance in 11 cases (representing 50% of the sample). Thus, for fully half of these nurse managers, values appear to be a source of previously unexplained variance in performance appraisals.

Analysis by Most Predictive Ladder

From each respondent's two ladders, the one with the highest $R^2$ term with respect to preference was chosen as that respondent's "best" ladder. The rationale for this was that in an exploratory study of this nature, the real question of interest is the potential for increased understanding afforded by a new technique. This potential can best be addressed by looking at the best of an individual's ladders. The following univariate CDA analysis was conducted on each individual's "best" ladder.

Figure 5 reports mean Tau-R values with preference and perceptual distance for respondents' best ladders. These results parallel those reported in Fig. 3. Tau-Rs with preference are in ascending order from attribute (.57) to consequence (.62) to value (.72). Thus, further support for Hypothesis 1b is provided.

With respect to perceptual distance judgments, the results are similar to those found in Fig. 3. Contrary to Hypothesis 1a, values exhibit a
stronger relationship to perceptual distance (mean $\text{Tau-R} = .39$) than do attributes (mean $\text{Tau-R} = .23$). Thus, two separate analyses have failed to support Hypothesis 1a.

**DISCUSSION**

This research has proposed a new theoretical and methodological approach to understanding performance appraisal decisions. Means-end theory provides a logical framework from which the performance appraisal process can be approached.

The triadic sorting and laddering procedures measured cognitive concepts important to performance rating. In fact, the data in Table 4 reveal that in only 1 of the 22 cases did triadic sorting fail to elicit at least one concept that accounted for significant variance in performance judgments. The laddering procedure was capable of eliciting higher order
cognitive constructs (values) when applied to the performance ratings of these nurse managers. These are especially encouraging in light of the fact that this is the first application of Gutman's model to performance appraisals.

The success with which the summary hierarchical value map (HVM) was able to represent the dominant performance-related orientations of these nurse managers was also encouraging. The explication of dominant hierarchies, such as those exhibited in Table 2, aids in the identification of managers who share certain perceptual orientations with respect to the rating task.

The acceptance of Hypothesis 1b (i.e., relationships with preference judgments were greatest for values, followed by consequences and attributes, respectively) coalesces the work in consumer behavior (e.g., Reynolds et al., 1984; Reynolds & Jamieson, 1984). The pattern of relationships found in prior consumer research was replicated in this study. Values appear to be more strongly related to preference than are consequences and attributes. Unfortunately, most performance appraisal research has focused on clarifying attributes and not on ascertaining values of the rater.4

Further support for the point of view that values are most predictive of preference is provided by the CDA regressions. Comparisons of CDA regressions within level of abstraction also indicated that values were the most predictive of preference judgments. Acceptance of Hypothesis 2b indicates that the $R^2$ terms for values are higher than those for lower level distinctions.

The results of the present study do not support the hypotheses (1a and 2a) concerning perceptual distance judgments. In the consumer behavior area, buyers apparently perceive products and other stimuli primarily at the attribute level. This does not appear to be the case with respect to rating the performance of employees. Results of the univariate CDA analyses and the regression analyses indicate that values, and not attributes, are the best predictors of perceptual distance judgments. This should not be surprising. There are, after all, considerable differences between products and employees when used as stimuli in means-end methodology. It seems reasonable to assume that managers are much more knowledgeable and involved with employees (as a stimulus class) than consumers are with consumer products. Because of this knowledge and

---

4 Locke (1976) and Locke, Schweiger, and Latham (1986) have noted a similar problem in the job satisfaction area. They comment that despite extensive theoretical and empirical support for the view that attaining one's values leads to job satisfaction, relatively little attention has been given to the study of values by job satisfaction researchers. One reason for the dearth of research is that employees' values are numerous and their means of fulfillment complex.
involvement, the hierarchical nature of memory might serve to elicit higher level distinctions even in the simpler and more general perceptual distance task. The important aspects of the stimulus nurses to the nurse managers are job related. Therefore, it is not surprising that higher level distinctions were reached even when they were asked to simply compare the general similarity of the stimulus nurses.

While the present research results support, in general, the efficacy of using means-end theory and CDA analysis to investigate cognitive structures in the performance-rating area, a number of issues remain to be addressed. These issues can be classified as methodological issues and performance-rating application issues. Among the methodological issues requiring further study are the following. First, the triadic sorting and laddering procedures are complex; they are, in fact, almost as much "art" as "science." Only through increased application to the performance rating problem will we be able to refine these data-gathering procedures.

Second, it is conceivable that the intrusive nature of these procedures created demand characteristics or response sets that artificially affected the obtained results. This sensitization may have put respondents in touch with their values and permitted us to capture the personal meanings in the respondent's own vernacular. While we cannot directly refute such an assertion at this time, a number of points seem relevant. Means-end theory suggests that if a person's involvement with a particular decision is high, and if the task is unstructured and requires considerable judgment, then values should be an important factor in that decision. This contention is supported by empirical research (Reynolds & Jamieson, 1984), which showed that the procedures employed here work more successfully with psychologically involving types of judgment decisions than with more calculative decisions. Making performance decisions is an unstructured task that requires a large degree of judgment on the part of the rater. Additionally, it is one that should be quite involving for the rater, because managers make decisions affecting their subordinates' careers and immediate rewards. Given the exigencies of the performance appraisal process, the results support Gutman's theory. Future research must try to better understand the dynamics of this issue.

Third, the relationship of hierarchical value structures to actual performance ratings, an issue not addressed by the present design, must be addressed. With the present design, we can only assume that the criterion measures (pairwise performance judgments) were isomorphic with actual organizational ratings. This is an empirical question that requires additional research.

With respect to future applications of means-end theory and these methods to performance rating, several additional questions remain to be
answered. First, how will samples drawn from other occupations respond to these techniques? Higher level managers, who primarily use intuition to make decisions, may be better able to articulate their distinctions in the interview setting and, therefore, provide even more accurate reproductions of their hierarchical value structures than lower level managers. These techniques may also add insights into differences in perceptions of performance across organizational levels. Borman (1974) found that persons at different levels within an organization have contrasting views of performance for particular jobs. Perhaps means-end theory and the laddering methodology can be used to uncover value level differences that could be causing these contrasting views. By comparing value hierarchies across organizational levels, such an examination may be possible.

Second, Williams, DeNisi, Blencoe, and Cafferty (1985) and Landy and Farr (1980), among others, have presented evidence that performance ratings vary according to the purpose of the rating. For example, raters give different ratings for administrative purposes than they do for research purposes. For administrative purposes, raters often give different evaluations depending on whether the rating is for developmental purposes (career development) or for evaluative (merit raises) purposes. In the field of consumer behavior, researchers have found evidence that indicates preference is multidimensional. One’s preference judgments of a group of convenience restaurants, for example, are modified by the time of day and other exigencies of the situation. Likewise, it seems reasonable to argue that preference judgments of ratees would be similarly modified by organizational situations. Laddering has proven to be effective in uncovering these situational differences for managers making evaluative decisions, but perhaps it can also be effective in explicating the different value structures elicited by managers for developmental purposes as well.

The third possible application of means-end theory and the laddering methodology in the performance appraisal area concerns the potential use of laddering to identify the cognitive structures that exist among raters in certain organizational units. Once these are identified, this information could serve as the basis for a training or management development program in which managers (raters) are made aware of their dominant rating orientations and how these affect their performance judgments. These actual orientations might also be compared to the orientations desired by the organization, and any differences could be dealt with appropriately.

In conclusion, the present study has served as an introductory look at the potential for application of means-end theory and laddering methodology to the area of performance appraisal. More research is needed to
refine the techniques for use in this area. But the potential for increased understanding of the cognitive structures that raters impose on the rating task seems great.

REFERENCES


RECEIVED: September 16, 1985