

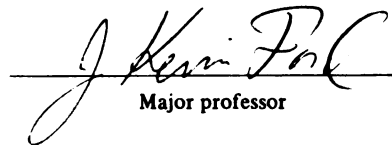
This is to certify that the
thesis entitled
The Effect of Information Quality and Type of Data on
Information Integration Strategies

presented by

Michael Paul Kirsch

has been accepted towards fulfillment
of the requirements for

Masters degree in Psychology


Major professor

Date November 13, 1985



RETURNING MATERIALS:
Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

1985

--	--

**The Effect of Information Quality and
Type of Data on Information Integration
Strategies**

by

Michael Paul Kirsch

A THESIS

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

MASTER OF ARTS

Department of Psychology

1985

18/180/

ABSTRACT

THE EFFECT OF INFORMATION QUALITY AND TYPE OF DATA ON INFORMATION INTEGRATION STRATEGIES

by

Michael Paul Kirsch

This research investigated the cognitive processes of raters when making performance evaluation decisions. Specific attention was focused on how raters integrated information which varied in terms of information quality, and was composed of either subjective or objective data. Analysis of the results showed that the manipulation of information quality was not successful. However, correlational analyses revealed that subjects were using the dimension of information quality in making their ratings, but this use was not related to the experimental conditions. The results also indicated that subjects had a strong bias towards using subjective data over objective data in making their ratings. An additional focus of the research was on examining subjects' knowledge and awareness of their rating policies. Analysis of the relationship between participants' subjective and statistical weighting schemes and their written protocols showed that subjects had fairly good notions of the policies they used in making the ratings.

Implications, limitations of the study, and future directions in the study of information integration and performance appraisal are discussed.

ACKNOWLEDGEMENTS

First, I would like to thank my committee members, Mary D Zalesny, Steve Kozlowski, and J.Kevin Ford for their guidance and suggestions which have helped improve the quality of my thesis. I would particularly like to express my appreciation to the chair of my committee, Kevin, as he stuck by me and pushed me when I needed it. This thesis would not have been possible without his help.

Secondly, I would like to thank my research assistants, Cathy Boroski and Dave Wearsch, for their help in data collection and coding. Their assistance enabled me to complete my thesis in a timely manner.

Finally, I would like to thank my family and friends for their encouragement and support throughout the Masters process. My parents, my brother Stuart and sister Debra, Micheline Bast, and Steve Fink all urged me on towards completion of my degree and gave me hell when I needed it. Thanks guys.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
INTRODUCTION	1
Policy Capturing Methodology	5
Steps in Policy Capturing Research	6
Research on Policy Capturing	9
Policy Capturing and Performance Assessment Research	11
Summary of Performance Appraisal Policy Capturing Studies	18
Decision-Making Research	21
Information Processing Approach	28
METHOD	36
Subjects	36
Procedure	36
Design	37
Instructions/Definitions of Performance Dimensions	37
Profiles of Officer Performance	40
Dependent Measures	42
Pretest Results	43
RESULTS	45
Quality of Information	45
Type of Data	52
Consistency	58
Statistical Weights vs. Self-report/Subjective Weights	62
Additional Analyses	66

DISCUSSION	69
Major Results	69
Additional Issues	74
Limitations	79
Future Directions for Research	81
APPENDICES	85
A. STIMULUS MATERIALS	85
Instructions	85
Definitions of the Performance Dimensions for Condition 1	87
Definitions of the Performance Dimensions for Condition 2	90
Definitions of the Performance Dimensions for Condition 3	93
Definitions of the Performance Dimensions for Condition 4	96
Instructions for Condition 5	99
Rating Strategy Questionnaire	101
Police Officer Evaluation Form 1	107
Police Officer Evaluation Form 2	108
Practice Police Officer Ratings Form	109
Police Officer Ratings Form	110
Zedeck and Cascio (1982) Algorithm	111
Police Officer Profiles: Dimension Scale Values	112
Police Officer Profiles: Job Performance Dimension Intercorrelations ...	114
Consent Form	115
Feedback Sheet	116
B. ADDITIONAL TABLES	117
Individual Rater Policies	117
Individual Rater Multiple R^2 for the Policy Capturing Analysis	121
Sum of the Relative and Subjective Weights by Type of Data	123
Spearman Rank-order Correlations between the Statistical and Subjective Weights	127
C. CODING SHEETS	129
For Rating Task	129
For Rating Strategy Questionnaire	130
For Open-Ended Questions	131

FOOTNOTE	133
REFERENCES	134

LIST OF TABLES

Table	Page
1. Pretest Data	44
2. Manipulation Check: Ratings on Credibility, Reliability, Quality by Condition by Dimension	46
3. Analysis of Variance Tests of the Manipulation Check Items by Condition	50
4. Relationship between Ratings of Information Quality and Relative Weights by Performance Dimension	51
5. Means of the Relative Weights by Dimension	53
6. Paired T-test between the Sum of the Relative Weights by Type of Data	55
7. Means of the Subjective Weights by Dimension ...	56
8. Paired T-test between the Sum of the Subjective Weights by Type of Data	57
9. Frequency Distribution of the R^2 Values Obtained from the Policy Capturing Analysis ..	60
10. Relationship between the R^2 Values and the Self-report Measure of Consistency	61
11. Frequency Distribution of the Spearman Rank-order Correlation Between the Relative and Subjective Weights	63
12. T-test between the Relative and Subjective Weights by Performance Dimensions	65

LIST OF FIGURES

Figure	Page
1. Conceptual Model of the Integration Process	20
2. Analytical Model of the Integration Process	20
3. Detailed Model of the Integration Process	29
4. Study Design	38

Introduction

Much of the previous research in the area of performance evaluation has focused solely on the psychometric characteristics of ratings forms (Ilgen & Feldman, 1983). This research has focused on improving the accuracy of performance appraisals and reducing such rater "errors" as halo, leniency, and central tendency. Several new scaling techniques have been developed, including Behaviorally Anchored Rating Scales (Smith & Kendall, 1963) and Behavioral Observation Scales (Latham & Wexley, 1977). Each was designed to reduce rating errors by providing the rater with specific behavioral information to evaluate. Despite the greater complexity in scale development over previous methods, the effects of these instruments on decreasing rater error and/or increasing rater accuracy produced results that were disappointingly similar to those obtained using more conventional scale formats (Bernardin, Alvarez, & Cranny, 1976; Borman & Dunnette, 1975; Dickinson & Zellinger, 1980).

In response to the unsatisfactory progress made in the instrumentation of performance appraisals, a new approach has been proposed that focuses on understanding the appraisal process (Feldman, 1981; Landy & Farr, 1980). Much of this work has taken an information processing

perspective, which adopts views first developed by cognitive psychologists and then adapted to person perception by social psychologists. This approach views the rater as an information processor and examines the cognitive tasks the rater must perform when making judgments concerning others.

According to Feldman (1981), raters must perform several non-independent cognitive tasks before performance appraisals are possible. These include the following:

- 1) Recognize and attend to relevant information about employees (Attention)
- 2) Organize and store information for later access. New information must also be integrated with previously gathered data (Encoding and Storage)
- 3) When judgments are required, relevant information must be recalled in an organized fashion (Recall)
- 4) At various times during the above stages, information must be integrated into some sort of summary judgment (Integration)

(Feldman, 1981, p. 128)

The goal of using an information processing framework to examine the performance appraisal process is to gain a greater understanding of the cognitive processes that a rater performs when making evaluations of others' performance. From this research, methods can be developed which may reduce the level of inaccuracy and/or biases in performance ratings. Given the increased legal requirements

on organizations to have "scientifically valid" personnel practices, this would seem to be an important outcome (Cascio, 1982).

This study will examine Feldman's (1981) fourth cognitive task, that of integration of information, which is thought to be the last task that a rater performs in the rating process. The task of information integration involves assigning weights and combining information gathered previously that concerns a target individual's characteristics or behaviors to form an overall judgment concerning that stimulus person. The key issues of integration include (1) examination of the kind of information a rater has at his/her disposal when making performance judgments, (2) the methods used to assign weights to this information, and (3) the manner in which this information is combined to form an overall judgment.

The paradigm that has been most commonly used to study the information integration process has been policy capturing analysis (Hoffman, 1960). This technique involves the use of multiple regression procedures to develop a statistical representation of an individual rater or group of raters' judgment strategies. Previous research in the application of policy capturing procedures to the performance appraisal domain has mainly been atheoretical, concerned more with demonstrating the efficacy of the

methodology for studying performance evaluation than with understanding the rating process itself. The lack of solid theory-driven research in this area (with the exception of Zedeck & Cascio, 1982) has yielded research which, while demonstrating the potential applicability of policy capturing to performance appraisal, has generally added little to our understanding of the integration process.

One of the major problems of this research has been the lack of attention given to understanding the nature of the information that is being used as inputs into the decisions being made and how different types of information may affect raters' integration processes. Research in social psychology and communication has consistently shown that various characteristics of information such as source credibility, reliability of information, and type of information can all impact on how raters use the information available to them in making attributions or other judgments. (Birnbbaum & Stegner, 1979; Surber, 1981; Weiss, 1979). Previous studies on the integration process in the performance appraisal domain have neglected to focus on this issue. This gap in the research has been noted by DeNisi, Cafferty, and Meglino (1984) who suggested that policy capturing studies should be undertaken to further our understanding of how raters utilize various kinds of information. Thus, the purpose of this study is to determine the effect of systematically

methodology for studying performance evaluation than with understanding the rating process itself. The lack of solid theory-driven research in this area (with the exception of Zedeck & Cascio, 1982) has yielded research which, while demonstrating the potential applicability of policy capturing to performance appraisal, has generally added little to our understanding of the integration process.

One of the major problems of this research has been the lack of attention given to understanding the nature of the information that is being used as inputs into the decisions being made and how different types of information may affect raters' integration processes. Research in social psychology and communication has consistently shown that various characteristics of information such as source credibility, reliability of information, and type of information can all impact on how raters use the information available to them in making attributions or other judgments. (Birnbaum & Stegner, 1979; Surber, 1981; Weiss, 1979). Previous studies on the integration process in the performance appraisal domain have neglected to focus on this issue. This gap in the research has been noted by DeNisi, Cafferty, and Meglino (1984) who suggested that policy capturing studies should be undertaken to further our understanding of how raters utilize various kinds of information. Thus, the purpose of this study is to determine the effect of systematically

varying the quality of information and type of data given to raters on their resulting information integration strategies.

Policy Capturing Methodology

The notion of using policy capturing procedures to model raters' integration and decision-making strategies was initially developed by Hoffman (1960). Hoffman suggested that mathematical models could be developed which link specified stimulus information to judgmental outcomes through the development of a multiple regression equation based on a pooled set of judgments by a rater. This regression equation does not directly model the rater's cognitive processes, but is a "paramorphic" description of the process. Hoffman borrowed the term paramorph from the field of mineralogy, where this term is used to describe a substance having crystalline structural properties which differ from those of another substance with the identical chemical composition. The mathematical representation of the judgment process is analogous to the situation in mineralogy in which two minerals can have identical chemical compositions, but differing underlying molecular structures. The level of analysis used to analyze the different elements determines the different underlying structures.

This notion of "paramorphic" representation of the

judgment process is an important one. Policy capturing procedures do not provide an exact representation of the cognitive processes of raters, but merely analog representations of how raters combine and integrate information on a given task. Process is inferred through the analysis of both the input variables (stimulus information) and the outcomes of the task (decisions made). Policy capturing procedures represent the decision-making process at a general level of understanding, while other methods such as verbal protocols and policy tracing may represent the process at a more specific level (Einhorn, Kleinmuntz, & Kleinmuntz, 1979). Researchers adopting the policy capturing methodology have tended to neglect the "paramorphic" representation notion from Hoffman's work thus obscuring some fundamental notions concerning the limits of policy capturing procedures (Schmitt & Levine, 1976).

Steps in Policy Capturing Research

The policy-capturing methodology involves the following sequence of steps. First, a number of profiles concerning various characteristics of hypothetical or real persons are developed or collected. The information contained in the profiles usually includes a limited number of pieces of information or cues (3 to 10) which are represented as numerical or categorical responses. Types of cues that have

been used in policy capturing studies previously include such cues as job performance dimensions (Hobson, Mendel, & Gibson, 1981; Naylor & Wherry, 1965; Stumpf & London, 1981; Taylor & Wilsted, 1974; Zedeck & Kafry, 1977; Zedeck & Cascio, 1982), MMPI profiles (Goldberg, 1971), stock-market data (Ebert & Kruse, 1978), and even information concerning safeguarding nuclear power plants (Brady & Rappaport, 1973). The number of profiles in a given set typically is a function of the number of cues contained in each profile, with at least a 10:1 ratio of profiles to pieces of information suggested as being necessary for stable results (Dawes, 1979). Once the profiles are developed (see later section for description of some of the problems associated with profile development), the rater's task is to analyze and integrate the information given in each profile into an overall assessment or decision. The typical procedure in performance appraisal policy capturing studies requires raters to make an overall rating of performance for each of the target ratees based on a set of job performance criteria.

After completion of the ratings, subjects are usually asked to state their "subjective" rating policy. The subjective rating policy is the rater's notion of the relative importance and weighting of cues which were used when making the judgments. The method for obtaining the

subjective weights typically requires raters to distribute 100 points among the sources of information available (cues) in such a way that this distribution reflects the relative importance of those variables to the final decision (Hoffman, 1960). The 100 point allocation method has been most commonly applied, although other methods have been used (Cook & Stewart, 1975; Doherty & Keely, 1972).

The major step in the analysis of the results of a policy capturing study is the development of a multiple regression equation for each rater by regressing the overall rating onto the values of the cue elements contained in each profile. The resulting multiple R^2 is evidence for how well the cues account for the linear portion of the variance in the overall ratings, while the beta weights obtained for each cue element paramorphically represent the weighting scheme used by the rater. A comparison is then made between the "objective" (beta) weights obtained through the multiple regression procedure and the "subjective" (stated) weights to determine how well a rater's stated rating policy matches the statistical policy. Finally, a clustering procedure, is often utilized to cluster raters based on the similarity of their rating strategies to determine if consistent rating patterns emerge across raters.

Research on Policy Capturing

Early attempts at using the policy capturing methodology were in the area of modeling clinicians' judgments. The majority of these efforts attempted to show that statistical combination methods were superior in prediction to clinical combination. In his original study, Hoffman (1960) presented raters with profiles of scores obtained from the Edwards Personal Preference Schedule and asked them to make judgments concerning stimulus persons' sociability and intelligence. He found that for the two judges used, the multiple R^2 were .837 and .937, and that the correlation between the best linear combination of predictor scores and actual judgments correlated .829 and .948 respectively. Hoffman also found that the subjective weights given by the judges differed markedly from the objective weights obtained from the regression equation. ¹

Using MMPI profiles as stimulus materials, Goldberg (1971) required judges to differentiate neurotic from psychotic patients. Goldberg found that linear models of clinicians' judgments had a multiple correlation of .78 for the average judge, and .89 when a composite judge was created by pooling the ratings made across judges.

Sawyer (1966) reviewed forty-five studies which directly compared the results obtained from policy capturing procedures with those obtained from clinical judgment. He

differentiated the mode of data collection (whether the data was collected by a clinician or obtained through analysis of records, tests, self-report inventories, etc.) from the mode of data combination (whether the data was combined by clinicians' judgment or combined through statistical techniques). He developed a 4 (mode of collection: clinical, mechanical, both, or miscellaneous) X 2 (mode of combination: clinical or mechanical) chart and found that predictions made on the basis of statistical combination techniques were clearly superior to those made by a clinical judgment/combination of information.

A number of other studies have applied the policy capturing methodology beyond clinical settings to decisions concerning the admission of students to graduate school. Dawes (1971) used undergraduate grade point average, quality of the undergraduate institution, and scores from the Graduate Record Exam to predict admission committee ratings (at the time of admission) and faculty evaluations of the students' performance in graduate school. The multiple correlation of these cues predicting the admissions committee rating was .78. In addition, the researchers employed "bootstrapping," a procedure in which a linear model of the judges' decision-making strategies is developed and then used in place of the actual judges in prediction. In this study, the mathematical representation of the

raters' policies outperformed the raters' actual judgments in predicting the results of first-year evaluations of a new set of students.

Policy Capturing and Performance Assessment Research

In recent years, the performance appraisal literature has moved towards taking a cognitive information processing approach to examining how raters make judgments. The introduction of the person perception literature into this area has highlighted the role of cognition as being an important component in the appraisal process. Researchers have begun drawing from this literature as well as the literature on decision-making to help further their understanding of the cognitive tasks a rater must perform when making judgments. The use of policy capturing procedures to model performance rating related decision strategies fits in nicely with this approach.

Several studies applying policy capturing methods have been conducted in the performance evaluation domain. The major focus of these studies has been on determining the raters' internal consistency in making ratings, as determined by individual judges' R^2 , clustering of raters who used similar rating policies, and comparison of raters' subjective/stated rating policies with their objective/statistically determined weights.

The first major application of policy capturing procedures to a performance appraisal context was a study conducted by Naylor and Wherry (1965). Naylor and Wherry developed 250 profiles of hypothetical stimulus persons, each containing scores for that individual on 23 job performance related traits. Fifty raters were asked to make overall judgments of performance for each of the profiles, in terms of the stimulus person's worth to the Air Force. Results indicated that R^2 values ranged from a high of .973 to a low of .569. Application of the JAN clustering procedure (Christal, 1968), in an effort to cluster raters with similar rating policies, yielded the result that the rater equations were basically homogeneous in nature, indicating that the group of raters were quite consistent in the cues they used to make their ratings.

Taylor and Wilsted (1974) modeled the rating policies of United States Air Force officers' ratings of cadets. The raters were asked to make overall judgments of cadet performance based on ten subjective performance factors (e.g. initiative, expression, cooperation, leadership, etc.). Twenty-five raters each rated twenty-five cadets and stepwise regression equations were developed for each rater. The results revealed that raters were internally consistent in their ratings, with R^2 ranging from .92 to .99. This implied that cadets rated by the same rater tended to be

The first major application of policy capturing procedures to a performance appraisal context was a study conducted by Naylor and Wherry (1965). Naylor and Wherry developed 250 profiles of hypothetical stimulus persons, each containing scores for that individual on 23 job performance related traits. Fifty raters were asked to make overall judgments of performance for each of the profiles, in terms of the stimulus person's worth to the Air Force. Results indicated that R^2 values ranged from a high of .973 to a low of .569. Application of the JAN clustering procedure (Christal, 1968), in an effort to cluster raters with similar rating policies, yielded the result that the rater equations were basically homogeneous in nature, indicating that the group of raters were quite consistent in the cues they used to make their ratings.

Taylor and Wilsted (1974) modeled the rating policies of United States Air Force officers' ratings of cadets. The raters were asked to make overall judgments of cadet performance based on ten subjective performance factors (e.g. initiative, expression, cooperation, leadership, etc.). Twenty-five raters each rated twenty-five cadets and stepwise regression equations were developed for each rater. The results revealed that raters were internally consistent in their ratings, with R^2 ranging from .92 to .99. This implied that cadets rated by the same rater tended to be

The first major application of policy capturing procedures to a performance appraisal context was a study conducted by Naylor and Wherry (1965). Naylor and Wherry developed 250 profiles of hypothetical stimulus persons, each containing scores for that individual on 23 job performance related traits. Fifty raters were asked to make overall judgments of performance for each of the profiles, in terms of the stimulus person's worth to the Air Force. Results indicated that R^2 values ranged from a high of .973 to a low of .569. Application of the JAN clustering procedure (Christal, 1968), in an effort to cluster raters with similar rating policies, yielded the result that the rater equations were basically homogeneous in nature, indicating that the group of raters were quite consistent in the cues they used to make their ratings.

Taylor and Wilsted (1974) modeled the rating policies of United States Air Force officers' ratings of cadets. The raters were asked to make overall judgments of cadet performance based on ten subjective performance factors (e.g. initiative, expression, cooperation, leadership, etc.). Twenty-five raters each rated twenty-five cadets and stepwise regression equations were developed for each rater. The results revealed that raters were internally consistent in their ratings, with R^2 ranging from .92 to .99. This implied that cadets rated by the same rater tended to be

rated on similar criteria. Rating policies, though, varied between raters, as the beta weights of the raters' regression equations showed that the dimensions that raters used to rate the cadets were not consistent across raters. Finally, raters' subjective or stated policies differed from their objective or statistical policies. In most cases, raters overestimated the number of performance factors used in making their decisions; the results suggested that the overall rating could be predicted quite well using only 3 out of the 10 cues.

Zedeck and Kafry (1977) used policy capturing procedures in assessing ratings of performance of public health and registered nurses. Profiles of forty hypothetical nurses were developed that contained information on nine criterion elements, with three levels of performance possible on each dimension. The profiles were developed such that the intercorrelation between dimensions approximated zero (see Hoffman, 1960 for a discussion of correlated dimensions). Raters were asked to make ratings for each of the stimulus profiles, and to assign subjective weights to each of the nine criterion elements. In addition, several cognitive and personality tests were administered to raters in an attempt to predict differences between clusters of raters. Results indicated that the R^2 values ranged from .41 to .77 for the public health nurses sample, and .20 to .90 for the

registered nurses sample. Comparison of the subjective and objective weights indicated that, like the Taylor and Wilsted (1974) study, the subjective and objective weights differed significantly. Finally, two clusters of raters were found using the JAN procedure. Attempts to relate these clusters to the background and individual difference measures were unsuccessful. The authors concluded that the sample was basically homogeneous on the characteristics measured.

Stumpf and London (1981) investigated managerial promotion policies with specific regard to 1) the extent to which judges used linear and non-linear composite criteria modes, 2) the existence of clusters of raters with similar policies, and 3) the similarity of statistically derived and subjectively evaluated weighting of the criterion elements. Forty-eight hypothetical candidates for promotion in a commercial bank were evaluated by separate groups of managers and students on the basis of five criterion elements: managerial potential ratings, recommendations, position, candidate weakness, candidate sex, and situational and individual differences factors.

A $2 \times 2 \times 3 \times 2 \times 2$ ANOVA was run and the results indicated that there were three significant main effects (potential, position, weakness) as well as a significant two-way interaction between potential and recommendation, and a

significant three-way interaction of potential by recommendation by weakness. Overall, these six effects accounted for 70% of the variance in raters' judgments. Multiple Discriminant Analysis procedures were used to cluster raters and the results revealed that six clusters of raters emerged for both the managerial and student samples. Individuals in the clusters were differentiated by the weighting of particular individual criterion elements. In other words, across clusters, raters differed in the criterion element which accounted for the largest portion of the variance. Finally, a Spearman rank-order correlation was computed between the subjective rankings of criterion importance and the empirically derived weights and yielded a correlation of .67. When the proportion of variance due to the interactional use of the cues was taken into account, the correlation increased to .74, representing a fairly high level of agreement. Overall, Stumpf and London (1981) concluded that raters can and do employ non-linear weighting of criterion elements (as indicated by the significant interactions), and that subjective rating procedures may be deficient to the extent that they do not allow raters to state or use such non-linear policies in reporting their subjective weighting policies. More on the issue of non-linear use of cues will be discussed below.

Hobson, Mendel, and Gibson (1981) developed a set of 100

hypothetical performance profiles concerning college faculty member performance. Fourteen performance cues were utilized covering four broad areas of performance: instruction, instructional support, professional activities, and interpersonal skills and image. Each of the fourteen dimensions were given three behavioral anchors, which were assigned values of below average, average, or above average performance. Raters (19 faculty members) were instructed to assign an overall rating to each of the stimulus profiles using a 1-9 scale. In addition, subjects were asked to assign subjective weights to each of the performance dimensions twice: once indicating the importance they felt should be attached to the performance elements, and a second time indicating the relative importance of the elements they felt their department head utilized in his ratings.

Results from the study revealed that, (1) the raters' R^2 ranged from .61 to .94 (median R^2 was .77), which indicated that raters were fairly consistent in their utilization of the cue information and (2) raters' subjective policies differed greatly from their objective rating policies, with raters overestimating the number of cues that they utilized in the rating task (13 vs. 9). In fact, three dimensions accounted for approximately 71% of the predictable variance in the raters' judgments. In addition, subjects apparently had relatively poor insight into their supervisors' rating

policy, as the dimensions of performance which the supervisor used in making his judgments were different from those that the subjects felt he used. Finally, clustering raters with like rating policies yielded four distinct clusters of raters who could be identified in terms of their orientation towards teaching vs. research vs. administration, which in turn, was related to the age and tenure of these individuals in the university.

Zedeck and Cascio (1982) looked at the effects of rater training and purpose of appraisal on performance appraisal decisions. They developed 33 one-paragraph descriptions of hypothetical supermarket checkers' performance, with the descriptions of performance containing information on five behavioral dimensions, with three levels of performance possible for each dimension. Subjects in this study were divided into two groups: one group which received rater training on the reduction of common rating errors (leniency, central tendency, halo, etc.) and a control group which did not receive the rater training. The purpose of appraisal factor (development, merit raise, retention) was nested within the training factor. The dependent variable used in this study was the rater's standard deviation of his or her evaluations across the 33 paragraphs. Zedeck and Cascio (1982) hypothesized that six clusters of raters would emerge, consistent with the 2 (training vs. no training) X 3

(purpose of appraisal) design.

An analysis of variance was run and the results showed that the only significant effect was for purpose of appraisal, with the differences coming between the merit raise vs. developmental and retention conditions. The R^2 for the individual raters ranged from .05 to .51, with the modal R^2 ranging from .21 to .25. ² These R^2 values are much lower than the other studies previously mentioned. The authors noted, however, that raters were quite accurate in their ratings; profiles which were described in terms of more positive performance were evaluated more positively on the criterion. The JAN clustering procedure was used to cluster raters with similar policies, and the results mirrored the ANOVA results, i.e., the individuals in the clusters differed by the purpose of appraisal condition. The authors concluded that this study demonstrated that policy capturing methodology could be used for hypothesis testing and/or providing insight into raters' judgment behaviors.

Summary of Performance Appraisal Policy Capturing Studies

As Hobson and Gibson (1983) have noted, there are a number of consistent findings with regard to the policy capturing studies in the performance appraisal domain. First, the general linear model has worked well in describing rater policies. With the exception of the Zedeck

and Cascio (1982) study, R^2 values have been consistently high. Second, there is evidence that raters' subjective policies are dissimilar to their statistical or objective rating policies. It has been typically found that raters' subjective policies overestimate the number of statistically significant cues obtained from the regression analysis. Third, differences between raters in their rating policies have been found when using clustering procedures to group raters with similar weighting/judgment policies. Numbers of different rating clusters found in the studies range from 1 to 6.

Although these findings are meaningful in demonstrating the efficacy of applying policy capturing procedures to the performance appraisal domain, little attention has been paid to understanding the cognitive processes that raters perform when making their judgments. These studies focused on the outputs of the rating process, i.e., prediction of overall judgments of performance, with little regard to both the inputs (the performance information used to form the judgments) and the mediating processes which intervene between inputs and outputs (Schmitt & Levine, 1977). (See Figure 1)

The distinction between inputs, processes, and outcomes, however, is not a clear one. The basic notion behind policy capturing research is that raters' processes can be inferred



Figure 1. Conceptual Model of the Integration Process

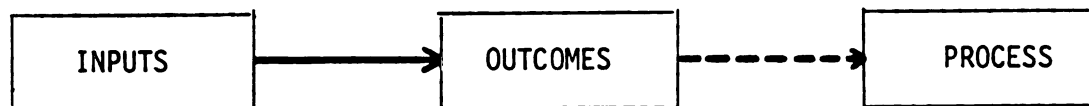


Figure 2. Analytical Model of the Integration Process

through a joint examination of both input factors and the outcomes or decisions made. The analysis (see Figure 2) differs from the conceptual model in that the rating process is inferred following assessment of inputs and outputs. Again, drawing on the paramorphic analogy, process can not be directly measured but is assessed post hoc through the analysis of hypotheses concerning the relationship between the inputs and outcomes. As Payne, Braunstein, and Carroll (1978) note, "Observations of the relations between inputs and outputs can be used to test process rules when different rules for transforming inputs imply different types of outputs" (p.18). It is presumed that various types of information inputs will lead to different types of decision outcomes, and that this occurs because different cognitive processes are operating.

Decision-Making Research

Researchers in the more generalized area of decision-making (see Slovic & Lichtenstein, 1971 for a review) have done some work on the issues concerning the nature of inputs and processes used by raters making judgments. Specific attention has been focused on analyzing the various effects of different amounts and kinds of stimulus information provided to the rater, as well as on the raters' combination processes.

One issue that has been investigated by decision theorists concerns the impact of the number of cues presented to the rater on the R^2 of the raters' regression equation (the consistency with which raters weight and combine the cues into an overall judgment). The higher the R^2 obtained, the greater consistency of the rater. Results of studies on this topic have been mixed. Einhorn (1971) compared subjects' R^2 obtained using two, four, or six cues as stimuli. He found that the values of R^2 increased with fewer numbers of cues. In addition, subjects reported that they felt the task was more difficult with an increasing number of cues. Cook and Stewart (1975) and Billings and Marcus (1983) also found higher R^2 with fewer cues, as compared to judgments with a greater number of cues. However, Anderson (1977) compared R^2 for tasks involving four, six, or eight cues and found no differences across conditions. Although no clear conclusions from this research can be drawn, one would expect that the R^2 would be higher given a small number of cues, since such tasks should be less cognitively complex than those with a greater number of cues. Large numbers of cues may overload our information processing capacity. Thus, it would be easier to weight or combine a small number of cues in a consistent fashion. When a large number of cues are given, subjects may cognitively reduce the set of cues to a more meaningful number so that

they are better able to process this information (Miller, 1956). This suggestion has been borne out in the research presented earlier in which the researchers found that a small subset of the presented cues usually account for large proportions of the variance in raters' judgments (e.g., Taylor & Wilsted, 1974).

A related issue concerns the level of intercorrelation among the stimulus material dimensions. Schenk and Naylor (1968) showed that as the amount of cue intercorrelation increases, subjects' responses become more systematically a linear function. In other words, as the dimension intercorrelations increase, the R^2 for each rater should increase accordingly, solely on the basis of this statistical artifact. Due to Schenk and Naylor's suggestions, research using policy capturing in the performance appraisal domain has artificially constrained the intercorrelations between dimensions to be zero, which probably does not accurately reflect ecological reality (Hobson & Gibson, 1983; Schmitt and Levine, 1977). Subjects may be operating on the basis of their intuitive notions of how performance dimensions vary and, thus, may not be sensitive to the "actual" degree of intercorrelation among the dimensions presented to them in the stimulus materials (Cooper, 1981; Kozlowski, Kirsch, & Chao, in press). Lane, Murphy, and Marques (1983) suggest that one way to avoid

this problem is to use raw score regression weights, which, because they remain constant across differing cue intercorrelation matrices, would enable researchers to employ perhaps a more realistic non-orthogonal cue structure.

A third issue relating to the nature of cues used in policy capturing tasks is that of cue format. In the only study that has directly addressed this issue, Anderson (1977) compared cues presented in verbal/paragraph form with cues presented in numerical form. Her results indicated that subjects were more consistent in their ratings when rating numerical cues than when rating verbal cues. One possible explanation for this finding might be that raters were not able to accurately determine the "true" level of performance of the stimulus ratees, since the behavioral examples used in this study did not unambiguously identify the true levels of performance. Additional evidence for this notion comes from a study conducted by Cotton, Jacobs, and Grogan (1983) which found that using individually scaled cue values resulted in judgment models which were more successful in reproducing the decision-makers' responses than models employing normatively scaled cues.

Research on the set of issues related to the process component have generally been concerned with discovering whether judges are combining cues in a linear or non-linear

fashion. The three major studies which have made direct comparisons between the efficacy of linear versus non-linear models ability for prediction have yielded conflicting results. Einhorn (1971) found that non-linear models outperformed the linear models, while Goldberg (1971) and Ogilvie and Schmitt (1979) found that linear models outperformed non-linear models. As Goldberg (1971) suggested, there were important differences in the nature of these studies which could have resulted in the conflicting conclusions. These differences include: the kind of judges, the type of task, the number of cues, the intercorrelations among the cues, type of responses required (rating vs. ranking), values for cues (discrete vs. continuous) and the number of cases being evaluated.

A number of these factors relate to the previous discussion regarding the nature of the inputs or cues utilized as stimulus materials, while several others draw attention to additional factors which might be relevant to the rating process. The notion that the type of task or the nature of the decision required can influence whether raters use cues in a linear or non-linear fashion is an important one. This notion is consistent with the work done using policy-tracing (verbal protocol) procedures which show that rating tasks require different kinds of cognitive processing than choice/preference tasks, in that the latter require

more configural use of cues than do the former (Billings & Marcus, 1982; Payne, Braunstein, & Carroll, 1978; Svenson, 1979).

Another issue that has been raised concerns the relationship between subjective or stated rating policies and the objective or statistical rating policies. Following Hoffman's (1960) suggestions, most of the previous studies using policy capturing procedures have used the method of asking subjects to distribute 100 points among the cues, in order to obtain subjects' subjective weighting of the dimensions. Using the statistical weights that contribute significantly to the regression equation as the measure of objective cue usage, results from these studies have consistently indicated that subjects overestimate the number of cues that they actually use in making their judgments. Questions have been raised as to whether this method allows raters the opportunity to state that they are using cues in a non-linear fashion. Cook and Stewart (1975) addressed this issue through comparison of seven different techniques for obtaining subjective weights. They compared the traditional method with both additional linear and non-linear methods and found that there were no major differences between the methods. The authors concluded that the 100 point allocation method was as good as any other, and recommended its continued use, primarily because it is probably the simplest

method to use.

A final issue concerns whether subjects should state their policies before or after completion of the rating task. Balzer, Rohrbaugh, and Murphy (1983) found that subjects who completed the rating task first had significantly higher reliabilities for their predictions based on their subjective policies than did subjects who completed their subjective policies before completing the rating task. The authors hypothesized that this result was due to raters having an opportunity to monitor their decision behavior and assess their strategies "in practice" when giving their subjective weighting policies after making the ratings.

As can be seen from the review of the decision-making literature, the focus of the decision-making researchers has been quite different from that of the researchers working in the performance appraisal domain. Much more attention in the decision-making literature has been given to discovering the effects of information inputs on judgment strategies. Although the performance appraisal literature has used such varied stimulus inputs as subjective performance dimensions (Taylor & Wilsted, 1974), behavior-oriented performance rating scales (Hobson, Mendel, & Gibson, 1981; Zedeck & Cascio, 1982; Zedeck & Kafry, 1977), and candidate promotion qualifications, including sex, position, managerial

potential ratings, recommendations (Stumpf & London, 1981), the focus of these studies was not on directly identifying how characteristics of these inputs might impact on the integration process.

Information Processing Approach

The information processing approach to performance appraisal suggests that a critical factor affecting the rating process is the kind of information a rater has at his/her disposal when making decisions (Feldman, 1981). The cognitive task of interest that a rater must perform is the development of a strategy for evaluation of that information. This would imply that some method for weighting and combining the information in a consistent manner must be developed. It is suggested that one important factor affecting the rules or policies developed by raters concerning information is based on characteristics of the information itself, and that these characteristics impact on how the information is used. The hypothesized integration process is illustrated in Figure 3.

There are a number of performance related characteristics of information that raters might attend to when making performance ratings including source of the information, reliability of the information, and type of information. Source credibility is defined as

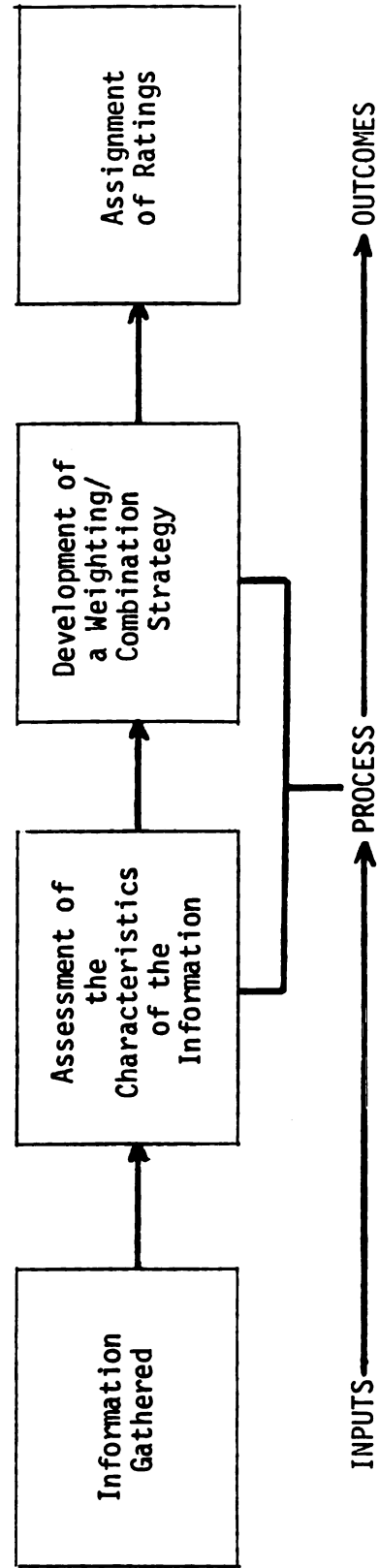


Figure 3. Detailed Model of the Integration Process

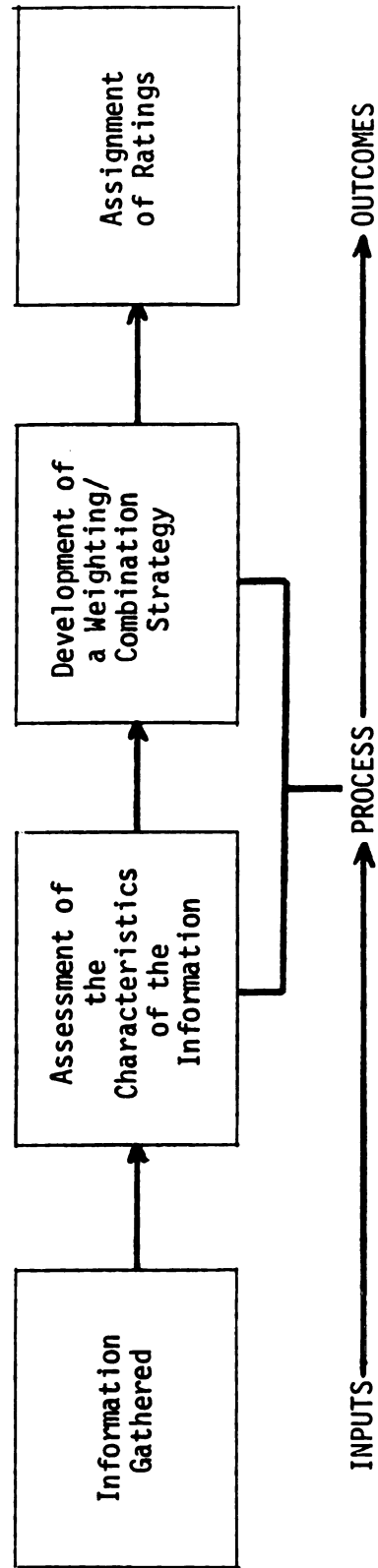


Figure 3. Detailed Model of the Integration Process

"believability" of the source of the information, and thus "believability" of the information itself. Birnbaum and Stegner (1979) identified two components of source credibility: expertise of the source and bias of the source. Expertise of the source refers to the perceived correlation between the source's report and the outcomes of interest, and is thought to be dependent on the training, experience, and/or ability of the source. Bias of the source refers to factors that are perceived to influence the expected algebraic difference between the source's report and the true state of nature. Research on source credibility in the social psychological literature has consistently shown that the values associated with information cues in a decision-making task are monotonically related to the credibility of the source of that information. More specifically, it has been found that the higher the credibility of the source of the information, the greater weight placed on those elements in judgment, as well as the less weight placed on other elements in that decision set (Birnbaum, Wong, & Wong, 1976; Birnbaum & Stegner, 1979; Rosenbaum & Levin, 1969).

Researchers studying organizational decision-making processes have also examined the impact of source credibility on the use of information. In a recent review of the literature on information usage, O'Reilly (1983) found

that managers have a bias towards using information from trustworthy or credible sources, such that information obtained from credible sources is more likely to be utilized in decisions made. Beach, Mitchell, Deaton, and Prothero (1978) examined the information use of subjects when evaluating the probability of success and the acceptability of hypothetical job candidates. Their results indicated that information use was related to information relevance and source credibility, as subjects used the information given to them to a greater extent when it was obtained from high rather than low credibility sources.

A second characteristic of information that might impact on a rater's integration process is reliability of the information. Reliability of information refers to its freedom from unsystematic errors of measurement or its consistency under different conditions that might introduce error into the scores (Aiken, 1979). Surber (1981) manipulated reliability of information concerning the effects of ability and effort on students' performance by varying the information given to subjects concerning the reliability of scores from an IQ test and the amount of time students spent studying. Surber found that the higher the reliability of the effort/ability information, the greater its effect was on judged performance. This suggests that the reliability of a piece of information affects the weight

that will be given to that piece of information by raters when integrating the information into an overall judgment.

For the purpose of the present study, source credibility and reliability of the information will be combined into a variable called information quality. (The effects of source credibility and reliability will not be examined separately, but will be assumed to covary.) Information quality will be defined here as the usefulness of a piece or set of information in describing job performance. It is hypothesized that the higher the quality of information of a particular information cue, the greater the weight placed on that element by raters when integrating information to make performance judgments.

H₁: Raters will utilize information of high quality to a greater extent than information of low quality when making performance rating decisions.

Another characteristic of information that might impact on a rater's integration process is the type of information the rater has available when making judgments. The standard distinction concerning type of data in the Industrial/Organizational psychology literature has been objective versus subjective data (Smith, 1976). Objective data are measures of the results of behavior or outcomes, such as production data (e.g. number of units produced,

number of errors, etc.), as well as personnel data (e.g. number of absences, turnover, tardiness, etc.). Although objective data require some level of judgment (Muckler, 1982), subjective data are considered to rely on human judgment to a much greater extent in determining the level of performance. Subjective data in the performance appraisal domain typically consist of a set of ratings concerning performance-related traits or behaviors made by the supervisor or peer of the target ratee, or by the ratee him/herself (Cascio, 1982).

Research examining the relationship of subjective and objective performance measures have found that the relationship is typically low. These results indicate that subjective and objective measures may be tapping different aspects of the construct of job performance (Alexander & Wilkins, 1982; Bass & Turner, 1973; Cascio and Valenzi, 1978). Raters, when attempting to combine information obtained from both subjective and objective data, must develop a strategy for the combination of this information.

Research on the integration of information obtained from both subjective and objective data sources has indicated that individuals seem to have a bias towards using information which has been obtained from subjective or personal experience sources over more abstract statistical data (Tversky and Kahneman, 1974). Hogarth (1980) posited

the following explanation for this phenomenon:

"Information that is concrete (based on subjective experience) is more salient in memory than information that is abstract. That is, information that is vivid, e.g. describing an experience or perhaps involving a personal incident, is more easily recalled than, for example, statistical summary data.... Data coded in memory by images and through several associations can become disproportionately salient " (p. 161).

Hogarth (1980) also adds that a mixture of the two types of information during the acquisition phase can lead to a concentration on one type of data to the exclusion of the other. Nisbett and Ross (1980) also noted that statistical information, by its very nature, may lack the "force" necessary for subjects to attend to and use, and that it is too abstract and dry for people to evaluate. An example of the bias towards the use of subjective data over objective data is the continued use of the interview in the selection process by organizations, even though research has shown that decisions made on the basis of objective measures are typically more valid than decisions made on the basis of interviews (Schmitt, 1976).

Thus, it is believed that raters, when in a situation in which integration of data obtained from subjective and objective data is required, will utilize the subjective information in making decisions to a greater extent than the

subjective information.

H₂: Raters will place greater weight on subjective information rather than objective information when making performance rating decisions.

In addition to examining the above hypotheses, several supplemental analyses will be performed to assess raters' knowledge and awareness of the rating policies that they used in assigning their ratings. First, raters subjective rating policies will be compared with their objective or statistically determined policies to determine if the weights subjects felt that they used in assignment of the ratings match the weights obtained from the policy capturing analysis. Second, a post-experimental rating strategy questionnaire will be administered to further investigate raters' awareness of the policies used in making the ratings. It is expected that this questionnaire will provide additional insight into the raters' integration processes and mirror the results obtained from the policy capturing procedure.

Method

Subjects

The subjects in this study were 104 undergraduate students enrolled in psychology and business courses at a large midwestern university. Subjects participated in the study either on a voluntary basis or for nominal course credit. Three subjects' responses were dropped from the analysis due to missing data, and one subject's responses were dropped due to incorrectly making the ratings. The effective sample size of the study was 100 persons.

Procedure

Subjects were tested in groups ranging in size from one to ten persons. Subjects in each session were randomly assigned to the same experimental condition or to the control group. Sessions were run over a five week period, until usable responses from 100 subjects were obtained.

Each subject initially received a packet of materials containing instructions, definitions of the performance dimensions, and forty-five profiles of police officer performance. (Copies of the experimental materials can be found in Appendix A). Upon entering the experimental room, subjects were seated at desks and handed the packet of materials. When all of the subjects for an experimental

session had arrived, the experimenter read the instructions and definitions of the performance dimensions out loud to the subjects. Subjects were then given an opportunity to ask questions before beginning the rating task. Subjects completed 5 practice ratings, before completing the ratings on the 40 experimental profiles. Upon completion of the ratings task, subjects were asked to complete the rating strategy questionnaire. Following completion of the questionnaire, subjects were debriefed and dismissed.

Design

There were four experimental groups in this study; each group received a different set of information concerning the officer performance dimensions. Figure 4 illustrates the experimental design used in this study. Two variables were systematically varied: quality of information and type of data, resulting in a 2 (quality of information: high or low) X 2 (Type of data: objective vs. subjective) design, with repeated measures on the second factor. In addition, a control group was used which did not receive any information concerning the performance dimensions.

Instructions/Definitions of Performance Dimensions

A one page description of the rating task was provided to participants. This description informed participants that

INFORMATION QUALITY	TYPE OF DATA	
	SUBJECTIVE	OBJECTIVE
	GROUP 1 High	High
	GROUP 2 High	Low
	GROUP 3 Low	High
	GROUP 4 Low	Low
	CONTROL GROUP -	-

Figure 4. Study Design

they were to play the role of police captain and that their task was to assign overall ratings of performance to each of forty police officers in their squad. Subjects were instructed to assign a rating of 1 to 9 to each police officer, based on the information contained in that officer's job performance information sheet.

Following this instructions sheet, definitions of the police officer performance dimensions were provided. The six performance dimensions used were composed of three dimensions identified as being obtained from supervisor ratings (job knowledge, initiative, and attitude) and three dimensions identified as being obtained from personnel records (number of arrests, number of absences, number of community grievances/complaints). The experimental manipulation of this study was the variation of information concerning the police officer job performance dimensions. Each group of subjects received information concerning the source and reliability of the subjective and objective performance dimensions (information quality), as well as definitions of the performance dimensions themselves. The descriptions of the dimensions were systematically varied such that each group received different information about the information quality of the performance dimensions. Examples of dimension descriptions of high and low information quality are provided below.

"Number of Arrests: This figure represents the number of arrests made by the officer during the previous six months. This figure has been standardized across officers, such that biasing factors which would artificially inflate/deflate this figure, such as location (i.e. suburbs vs. the inner city) and assignment (i.e. foot patrol vs. squad car) have been taken into account. This standardization procedure allows for comparisons across officers to be made. This measure is considered by the officers to be a good measure of an officer's productivity. (High Information Quality)

"Number of Arrests": This figure represents the number of arrests made by the officer during the previous six months. The number of arrests made is subject to fluctuations due to location of assignment (i.e. suburbs vs. the inner city) and type of assignment (i.e. foot patrol vs. squad car). Because of these variations, which are not taken into account in this measure, comparisons across officers are hard to make. In addition, because this measure is used as the basis for promotion decisions, some officers "pad" their arrest figures by making arrests for minor violations, many of which are subsequently thrown out of court. Due to these problems, police officers do not consider the number of arrests to be a good measure of an officer's productivity. (Low Information Quality)

Profiles of Officer Performance

Each subject received the same forty-five profiles of police officer performance. Each profile contained ratings for the following six dimensions of officer performance: job knowledge, initiative, attitude, number of arrests, number of absences, and number of community complaints/grievances. The first three dimensions comprised the subjective data. The definitions of these dimensions were obtained from a performance appraisal scale developed by Landy and Farr (1975). The second three dimensions comprised the objective data, and were based on a study completed by Cascio and

Valenzi (1978). To prevent possible order effects of the dimensions, there were two sets of profiles. One half of the subjects' profiles presented the supervisor rating dimensions first, followed by the personnel data dimensions; the other half of the subjects' profiles presented the personnel data dimensions first, followed by the supervisor rating dimensions. Subjects randomly recieved one or the other set of profiles depending on their subject identification number (Persons with odd identification numbers received profiles with the supervisor rating dimensions presented first, while persons with even identification numbers received profiles with the personnel data dimensions presented first.).

The profiles were developed such that the intercorrelations among dimensions approximated zero, and the mean values for the dimension scores across profiles were approximately 5.0 (See Appendix A for a copy of the algorithm which was used for the development of the scores as well as the actual dimension scale values that were used in the study). The algorithm was based on previous work done by Cascio and Zedeck (1982). Each performance dimension received a score ranging from 1 to 9, with 1 representing poor performance, 5 representing average performance, and 9 representing above average performance. The first 5 profiles of officer performance in the packet were used to

familiarize subjects with the task.

Dependent Measures

There were three major dependent variables used in this study: relative weights, subjective weights, and responses obtained through a self-report questionnaire of rating policy. The relative weights were obtained through a policy capturing analysis, in which the overall ratings of performance made on each profile were regressed onto the dimension scale values for each individual. The Beta weights were then transformed into relative weights using Hoffman's (1960) formula. Raters' subjective policy weights were obtained through the subjects' distribution of 100 points among the six dimensions of job performance according to the weighting scheme they used when making the ratings. A self-report rating strategy questionnaire (see Appendix A) was developed to further assess subjects' knowledge and awareness of their rating policies. There were two open-ended questions regarding subjects' rating strategies. A coding scheme was developed to content analyze these questions (see Appendix C) and individual subject's responses were coded and tabulated. In addition to these questions, there were a number of close-ended questions which further assessed raters' perceptions of the strategies used when completing the rating task. Responses to the

self-report rating questionnaire were analyzed through a frequency distribution of responses. Finally, there was a manipulation check regarding the subjects' perceptions concerning the quality of information manipulation. Subjects were asked to rate each dimension of performance in terms of the reliability, source credibility, and information quality of that dimension.

Pretest Results

The manipulation of information quality was pretested on a sample of 30 undergraduate psychology students. Subjects were given the description of the performance dimensions of both high and low information quality and asked to give ratings on a scale from 1 to 5 on the source credibility, reliability, and information quality of each of the descriptions. Means and standard deviations of the ratings by performance dimension are presented in Table 1. As can be seen, the manipulation was effective except for the dimensions of arrests and community complaints/grievances. The descriptions of these dimensions were rewritten in an attempt to make the manipulation more salient to the subjects in the study.

Table 1. Pretest Data

Dimension	Source Credibility		Reliability		Information Quality	
	Mean	SD	Mean	SD	Mean	SD
Supervisor Ratings (HI)	4.58	.70	4.35	.89	4.50	.99
Supervisor Ratings (LO)	1.81	.93	1.65	.68	1.92	1.02
Arrests (HI)	3.81	.69	3.77	.95	3.54	.98
Arrests (LO)	3.81	.98	2.69	1.29	2.69	1.19
Absences (HI)	4.46	.58	4.42	.75	4.12	.99
Absences (LO)	2.12	.90	1.85	.83	1.89	.81
Complaints (HI)	3.46	1.17	2.96	.95	3.27	.91
Complaints (LO)	2.07	1.06	1.77	.81	1.81	.93

Results

There were a number of issues that were examined in this study. First, the effects of the two major variables of interest, information quality and type of data, on rater integration strategies were examined. Second, raters' consistency in making their ratings was investigated. Third, the results of the statistical policy-capturing analysis were compared with the results of the self-report of policy information. (Particular emphasis was focused on the relationship between the statistical weights obtained from the policy capturing analysis and the subjective weights obtained from the rating strategy questionnaire). These topics were investigated using information obtained from three different sources: the relative weights obtained from the policy capturing analysis, the subjective weights obtained from the rating strategy questionnaire, and the self-reporting of rating policy information obtained through both open-ended and close-ended questions in the rating strategy questionnaire.

Quality of Information

The means and standard deviations of the manipulation check items by experimental condition are presented in Table 2. Univariate analysis of variance tests were conducted to test for differences in perceptions of source credibility,

Table 2. Manipulation Check: Ratings on Credibility, Reliability, Quality by Condition by Dimension

Dimension	Credibility		Reliability		Quality	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Dim 1						
Job Knowledge						
Cond 1	3.68	.82	3.47	.96	3.84	1.02
Cond 2	3.85	.58	3.70	.80	3.90	.87
Cond 3	3.37	1.01	3.26	1.04	3.63	.95
Cond 4	3.65	.81	3.45	1.05	4.05	.94
Cond 5	3.84	.95	3.74	.87	3.90	.93
Overall	3.68	.84	3.53	.94	3.87	.90
Dim 2						
Initiative						
Cond 1	3.26	.56	3.21	.71	3.47	1.02
Cond 2	3.65	.67	3.35	.81	3.65	.87
Cond 3	2.74	.87	2.95	.62	3.00	.74
Cond 4	3.20	.76	3.20	1.01	3.55	.94
Cond 5	3.32	1.00	2.95	.97	3.68	1.05
Overall	3.24	.82	3.13	.83	3.47	.94

(table continues)

Dimension	Credibility		Reliability		Quality	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Dim 3						
Attitude						
Cond 1	3.42	.83	3.16	.76	3.68	.94
Cond 2	3.75	.63	3.45	1.05	3.80	.89
Cond 3	3.16	1.07	3.37	1.07	3.58	1.12
Cond 4	3.25	.96	3.15	1.04	3.70	.92
Cond 5	3.47	1.02	3.36	1.06	3.95	.91
Overall	3.41	.92	3.30	.99	3.74	.95
Dim 4						
Arrests						
Cond 1	4.63	.68	4.47	.90	3.79	.31
Cond 2	3.60	1.57	3.45	1.50	2.45	1.40
Cond 3	3.90	1.29	3.84	1.12	2.90	1.45
Cond 4	3.70	1.38	3.70	1.42	2.65	1.31
Cond 5	4.32	.82	4.21	1.03	3.68	1.29
Overall	4.02	1.24	3.93	1.25	3.08	1.44

(table continues)

Dimension	Credibility		Reliability		Quality	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Dim 5						
Absences						
Cond 1	4.79	.53	4.42	.96	3.95	1.18
Cond 2	3.90	1.12	3.84	1.17	3.50	1.10
Cond 3	4.21	1.03	3.79	1.13	3.42	1.39
Cond 4	3.85	1.14	3.60	1.23	3.20	1.01
Cond 5	4.26	.87	4.26	.87	3.21	1.27
Overall	4.20	1.02	3.98	1.11	3.45	1.20
Dim 6						
Complaints						
Cond 1	4.00	1.20	3.79	1.18	3.74	1.15
Cond 2	3.20	1.32	3.25	1.25	3.00	.97
Cond 3	3.95	.97	3.53	1.02	3.00	1.29
Cond 4	3.35	1.42	3.45	1.36	3.05	1.47
Cond 5	3.74	.99	3.47	1.22	2.84	1.12
Overall	3.64	1.22	3.50	1.20	3.12	1.23

Note. Scale values ranged from 1 to 5.

reliability, and quality of the information by condition and by performance dimension. The results of the Anovas, presented in Table 3, indicated no significant differences in perceptions across conditions. Although the stimulus materials were pre-tested for the saliency of the manipulation, subjects participating in the study did not perceive any differences among the performance dimensions in terms of information quality. It would not be meaningful, therefore, to use the experimental groups to test hypotheses regarding information quality.

An examination of the means and standard deviations in Table 2 do, however, indicate variability in individual's perceptions of the information quality condition. It is possible that these individual variations are related to the weights used. To test this notion, a correlation analysis was performed to examine the relationship between the relative weights obtained from the policy capturing analysis and the manipulation check ratings of information quality. Table 4 shows the results of this analysis. Correlations between the relative weights and the information quality ratings were significant for all six performance dimensions. This indicates that although the experimental conditions did not have any impact on rater integration behavior, subjects were somehow incorporating perceptions of information quality in their ratings of police officer performance.

Table 3. Analysis of Variance Tests of the Manipulation Check Items by Condition

Dimension	Ratings of:	Degrees of Freedom	F Value	Significance Level
Job Knowledge	Credibility	4, 92	0.470	.758
	Reliability	4, 92	0.364	.834
	Quality	4, 92	0.632	.641
Initiative	Credibility	4, 92	1.984	.104
	Reliability	4, 92	1.500	.209
	Quality	4, 92	0.890	.473
Attitude	Credibility	4, 92	0.616	.652
	Reliability	4, 92	1.109	.357
	Quality	4, 92	0.262	.902
Arrests	Credibility	4, 92	2.039	.095
	Reliability	4, 92	1.299	.276
	Quality	4, 92	1.183	.324
Absences	Credibility	4, 91	0.660	.622
	Reliability	4, 91	0.584	.675
	Quality	4, 91	1.160	.334
Community Complaints/ Grievances	Credibility	4, 92	0.506	.732
	Reliability	4, 92	0.273	.894
	Quality	4, 92	0.508	.730

Table 4. Relationship Between Ratings of Information Quality^a
and Relative Weights^b by Performance Dimension

Performance Dimension	Information Quality	
	r	p
Job Knowledge	.170	.048
Initiative	.212	.018
Attitude	.372	.001
Arrests	.289	.002
Absences	.299	.001
Community Complaints	.296	.002

^aRatings of information quality obtained from the manipulation check items.

^bRelative weights obtained from the policy-capturing analysis.

Type of Data

The second issue investigated was an examination of the effect of type of data on raters' integration strategies. This involved a comparison of the subjects' weighting schemes for the two types of data, the supervisor ratings and the personnel data. The results of the policy capturing analysis (see Appendix B) indicated that subjects gave much more weight in making their ratings to the supervisor rating dimensions than to the personnel data dimensions. Table 5 presents the means of the relative weights obtained by performance dimension. The three supervisor rating dimensions of job knowledge, initiative, and attitude accounted for 68.4% of the variance in the subjects' decision-making. To examine the subjects' use of the two types of data further, the relative weights for the supervisor ratings and the personnel data were separately summed to form an index of the subjects' weighting schemes by type of data. The results (see Appendix B) indicated that the sum of the relative weights for the supervisor rating dimensions was greater than the sum of the relative weights for the personnel data for 85 of the 100 subjects in the study. A paired t-test was performed to test for differences between the sum of the relative weights for the personnel data dimensions and the sum of the relative weights for the supervisor rating dimensions. The results, presented in

Table 5. Means of the Relative Weights by Dimension

Job Performance Dimension	Mean	Standard Deviation
Job Knowledge	29.79	21.84
Initiative	17.85	10.88
Attitude	20.73	13.89
Number of Arrests	14.67	16.04
Number of Absences	8.81	11.11
Number of Community Complaints/Grievances	8.10	10.92

Table 6, show that the two sets of weights are significantly different from one another ($t(99) = 8.03, p < .001$).

A similar analysis was conducted for the subjective weights obtained (See Appendix B). The mean subjective weights by dimension are presented in Table 7. The results are consistent with those obtained for the relative weights in that the three supervisor rating dimensions accounted for 60.7% of the variance in the subjects' ratings. Table 8 presents a paired T-test between the sum of the subjective weights for the supervisor rating dimensions and the personnel data dimensions. The results indicated that there were significant differences between the two ($t(99) = 7.68, p < .001$).

Subjects were also asked questions in the rating strategy questionnaire regarding their use of the personnel data and supervisor ratings, as well as their perceptions of the accuracy and usefulness of the two types of data. Participants were asked to report how they reconciled differences which were found in the experimental profiles between the level of performance obtained from the supervisor ratings with those from the personnel data. Twenty-nine percent of the subjects said that they used both dimension sets equally, 44% reported that they used both, but weighted the supervisor ratings more heavily, 24% said that they used both, but weighted the personnel data more

Table 6. Paired T-test Between the Sum of the Relative Weights by Type of Data

	Mean	SD	T-value	Significance
Sum of the Supervisor Rating Relative Weights	68.36	23.01	8.03	.000
Sum of the Personnel Data Relative Weights	31.58	22.82		

Table 7. Means of the Subjective Weights by Dimension

Job Performance Dimension	Mean	Standard Deviation
Job Knowledge	22.79	9.27
Initiative	17.28	6.79
Attitude	20.65	8.33
Number of Arrests	12.51	7.54
Number of Absences	13.90	7.89
Number of Community Complaints/Grievances	12.95	8.23

Table 8. Paired T-test Between the Sum of the Relative Weights by Type of Data

	Mean	SD	T-value	Significance
Sum of the Supervisor Rating Subjective Weights	60.720	13.87	7.68	.000
Sum of the Personnel Data Subjective Weights	39.360	14.09		

heavily, and 1% reported that they used the supervisor ratings only.

Subjects' were also asked to report on their perceptions concerning the accuracy of information obtained from personnel data or supervisor ratings, as well as which type of data would be better for making comparisons across individuals and for assigning ratings to the ratees. Thirty-one percent felt that the supervisor ratings would be more accurate, 31% felt that the personnel data would be more accurate, and 38% felt that supervisor ratings and the personnel data would be equally likely to be accurate. As to which type of data would be better for making comparisons across individuals, 29% responded that the supervisor ratings would be better, 26% responded that the personnel data would be better, and 45% felt that both would be equally good. When asked which type of information they would prefer to use when making a set of ratings similar to those they made earlier, 28% preferred supervisor ratings, 16% preferred personnel data, and 56% said that they would use the two equally.

Consistency

The third issue examined in this study concerned the consistency with which raters made their ratings. The multiple R² obtained from the policy capturing analysis was

used as evidence for rater consistency. The mean R^2 value for the sample was 0.761, with a range of values from 0.303 to 0.954. Table 9 presents a frequency distribution of R^2 for the 100 raters. Most raters were consistent in using their rating strategy across the 40 profiles. The few raters whose R^2 were lower probably either did not correctly understand the task to be performed, or merely responded in a capricious fashion.

Subjects answered a question in the rating strategy questionnaire concerning how consistent they felt they were in making their ratings. Six percent felt that they were highly inconsistent, 11% felt that they were somewhat inconsistent, 70% felt that they were consistent with most profiles, while 13% reported that they were highly consistent in making their ratings. This self-report information on consistency relates favorably with the statistical findings concerning the R^2 obtained. In order to examine this relationship more closely, mean R^2 values were computed for the self-report of consistency response categories. Table 10 presents the means and standard deviations for the R^2 by response category. In general, higher R^2 values are associated with self-report responses of higher consistency. An analysis of variance was run to test for differences between the groups, but the result, while in the predicted direction, was not significant ($F_{3,96}$

Table 9. Frequency Distribution of the R^2 Values Obtained from the Policy-Capturing Analysis

Range	Frequency
0.00 - .09	0
.10 - .19	0
.20 - .29	0
.30 - .39	3
.40 - .49	3
.50 - .59	7
.60 - .69	12
.70 - .79	26
.80 - .89	38
.90 - .99	11

Table 10. Relationship Between R^2 Values and Self-report Measure of Consistency

Self-rating	R^2	
	Mean	SD
Highly Inconsistent	.843	.076
Somewhat Inconsistent	.700	.142
Consistent with Most Profiles	.754	.145
Highly Consistent	.816	.097

= 2.22, $p=.091$).

Statistical Weights vs. Self-report/Subjective Weights

The fourth issue investigated was the relation between the statistical information obtained through the policy capturing analysis and the self-report information obtained through the rating strategy questionnaire. The major focus was on the comparison of weighting schemes obtained through the regression analysis with subjective weights provided by the subjects. The relation between the two sets of weights was assessed through several different methods. First, individual-level Spearman rank order correlations were computed between the two sets of weights. Because the two sets of weights were on a common metric (where both the relative and subjective weights summed to 100), the correlations could be computed directly (Zedeck & Kafry, 1977). Table 11 presents the frequency distribution for the Spearman rank order correlations obtained. The mean correlation was .489, with a range of values from -0.828 to 1.0 (see Appendix B for individual values). As can be seen, there was relatively good agreement between the two sets of weights, indicating that most subjects had a were aware of the policies that they used when making their judgments. Fifty-two of the 100 subjects had correlations of .50 or better, and only 15 of the 100 subjects had correlations

Table 11. Frequency Distribution of the Spearman Rank-order
Correlation Between the Relative and Subjective Weights

Range	Frequency
-1.0 - -.80	1
-.79 - -.60	2
-.59 - -.40	1
-.39 - -.20	4
-.19 - .00	7
.01 - .19	2
.20 - .39	15
.40 - .59	15
.60 - .79	24
.80 - 1.0	29

that were negative in sign.

The relation between the statistical and subjective weights was also analyzed by computing t-tests between mean differences in the weightings for individual performance dimensions using the relative and subjective weights. Table 12 presents the results of this analysis. There were significant differences between the statistical and subjective weights for the dimensions of job knowledge ($t(99) = 3.89, p < .05$), absences ($t(99) = -5.77, p < .05$), and community complaints/grievances ($t(99) = -4.92, p < .05$). Subjects underestimated their use of job knowledge in making their ratings (the mean subjective weight for this dimension was less than the mean relative weight), while they overestimated the amount of weight given to the dimensions of absences and community complaints. The statistical and subjective weights for the dimensions of initiative ($t(99) = 0.57, p > .05$), attitude ($t(99) = 0.06, p > .05$), and arrests ($t(99) = 1.47, p > .05$) were not significantly different from one another. Overall, the results suggest that although the pattern or shape of the distributions of the statistical and subjective weights were similar (as evidenced by the Spearman rank-order correlations), the magnitude of the dimension weights were somewhat dissimilar.

Table 12. T-test Between the Relative and Subjective Weights by Performance Dimension

Performance Dimension	Relative Weight Mean	Relative Weight SD	Subjective Weight Mean	Subjective Weight SD	T Value	Significance Level
Job Knowledge	29.8	21.8	22.8	9.3	3.89	.000
Initiative	17.9	10.9	17.3	6.8	0.57	.570
Attitude	20.7	13.9	20.7	8.3	0.06	.954
Arrests	14.7	16.0	12.5	7.5	1.47	.144
Absences	8.8	11.1	13.9	7.9	-5.77	.000
Complaints	8.1	10.9	13.0	8.2	-4.92	.000

Additional Analyses

Another source of information concerning the subjects' rating policies was information obtained from the open-ended questions. One question concerned whether or not subjects ignored or disregarded information from one or more of the performance dimensions presented to them. It was found that 32% of the subjects reported that they disregarded or eliminated some of the dimensions from consideration in making their assignment of ratings. Four percent reported that they eliminated job knowledge, 7% eliminated initiative, 3% eliminated attitude, 24% eliminated arrests, 10% eliminated absences, and 17% eliminated community complaints/grievances. It is clear that subjects disregarded the personnel data dimensions more often than they disregarded the supervisor rating dimensions. Subjects, however, underestimated the number of cues which did not receive any weight in the policy-capturing analysis. Thirty-one of the 100 subjects had relative weights of 0 for one or more of the dimensions, while only 2 of the 100 subjects gave weightings of 0 to dimensions in their subjective weights. This is inconsistent with the self-report information obtained from the open-ended questions presented above. It could be that subjects felt that they had to assign at least some weight to each dimension in making the subjective weightings of the

dimensions. This points to a possible methodological flaw in the procedure for obtaining the subjective weights.

Specific rating strategies mentioned in the open-ended questionnaire were also examined. It was found that 15% of the subjects reported that they simply averaged across all the dimensions to formulate their overall ratings, while 25% reported that they used an averaging strategy for only a subset of the dimensions. The use of a conjunctive strategy, which was defined as the subjects reporting that they looked at a couple of the dimensions in detail first, and then proceeded to check others to see if they met some minimum criterion (either the values were very high or low) was also examined. Fifty-three percent of the subjects reported that they used such a strategy. This is particularly interesting in light of the high R^2 values obtained in this study, because R^2 is considered to be a measure of the linear use of cues.

Finally, subjects were asked about changes in their rating policies that took place over the course of the experiment. Forty-five percent of the subjects reported that their rating policies change over time. Forty-nine percent felt that the task became easier over time, and 14% felt that the task became more difficult. Of those who felt that the task became easier over time, 74% felt that the task became easier because they developed a better notion of

their rating policy.

Discussion

This study investigated two factors which might impact on how raters integrate information from various sources in making overall judgments of performance. These factors were quality of information and type of information. Policy capturing analyses and self-report questionnaires of rating policy were employed to explore the effects of these variables on raters' integration strategies.

Major Results

This study attempted to manipulate perceptions of the quality of information presented to raters for the purpose of assigning overall ratings of performance to 40 hypothetical profiles of police officer performance. The manipulation failed as subjects across the five experimental conditions did not perceive any differences in the quality of information of the performance dimensions. There are a number of possible reasons for the lack of experimental effects on this variable. First, the manipulation may not have been salient enough for the subjects. Because the information concerning the quality of information variable was presented before the subjects made any ratings, it is possible that the subjects simply neglected this information by the time they actually made their ratings and completed

the rating strategy questionnaire. The manipulation check items were the last items on the rating strategy questionnaire, while the manipulation concerning the information quality was the first information presented to subjects. The lapse of time, as well as the information presented not being emphasized strongly enough could have contributed to the lack of experimental effects. Secondly, the fact that subjects were participating in the study simply for class credit or on a volunteer basis could have reduced their motivation for carefully attending to all aspects of the study.

Third, the lack of results found for the manipulation could also have been a valid response to the stimuli used in this study. Subjects may not use information quality information as cues in situations where they have at least some information concerning the nature of the job under study. In those situations, individuals may rely on pre-formulated "schema" based on information gathered from previous life experiences concerning the stimulus job. The correlational analysis provides some evidence for this notion in that it was found that people seemed to have their own perceptions regarding information quality which were unrelated to the experimental conditions. Performance dimensions which were given higher ratings of information quality received higher relative weights in the policy

capturing analysis. These pre-conceived notions of the information quality of particular dimensions may have been so strong that they overrode the experimental manipulation. In any case, perceptions of information quality were related to the weights used.

An alternative explanation for the correlational results might be that the subjects were trying to be self-consistent with respect to the ratings made. The ratings of information quality, which were gathered after completion of the rating task, might have been merely reflecting an awareness of the dimensions used in assigning the overall ratings. Since causality can not be addressed with a correlational analysis, it is not known whether subjects' perceptions of information quality influenced the ratings made or whether awareness of rating policies affected the information quality ratings. Future research efforts designed at analyzing raters' integration strategies should first investigate raters' implicit evaluation schema regarding the importance of dimensions on various jobs before examining their effects on rating behavior.

The second major variable of interest in this study was the effect of type of information on raters' integration strategies. This study examined how raters combined information obtained from subjective data (supervisor rating dimensions) and objective data (personnel data dimensions).

The results indicated that subjects had a strong bias towards using the information from the supervisor rating dimensions over the information from the personnel data dimensions. Results from all three methods, the policy capturing analysis, analysis of the subjective weights, and analysis of the rating strategy questionnaire yielded the consistent finding that subjects placed greater weight on the supervisor rating dimensions than they did on the personnel data dimensions when assigning ratings to the police officer profiles. This finding fits nicely with the findings of other decision-making research studies (e.g. Tversky and Kahneman, 1974) in that subjects did use subjective data to a greater extent than objective data in making their ratings.

The bias towards subjective data raises several important issues. The first issue concerns whether this finding is pervasive phenomenon among individuals or whether this finding is simply a function of the job selected and the performance dimensions used in this study. If a bias towards using subjective data is generalizable, raters in actual industrial rating situations may tend to ignore or discount the personnel data information that is given to them, and rely more on intuitive judgment or "soft" criterion in making their ratings. It would be interesting to investigate possible reasons for such a bias (if it

indeed exists) in a field setting with organizational decision-makers. One possible reason for such a bias might be that raters have pre-determined notions that objective data is influenced to a greater extent by non-performance related factors, such as criterion unreliability or contamination than are ratings made by others.

An alternative explanation is that the nature of the position being rated influences which type of information is more important or relevant for rating purposes. For the job of police officer, the supervisor ratings may provide more "important" information concerning police officer performance, while for another job, such as a machine worker, objective criterion would be more useful to raters in helping to determine overall performance ratings. Thus, characteristics of jobs might interact with the source of information to influence how raters integrate performance information. Identification of those characteristics of jobs which are important in this process might help elucidate this phenomenon.

Tacit support for the perspective that the source of the information interacts with the type of job being rated could be found in the responses obtained from the self-report questionnaire. When subjects were asked which type of information in general they felt was more accurate, and more useful for making comparisons across individuals as well as

assigning overall ratings, the bias towards the supervisor ratings over the personnel data disappeared. This would suggest that when raters were not considering job-specific situations, supervisor ratings and personnel data are considered to be of equal value or use in the assignment of ratings. However, in rating performance of individuals on specific jobs, the bias towards the use of one type of information over the other may appear.

Additional Issues

Another issue examined in this study concerned the consistency with which raters rated the 40 profiles of police officer performance. The results indicated that the raters were fairly consistent in making their ratings. The high R^2 values found in this study compare favorably with the results found in other studies which have used numerical cues as stimulus values (Anderson, 1977; Hobson, Mendel, & Gibson, 1981) and are considerably higher than those which have used verbal descriptions of behaviors as cues (Zedeck & Cascio, 1982). The high R^2 values obtained indicate that raters were employing a consistent strategy in making the ratings across the profiles. Although many raters did indicate in the rating strategy questionnaire that their rating policies had changed over time, it seems that they may not have actually changed their policies, but that the

policies became more clear to them, and thus were able to use it consistently. Support for this notion was evidenced by the finding that 74% of the subjects who felt that the task became easier, felt that it did so because they developed a better notion of their policies.

It was also interesting that there did seem to be a trend for raters with higher R^2 values to mention in the self-report data that they were more consistent in making their ratings. Although the analysis was not significant, it does appear that raters do have at least some type of knowledge of the consistency with which they are applying their rating policies.

A second issue concerning the R^2 values obtained relates to subjects' perceptions of whether they used a linear or a non-linear strategy in weighting the cues to make their ratings. Dawes (1979) has pointed out that even when subjects are combining cues in a non-linear fashion, multiple regression techniques are so robust that this has little influence on the R^2 values obtained. From the self-report questionnaire it was found that 40% of the subjects reported that they used some sort of linear, additive weighting of the cues, whereas 53% of the subjects reported that they used conjunctive or non-linear strategies in making their ratings. Although over half of the subjects reported that they used a non-linear strategy in rating the

profiles, the R^2 values obtained were still very high. This suggests that policy-capturing methods may be lacking in the information that they provide concerning the more detailed aspects of raters' integration strategies. It is suggested that methods such as self-report questionnaires or verbal protocols be used to supplement the information obtained from policy capturing procedures to obtain this more detailed information.

The final issue investigated in this study relates to the relationship between the statistical information concerning the subject's rating policies and the self-report information of rating policies obtained from the generation of subjective weights and the questions in the rating strategy questionnaire. Results from the Spearman rank order correlation analysis yielded the finding that the pattern of weighting of the cues between the two sets of weights were fairly similar. Most subjects had a good notion of the ranking of the importance with which the performance dimensions were used in making their ratings. The relation between the statistical and subjective weights was also examined in terms of the magnitude of the weights placed on individual elements. Here subjects were less cognizant of the weights used. They underestimated their weighting of the job knowledge dimension, while they overestimated their weighting of the absences and community complaints

dimensions. It appears that subjects can estimate the rank ordering of the dimensions that they use in assigning ratings, but are less accurate in estimating the statistical weights used. For example, subject 12 had a value of 99 for the sum of the relative weights for the supervisor rating dimensions and a value of 1 for the sum of the relative weights for the personnel data dimensions, while he/she had a value of 50 for the sum of the subjective weights for the supervisor rating dimensions and a value of 40 for the sum of the subjective weights for the personnel data dimensions. Clearly this subject was not cognizant of the magnitude of the weights placed on the various dimensions in applying his/her rating policy. However, the Spearman rank order correlation between the relative and subjective weights for subject 12 was .515, indicating a fairly high degree of knowledge of the ranking of the performance dimensions.

The relation between the statistical and subjective weights was examined through analysis of the subjects' awareness of both the magnitude of weights placed on the performance dimensions as well as the pattern or ranking of the performance dimensions in terms of their importance to overall decisions made. Previous research efforts have either used the Spearman rank order correlations (e.g. Stumpf & London, 1981; Taylor & Wilsted, 1974) or t-tests between the statistical and subjective weights (e.g. Zedeck

& Kafry, 1977), but not both. Studies which have focused solely on the magnitude of weights placed on the dimensions have yielded the consistent finding that subjects underestimate the magnitude of weights placed on the major cues used, and overestimate the magnitude of weights placed on the minor cues (Slovic & Lichtenstein, 1971). The results of the present study are consistent with the previous research, in that subjects underestimated the weights placed on the supervisor rating dimensions and overestimated the weights placed on the personnel data in making their ratings. The conclusion typically drawn from studies with similar findings has been that raters have little insight into their rating policies.

The results from the analysis of subjects' awareness of the rank ordering of the weights used, however, provide evidence which would lead to the opposite conclusion. This study as well as the study by Stumpf and London (1981) found that subjects did have fairly accurate notions of the rank ordering of the cues used in assigning the ratings. It appears, then, that the two methods are providing different kinds of information concerning how cognizant the subjects are of the rating policies used. Future research should focus on what the practical and theoretical significance of these differences are, as well as the importance of this phenomenon to the rating process. One suggestion that

appears obvious is that the method of evaluating the relationship between statistical and subjective weights has an impact on the results obtained.

There is one issue that should be raised concerning the procedure for the collection of the subjective weights. In the present study, subjects were not explicitly instructed that they could assign a weight of zero to performance dimensions that they did not use in making their ratings. Only 2 out of the 100 subjects assigned a weight of zero to a particular dimension, even though many more subjects had statistical (relative) weights of zero for the policy capturing analysis. It is suggested that future research include explicit instructions to subjects that they can assign a weight of zero to dimensions which they eliminated or disregarded in making their ratings.

Limitations

There are some limitations to this study. First, as mentioned above, the manipulation of the information quality variable was not successful. Future research on this issue should make such a manipulation more salient to subjects. Possible suggestions to increase the saliency include having a group discussion regarding information quality issues, have subjects complete the manipulation check immediately following reading of the job performance dimension

definitions, and repackaging of the experimental materials so that the subjects can have the definitions of the dimensions more easily accessible to them when actually making the ratings.

A second issue concerns the chart included in the stimulus materials which translated the objective data in raw frequency count terms into values ranging from one to nine (see Appendix A). This table was intended to be merely a guide to the behavioral meaning of the objective data, but the meaning of the values in the chart may have been misinterpreted by the subjects in the study. It was noted that several subjects thought that the numbers inside the chart were the numbers which were written on the job performance profiles. For example, if a ratee received a rating of 8 for absences (which represents high performance), some subjects might have looked in the chart and saw that "8 absences" were low performance. This could have severely impacted on the results obtained in the study, as subjects would have misinterpreted the values of the performance profile dimensions and based their ratings on values that were not intended to be used in that fashion. The impact on the results would not be able to be detected unless subjects identified which values they had used. Future research should either eliminate the chart or make the meaning of the chart more clear to subjects.

definitions, and repackaging of the experimental materials so that the subjects can have the definitions of the dimensions more easily accessible to them when actually making the ratings.

A second issue concerns the chart included in the stimulus materials which translated the objective data in raw frequency count terms into values ranging from one to nine (see Appendix A). This table was intended to be merely a guide to the behavioral meaning of the objective data, but the meaning of the values in the chart may have been misinterpreted by the subjects in the study. It was noted that several subjects thought that the numbers inside the chart were the numbers which were written on the job performance profiles. For example, if a ratee received a rating of 8 for absences (which represents high performance), some subjects might have looked in the chart and saw that "8 absences" were low performance. This could have severely impacted on the results obtained in the study, as subjects would have misinterpreted the values of the performance profile dimensions and based their ratings on values that were not intended to be used in that fashion. The impact on the results would not be able to be detected unless subjects identified which values they had used. Future research should either eliminate the chart or make the meaning of the chart more clear to subjects.

A final limitation of the study concerns the generalizability of the findings. It should be noted that the results of the study may be limited by the sample used and the nature of the experimental task. Many of the undergraduate students who were participants in this study may not have had previous performance rating experience and thus the processes by which they made their ratings may be dissimilar to the processes by which managers in organizations make their ratings. In addition, the task, which required subjects to make ratings of police officer performance, may have been such that subjects were responding in a different way than persons who are more familiar with the job being rated would respond. Conducting the experiment in an artificial setting using hypothetical rather than actual performance profiles also reduces the generalizability of the findings. The effect of these limitations on the results of the study, however, are unknown. It is suggested that future research be conducted in a field setting to replicate the results found.

Future Directions for Research

This study yielded some important insight into the integration stage of the appraisal process. Future research on the rater behavior should continue this focus, investigating the cognitive processes by which raters make

their judgments concerning others. Specifically, more research needs to be undertaken on factors which might impact on how raters integrate information. The integration of information obtained from different sources and different types of data is an important issue which should be examined so that researchers can gain more knowledge of the processes and biases that raters might have when making a set of ratings. If researchers can identify critical factors which influence raters' cognitive processing of information, then this knowledge can then be applied towards the development of appraisal instruments and training programs which would increase raters' abilities to make more accurate judgments concerning others.

An important outcome of this study has been a demonstration of the ability of policy capturing techniques to test hypotheses concerning factors which influence rater integration strategies. It is suggested that future researchers using policy capturing techniques move away from studies which merely demonstrate the effectiveness of policy capturing procedures for obtaining information on performance-related decisions and move toward the use of such procedures for theory building and theory testing. Policy capturing clearly offers a unique opportunity for researchers to control the variables of interest and measure specified outcomes of the appraisal process.

Although policy capturing procedures are useful for testing the error of linear regression models, conducting tests of statistical significance, and determining the relative importance of cues, policy capturing is limited in the detail of information it can provide concerning raters' integration strategies (Einhorn, Kleinmuntz, & Kleinmuntz, 1979). It is suggested that another direction for future research be the use of verbal and/or written protocols as a supplement to the policy capturing procedures. Important information concerning raters' integration strategies can be obtained through such methods, as they provide much more rich, detailed information than that obtained from policy capturing procedures. Such information as the order in which subjects looked at cues, the number of cues looked at, and the ability to show that one can attend to cues and feel that one has used them, without such cues receiving significant weight in the regression equation can all be examined with a process tracing approach (Einhorn, Kleinmuntz, & Kleinmuntz, 1979). It is suggested that future research employ such data collection techniques as standard procedure.

Third, this study yielded some interesting findings regarding the effects of type of information on rater integration strategies. The finding that raters may have a bias towards using subjective data over objective data is an

intriguing one. Research should focus more specifically on this issue, investigating raters' use of different types of information when rating different kinds of jobs. Identification of pre-existing rater cognitive schema concerning the nature of performance and the meaning of performance dimensions is an area that has not been previously investigated and would help in understanding the rating process.

Finally, additional research needs to be undertaken regarding the impact of quality of information on rater integration strategies. A follow-up study in which the manipulation of information quality is made more salient to subjects would be useful in identification of its effects on rater integration of information.

Appendix A

Appendix A

INSTRUCTIONS

The following task requires that you make a set of ratings concerning the performance of a group of police officers. As you probably know, organizations usually require managers to make ratings of their subordinates' performance on a periodic basis. Your job for this task is to play the role of Police Captain and assign an overall rating of performance to each police officer in the packet given to you. Each officer that you will rate will have a job performance information sheet which contains information on six dimensions of police officer performance. These dimensions include job knowledge, initiative, attitude, number of arrests, number of community complaints/grievances, and number of absences. Information on the meaning of these performance dimensions is listed on the following page. The second packet that you received contains a ratings form on which you should mark your ratings. Please assign an overall rating from 1 to 9 for each of the 45 police officers in your packet.

Following the rating task, you are to answer the questions following the ratings form in packet 2. These questions pertain to how you went about making your ratings of the police officers. Please try to answer as best you can.

T

b

f

v

r

s

The actual values for the last three dimensions of performance have been translated from actual numerical counts to values which range from 1 to 9, with values of 1 to 3 representing LOW performance, values of 4 to 6 representing AVERAGE performance, and values of 7 to 9 representing HIGH performance. The chart below indicates how the values for the dimensions were translated.

	LOW PERFORMANCE			AVERAGE PERFORMANCE			HIGH PERFORMANCE		
Dimension	1	2	3	4	5	6	7	8	9
Arrests	0-7	8-14	15-21	22-28	29-35	36-42	43-49	50-56	57+
Absences	11+	9-10	8	6-7	5	3-4	2	1	0
Grievances	8+	7	6	5	4	3	2	1	0

Definitions of the Performance Dimensions - Condition 1

PLEASE READ THIS VERY CAREFULLY BEFORE MAKING ANY RATINGS

Definitions of Police Officer Performance Dimensions

Supervisor Performance Ratings

The first three dimensions of officer performance were obtained from ratings made by the lieutenant (supervisor) of each of the officers. The Lieutenant who made the ratings has been on the force for 10 years and has supervised this group of officers for the past 5 years. The ratings form itself has been found to be a highly accurate measure of police officer performance, due to the fact that all the lieutenants have undergone training in the use of the form and in understanding the behaviors which represent each of the the performance dimensions. In addition, it has been found that the ratings made by one lieutenant are very similar to those made by another lieutenant, when both rate the same police officer.

Definitions of the Dimensions of Officer Performance obtained
From the Lieutenant's Ratings

1. Job Knowledge:
 - Awareness of procedures, laws, and court rulings and changes in them.
2. Initiative:
 - Individual personal performance conducted without either direct supervision or commands, including suggestions for improved departmental procedures.
3. Attitude:
 - General orientation towards the law enforcement profession and the department.

Definitions of the Dimensions of Officer Performance obtained
from Personnel Data and Record Keeping Procedures

4. Number of Arrests:

-This figure represents the number of arrests made by the officer during the previous six months. This figure has been standardized across officers, such that biasing factors which would artificially inflate/deflate this figure, such as location (i.e. suburbs vs. the inner city) and assignment (i.e. foot patrol vs. squad car) have been taken into account. This standardization procedure allows for comparisons across officers to be made. This measure is considered by the officers to be a good measure of an officer's productivity.

5. Number of Community Complaints/Grievances:

-This figure represents the number of substantiated community complaints to the precinct concerning the police officer. As a result of community interest in policing, a special hot-line was set up by the precinct to receive community complaints. All calls received are investigated by a special task force of detectives specifically assigned to monitor community complaints and grievances. Only reports which resulted in an official reprimand or suspension of an officer from active duty were included in this figure.

6. Number of Absences:

-This figure represents the number of non-medical or non-sickness related absences of an officer during the past 6 months. This information was obtained from personnel department records. The personnel department requires a certified statement from the precinct doctor verifying that the officer has been examined, in order for an officer to be officially excused. In addition, the personnel department has instituted a verification policy, in which the officer's home is called upon his/her absence at roll call, to make sure the officer is really sick.

Definitions of the Performance Dimensions - Condition 2

PLEASE READ THIS VERY CAREFULLY BEFORE MAKING ANY RATINGS

Definitions of Police Officer Performance Dimensions

Supervisor Performance Ratings

The first three dimensions of officer performance were obtained from ratings made by the lieutenant (supervisor) of each of the officers. The Lieutenant who made the ratings has been on the force for 10 years and has supervised this group of officers for the past 5 years. The ratings form itself has been found to be a highly accurate measure of police officer performance, due to the fact that all the lieutenants have undergone training in the use of the form and in understanding the behaviors which represent each of the the performance dimensions. In addition, it has been found that the ratings made by one lieutenant are very similar to those made by another lieutenant, when both rate the same police officer.

Definitions of the Dimensions of Officer Performance obtained
From the Lieutenant's Ratings

1. Job Knowledge:
 - Awareness of procedures, laws, and court rulings and changes in them.
2. Initiative:
 - Individual personal performance conducted without either direct supervision or commands, including suggestions for improved departmental procedures.
3. Attitude:
 - General orientation towards the law enforcement profession and the department.

Definitions of the Dimensions of Officer Performance obtained
from Personnel Data and Record Keeping Procedures

4. Number of Arrests:

-This figure represents the number of arrests made by the officer during the previous six months. The number of arrests made is subject to fluctuations due to location of assignment (i.e. suburbs vs. the inner city) and type of assignment (i.e. foot patrol vs. squad car). Because of these variations, which are not taken into account in this measure, comparisons across officers are hard to make. In addition, because this measure is used as the basis for promotion decisions, some officers "pad" their arrest figures by making arrests for minor violations, many of which are subsequently thrown out of court. Due to these problems, police officers do not consider the number of arrests to be a good measure of an officer's productivity.

5. Number of Community Complaints/Grievances:

-This figure represents the number of community complaints to the precinct concerning the police officer. This information was obtained through an informal log of calls to the precinct from citizens in the community. The level of performance on this dimension has been known to vary widely depending on the amount of community interest in policing. In addition, there are often a few individuals in the community who frequently call the precinct to make complaints, many of which are unjustified.

6. Number of Absences:

-This figure represents the number of non-medical or non-sickness related absences of an officer during the past six months. This information was obtained from a self-report diary of absences filled out by the officer. There has been a problem in the past with officers exaggerating the number of illness related absences, as some officers have been using sick days as extra vacation days.

Definitions of the Dimensions of Officer Performance obtained
from Personnel Data and Record Keeping Procedures

4. Number of Arrests:

-This figure represents the number of arrests made by the officer during the previous six months. The number of arrests made is subject to fluctuations due to location of assignment (i.e. suburbs vs. the inner city) and type of assignment (i.e. foot patrol vs. squad car). Because of these variations, which are not taken into account in this measure, comparisons across officers are hard to make. In addition, because this measure is used as the basis for promotion decisions, some officers "pad" their arrest figures by making arrests for minor violations, many of which are subsequently thrown out of court. Due to these problems, police officers do not consider the number of arrests to be a good measure of an officer's productivity.

5. Number of Community Complaints/Grievances:

-This figure represents the number of community complaints to the precinct concerning the police officer. This information was obtained through an informal log of calls to the precinct from citizens in the community. The level of performance on this dimension has been known to vary widely depending on the amount of community interest in policing. In addition, there are often a few individuals in the community who frequently call the precinct to make complaints, many of which are unjustified.

6. Number of Absences:

-This figure represents the number of non-medical or non-sickness related absences of an officer during the past six months. This information was obtained from a self-report diary of absences filled out by the officer. There has been a problem in the past with officers exaggerating the number of illness related absences, as some officers have been using sick days as extra vacation days.

Definitions of the Performance Dimensions - Condition 3

Definitions of the Performance Dimensions - Condition 3

PLEASE READ THIS VERY CAREFULLY BEFORE MAKING ANY RATINGS

Definitions of Police Officer Performance Dimensions

Supervisor Performance Ratings

The first three dimensions of officer performance were obtained from ratings made by the Lieutenant (supervisor) of each of the officers. The lieutenant who made the ratings transferred into the precinct two weeks ago, and thus may not have a good knowledge of the officers' performance. In addition, the ratings form itself is known to be subject to the biases of the person who is making the ratings (i.e. If the rater likes the ratee, he/she is probably going to assign that person higher ratings, regardless of the officer's performance). Finally, it has been found that two lieutenants rating the same officer often assign very different ratings to that officer.

**Definitions of the Dimensions of Officer Performance obtained
From the Lieutenant's Ratings**

1. **Job Knowledge:**
 - Awareness of procedures, laws, and court rulings and changes in them.
2. **Initiative:**
 - Individual personal performance conducted without either direct supervision or commands, including suggestions for improved departmental procedures.
3. **Attitude:**
 - General orientation towards the law enforcement profession and the department.

Definitions of the Dimensions of Officer Performance obtained
from Personnel Data and Record Keeping Procedures

4. Number of Arrests:

-This figure represents the number of arrests made by the officer during the previous six months. This figure has been standardized across officers, such that biasing factors which would artificially inflate/deflate this figure, such as location (i.e. suburbs vs. the inner city) and assignment (i.e. foot patrol vs. squad car) have been taken into account. This standardization procedure allows for comparisons across officers to be made. This measure is considered by the officers to be a good measure of an officer's productivity.

5. Number of Community Complaints/Grievances:

-This figure represents the number of substantiated community complaints to the precinct concerning the police officer. As a result of community interest in policing, a special hot-line was set up by the precinct to receive community complaints. All calls received are investigated by a special task force of detectives specifically assigned to monitor community complaints and grievances. Only reports which resulted in an official reprimand or suspension of an officer from active duty were included in this figure.

6. Number of Absences:

-This figure represents the number of non-medical or non-sickness related absences of an officer during the past 6 months. This information was obtained from personnel department records. The personnel department requires a certified statement from the precinct doctor verifying that the officer has been examined, in order for an officer to be officially excused. In addition, the personnel department has instituted a verification policy, in which the officer's home is called upon his/her absence at roll call, to make sure the officer is really sick.

Definitions of the Performance Dimensions - Condition 4

PLEASE READ THIS VERY CAREFULLY BEFORE MAKING ANY RATINGS

Definitions of Police Officer Performance Dimensions

Supervisor Performance Ratings

The first three dimensions of officer performance were obtained from ratings made by the Lieutenant (supervisor) of each of the officers. The lieutenant who made the ratings transferred into the precinct two weeks ago, and thus may not have a good knowledge of the officers' performance. In addition, the ratings form itself is known to be subject to the biases of the person who is making the ratings (i.e. If the rater likes the ratee, he/she is probably going to assign that person higher ratings, regardless of the officer's performance). Finally, it has been found that two lieutenants rating the same officer often assign very different ratings to that officer.

**Definitions of the Dimensions of Officer Performance obtained
From the Lieutenant's Ratings**

1. **Job Knowledge:**
 - Awareness of procedures, laws, and court rulings and changes in them.
2. **Initiative:**
 - Individual personal performance conducted without either direct supervision or commands, including suggestions for improved departmental procedures.
3. **Attitude:**
 - General orientation towards the law enforcement profession and the department.

Definitions of the Dimensions of Officer Performance obtained
from Personnel Data and Record Keeping Procedures

4. Number of Arrests:

-This figure represents the number of arrests made by the officer during the previous six months. The number of arrests made is subject to fluctuations due to location of assignment (i.e. suburbs vs. the inner city) and type of assignment (i.e. foot patrol vs. squad car). Because of these variations, which are not taken into account in this measure, comparisons across officers are hard to make. In addition, because this measure is used as the basis for promotion decisions, some officers "pad" their arrest figures by making arrests for minor violations, many of which are subsequently thrown out of court. Due to these problems, police officers do not consider the number of arrests to be a good measure of an officer's productivity.

5. Number of Community Complaints/Grievances:

-This figure represents the number of community complaints to the precinct concerning the police officer. This information was obtained through an informal log of calls to the precinct from citizens in the community. The level of performance on this dimension has been known to vary widely depending on the amount of community interest in policing. In addition, there are often a few individuals in the community who frequently call the precinct to make complaints, many of which are unjustified.

6. Number of Absences:

-This figure represents the number of non-medical or non-sickness related absences of an officer during the past six months. This information was obtained from a self-report diary of absences filled out by the officer. There has been a problem in the past with officers exaggerating the number of illness related absences, as some officers have been using sick days as extra vacation days.

Instructions for Condition 5

INSTRUCTIONS

The following task requires that you make a set of ratings concerning the performance of a group of police officers. As you probably know, organizations usually require managers to make ratings of their subordinates' performance on a periodic basis. Your job for this task is to play the role of Police Captain and assign an overall rating of performance to each police officer in the packet given to you. Each officer that you will rate will have a job performance information sheet which contains information on six dimensions of police officer performance. These dimensions include job knowledge, initiative, attitude, number of arrests, number of community complaints/grievances, and number of absences. The second packet that you received contains a ratings form on which you should mark your ratings. Please assign an overall rating from 1 to 9 for each of the 45 police officers in your packet.

Following the rating task, you are to answer the questions following the ratings form in packet 2. These questions pertain to how you went about making your ratings of the police officers. Please try to answer as best you can.

RATING STRATEGY QUESTIONNAIRE

•

If so, describe the difference between the method used at the beginning vs. the end of the task. To what extent did the task become easier or more difficult over time? Why?

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

3. At what point in making the ratings did you develop the rating strategy that you used for this research project?
- a. I relied on a strategy I have used in other situations
 - b. I developed my strategy after reading the instructions and definitions of the performance dimensions
 - c. I developed my strategy after making ratings on the practice officers
 - d. I developed my strategy after making ratings on a few of the police officers
 - e. Other (Please Specify): _____

4. How consistent do you feel you were in using this strategy to make all of the 45 ratings?
- a. Highly inconsistent
 - b. Somewhat inconsistent
 - c. Consistent with most officers, there were some exceptions, however
 - d. Highly consistent across officers
5. If the supervisor's ratings and the personnel data had different levels of performance (i.e. 6,7,7 for the ratings and 2,3,3 for the personnel data) for an officer, what strategy did you use to reconcile differences between the performance levels obtained from the supervisor's ratings and the performance levels given for the personnel data?
- a. I used both the ratings and the personnel data equally
 - b. I used both the ratings and the personnel data, but I gave more weight to the supervisor's ratings
 - c. I used both the ratings and the personnel data, but I gave more weight to the personnel data
 - d. I relied on the supervisor's ratings only
 - e. I relied on the personnel data only
 - f. Other (Please specify): _____

6. Suppose you are given a set of performance data which contains scores on performance dimensions which were obtained from supervisor's ratings, as well as scores on performance dimensions which were obtained from personnel data. If you had no instructions which describe the performance dimensions, and no information concerning the source of the information or the quality of the information:

A. Which type of information would you be more likely to believe is accurate?

- a. the supervisor ratings
- b. the personnel data
- c. They would be equally likely to be accurate

B. Which type of information would be better for making comparisons across individuals (in terms of their performance levels)?

- a. the supervisor ratings
- b. the personnel data
- c. They both would be equally good for making comparisons

C. Which type of information would you be more likely to use in making a set of ratings similar to the ratings you made earlier?

- a. the supervisor ratings
- b. the personnel data
- c. I would probably use both equally

7. Consider the strategy you used to assign overall performance ratings to the 45 police officers.

Please distribute 100 points among the officer performance dimensions so that the distribution reflects the importance you placed on the dimensions in making your ratings.

Example: Suppose you felt that you used the dimensions of job knowledge and number of arrests to the greatest extent in making your ratings, followed by initiative, number of community complaints, attitude, and lastly, to a lesser extent, number of absences. Your ratings might be:

Job knowledge.....	30
Initiative.....	15
Attitude.....	10
Number of Arrests.....	30
Number of Absences.....	5
Number of Community Grievances/Complaints...	10
	100

Now, complete the form indicating the way you used the dimensions in making your ratings. Please remember that the sum of the weights you assign must add up to 100.

JOB KNOWLEDGE.....	_____
INITIATIVE.....	_____
ATTITUDE.....	_____
NUMBER OF ARRESTS.....	_____
NUMBER OF ABSENCES.....	_____
NUMBER OF COMMUNITY GRIEVANCES/COMPLAINTS..	_____

100

8. For each dimension listed below, please circle the number which represents how credible the source of the dimension was, how reliable the information from the dimension was, as well as the information quality of that information.

A. Source Credibility is defined as believability of the source of the information. This includes expertise of the source (whether the source is knowledgeable or not) and bias of the source (whether the source is providing accurate information regarding the police officer's performance).

B. Reliability is defined as a measure's consistency under different conditions. This includes such factors as whether two judges would give the same rating to the same individual, whether factors unrelated to job performance would influence an individual's score, and whether an individual's score is likely to be consistent over time (.i.e. be consistent from month to month.)

C. Quality of Information is the usefulness of a piece of information in describing an individual's job performance.

Please use the following scale to make your ratings:

1-----2-----3-----4-----5
 Low Moderate High

Performance Dimension	A. Credibility of the Information Source					B. Reliability of the Information					C. Quality of the Information				
Job Knowledge..	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Initiative.....	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Attitude.....	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Number of Arrests.....	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Number of Absences.....	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Number of Community Complaints.....	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

POLICE OFFICER PERFORMANCE EVALUATION FORM A16-11B/F1

OFFICER # _____

POLICE DIVISION 14
SPRING 1985

<u>JOB PERFORMANCE DIMENSION</u>	<u>RATING</u>
NUMBER OF ARRESTS.....	_____
NUMBER OF ABSENCES.....	_____
NUMBER OF COMMUNITY GRIEVANCES/COMPLAINTS..	_____
JOB KNOWLEDGE.....	_____
INITIATIVE.....	_____
ATTITUDE.....	_____

POLICE OFFICER PERFORMANCE EVALUATION FORM A16-11B/F2

OFFICER # _____

POLICE DIVISION 14
SPRING 1985

<u>JOB PERFORMANCE DIMENSION</u>	<u>RATING</u>
--	---------------

JOB KNOWLEDGE..... _____

INITIATIVE..... _____

ATTITUDE..... _____

NUMBER OF ARRESTS..... _____

NUMBER OF ABSENCES..... _____

NUMBER OF COMMUNITY
GRIEVANCES/COMPLAINTS.. _____

Practice Police Officer Ratings Form

Please use the following scale to rate the police officers:

1-----	2-----	3-----	4-----	5-----	6-----	7-----	8-----	9
Poor				Average				Outstanding
Performance				Performance				Performance

<u>Officer</u>	<u>Rating</u>
----------------	---------------

A..... _____

B..... _____

C..... _____

D..... _____

E..... _____

Police Officer Ratings Form

Please use the following scale to rate the police officers:

1-----2-----3-----4-----5-----6-----7-----8-----9
 Poor Average Outstanding
 Performance Performance Performance

Officer Overall
 Rating

1.....	_____
2.....	_____
3.....	_____
4.....	_____
5.....	_____
6.....	_____
7.....	_____
8.....	_____
9.....	_____
10.....	_____
11.....	_____
12.....	_____
13.....	_____
14.....	_____
15.....	_____
16.....	_____
17.....	_____
18.....	_____
19.....	_____
20.....	_____
21.....	_____
22.....	_____
23.....	_____
24.....	_____
25.....	_____
26.....	_____
27.....	_____
28.....	_____
29.....	_____
30.....	_____
31.....	_____
32.....	_____
33.....	_____
34.....	_____
35.....	_____
36.....	_____
37.....	_____
38.....	_____
39.....	_____
40.....	_____

Zedeck and Cascio (1982) Algorithm

Case	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9
1	2	2	2	2	2	2	2	2	2
2	2	1	3	2	1	3	2	2	1
3	1	2	3	1	2	2	3	2	1
4	1	1	1	1	1	1	1	1	1
5	3	2	1	2	3	1	2	3	2
6	2	3	1	3	2	2	1	2	3
7	3	3	3	3	3	3	3	3	3
8	1	3	2	2	1	2	3	3	1
9	3	1	2	1	2	3	1	2	2
10	2	2	2	3	3	1	1	1	3
11	1	2	1	3	1	3	2	2	2
12	2	1	3	2	3	2	3	1	2
13	3	2	2	1	1	2	3	3	3
14	2	3	1	1	3	3	2	1	1
15	1	1	2	3	2	1	2	3	1
16	3	3	3	2	2	1	1	2	1
17	2	2	3	1	2	1	3	2	3
18	1	3	2	2	3	3	1	2	2
19	2	1	2	1	3	1	3	2	2
20	3	2	1	3	2	3	3	1	1
21	2	3	1	1	1	2	2	3	2
22	3	1	1	2	2	2	3	1	3
23	2	2	2	3	1	2	1	3	2
24	1	2	3	2	3	2	2	2	3
25	1	1	3	3	1	1	2	1	3
26	3	3	2	2	1	1	3	2	2
27	2	2	2	2	2	3	1	3	1
28	1	3	2	1	2	2	2	1	3
29	3	1	2	3	3	2	2	2	1
30	2	1	1	2	2	3	2	3	3
31	1	2	1	2	3	1	1	3	2
32	2	3	3	3	2	1	2	1	2
33	3	2	3	1	1	3	1	1	1
34	2	2	2	2	1	3	3	2	3
35	1	1	1	1	3	2	1	2	1
36	3	3	3	3	3	2	1	3	2
37	2	1	3	2	1	2	3	3	1
38	1	2	2	3	2	3	3	2	2
39	3	2	1	2	2	1	2	2	3
40	2	3	2	1	3	2	2	1	2

Police Officer Profiles:
Dimension Scale Values

Case	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
1	5	5	6	6	5	5
2	5	3	7	4	2	7
3	1	4	8	2	5	5
4	2	1	1	2	3	2
5	8	5	3	6	7	3
6	5	7	2	9	6	6
7	8	8	9	8	7	7
8	2	7	6	5	1	5
9	9	1	5	3	6	8
10	5	5	5	7	8	2
11	1	6	1	8	2	7
12	4	3	7	6	9	6
13	9	5	6	2	1	4
14	6	9	1	2	8	7
15	2	1	5	9	5	2
16	8	7	9	6	5	1
17	6	5	9	1	5	3
18	2	8	7	5	7	7
19	6	1	6	1	9	2
20	9	4	2	8	4	7
21	4	8	2	3	3	6
22	9	1	3	5	5	5
23	4	4	4	8	1	6
24	2	5	8	6	7	6
25	1	2	8	7	2	3
26	8	8	7	6	2	1
27	4	5	6	5	5	9
28	3	8	5	1	4	4
29	8	2	6	8	9	5
30	6	2	3	5	5	9

(table continues)

Case	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
31	1	6	2	6	7	2
32	6	8	7	9	4	1
33	7	4	9	2	2	7
34	6	6	6	5	2	8
35	1	2	1	3	8	2
36	8	9	9	8	8	6
37	4	1	9	6	2	5
38	3	4	6	9	6	8
39	8	6	1	5	6	3
40	6	8	5	3	8	6

Police Officer Profiles:
Job Performance Dimension Intercorrelations

	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
Job Knowledge	1.0000	.0954	.1065	-.0019	.1011	.0663
Initiative	.0954	1.0000	.0557	.0745	.0330	.0415
Attitude	.1065	.0557	1.0000	.0348	-.0751	.0066
Arrests	-.0019	.0745	.0348	1.0000	.0406	.0418
Absences	.1011	.0330	-.0751	.0406	1.0000	-.0484
Complaints	.0663	.0415	.0066	.0418	-.0484	1.0000

Explanation and Consent

We do not anticipate any risks to you from the questions in this study. Your responses will be completely confidential and anonymous -- we are not interested in your identity, so please do not put your name on any of these forms. Participation in the study will require approximately one hour of your time.

Before you answer any of the following, be sure to read and understand the following:

1. Your participation is completely voluntary and you may withdraw from the study at any time.
2. The ratings are for research purposes only.
3. The research is being conducted by Mr. Michael P. Kirsch and his assistants in order to fulfill requirements for his Master's Degree.
4. Overall results of the investigation will be available to you at the conclusion of the study, if you so desire.

The return of the completed questionnaire indicates your understanding of the study and your consent to participate in the project. If you have read the above and agree to participate, please sign the form at the bottom of the page. Thank you very much.

I have read the above consent form and agree to participate in this study.

Signature

Date

FEEDBACK FOR POLICE OFFICER RATING EXPERIMENT

Almost every organization has some sort of method for evaluating the performance of its employees. The information obtained from these performance evaluations is used for making many organizational decisions, such as determining pay increases, and promotions. Because organizations are now required to document all personnel-related decisions, increased attention in the field of Industrial/Organizational psychology has been paid towards understanding how people make these performance evaluations. It is hoped that as a result of the information gained from this research that performance evaluations can become more accurate and unbiased, and that organizational members can be evaluated more fairly.

The study you just participated in examines a small part of the appraisal process: how people combine information from various sources. We are interested in how people will integrate information that comes from supervisor ratings vs. personnel data, and whether the quality of the information (i.e. the reliability and the source credibility) impacts on how you came up with your overall judgments of performance. The results from your overall ratings will allow us to determine what information you looked at during the task, the importance you placed on specific dimensions, and how you combined all the information to form an overall judgment. The researchers have hypothesized that the higher the quality of information of the performance dimension, the more weight subjects will place on that dimension when making their overall judgements.

Appendix B

Individual Rater Policies*

Case	Job Performance Dimensions					
	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
1	12.2	2.3	13.1	61.9	9.3	1.1
2	4.6	1.0	10.3	18.5	27.8	28.7
3	15.8	17.3	30.5	15.1	14.4	7.0
4	59.3	9.5	7.8	9.3	3.0	11.2
5	80.2	9.1	8.3	2.3	0.0	0.0
6	15.1	17.8	33.5	8.1	20.5	5.1
7	51.4	23.7	18.7	1.0	4.1	1.3
8	49.2	9.3	19.2	18.8	3.6	0.0
9	67.4	9.7	14.7	1.7	0.0	6.6
10	32.3	15.5	22.3	2.5	9.1	18.2
11	31.4	19.7	36.9	3.5	6.9	1.7
12	83.7	9.8	5.4	0.0	0.0	1.0
13	9.5	11.1	29.5	11.3	20.8	17.7
14	6.5	8.6	16.4	64.5	4.0	0.0
15	18.3	0.0	3.9	62.1	14.9	4.3
16	34.4	6.2	18.5	1.0	6.5	33.6
17	10.7	34.0	19.5	13.9	2.0	20.0
18	14.8	13.5	28.2	31.5	4.2	7.8
19	27.2	22.5	25.9	14.5	4.7	5.1
20	32.8	14.2	34.2	14.4	1.0	3.8
21	77.6	5.9	8.0	0.0	3.4	4.8
22	41.3	18.6	16.7	2.8	2.4	18.2
23	47.6	10.1	13.0	22.2	7.4	0.0
24	25.0	30.6	31.3	2.1	9.6	1.3
25	64.4	18.4	0.0	14.8	1.9	1.0
26	36.1	35.5	20.8	0.0	1.7	5.5
27	5.9	18.4	71.0	4.5	0.0	0.0
28	22.0	34.3	40.8	0.0	1.0	2.0
29	33.2	34.5	30.4	0.0	1.0	1.2
30	9.2	15.5	34.7	33.3	6.1	1.1

(table continues)

Case	Job Performance Dimensions					
	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
31	31.6	29.4	25.6	1.0	5.7	6.9
32	31.6	25.6	10.1	12.0	16.9	3.9
33	11.0	17.6	12.3	2.6	46.2	10.3
34	6.6	11.8	42.1	18.7	18.4	2.5
35	21.3	23.3	11.7	35.7	1.2	6.7
36	26.0	24.1	28.9	9.9	6.6	4.6
37	29.6	26.2	25.7	7.5	4.5	6.5
38	37.4	28.9	25.7	4.4	3.4	0.0
39	33.4	19.6	26.4	9.6	8.9	2.1
40	63.0	13.9	20.2	2.6	0.0	0.0
41	19.0	13.4	34.5	27.3	4.2	1.4
42	9.8	24.3	18.1	32.1	11.3	4.4
43	26.3	22.2	31.3	12.3	6.0	1.9
44	69.1	14.9	5.4	1.0	3.5	6.2
45	20.8	9.3	18.4	51.6	0.0	0.0
46	73.1	16.2	4.6	2.1	4.1	0.0
47	79.6	9.2	11.3	0.0	0.0	0.0
48	56.2	2.2	1.0	40.8	0.0	0.0
49	6.7	2.2	24.8	21.5	37.2	7.7
50	89.4	5.4	4.9	0.0	0.0	0.0
51	40.0	25.6	27.3	5.3	6.6	0.0
52	5.3	2.2	70.2	15.1	2.4	4.7
53	19.7	24.6	20.1	8.2	0.0	27.4
54	11.8	32.8	11.2	20.5	17.5	6.3
55	5.8	4.9	6.0	0.0	80.8	2.4
56	12.1	16.7	36.8	15.4	0.0	18.5
57	15.6	9.0	6.5	27.8	22.0	19.0
58	3.8	8.2	57.9	20.1	7.5	2.5
59	3.4	5.6	3.0	8.8	20.1	59.1
60	14.6	16.0	25.4	12.1	19.4	12.6
61	92.3	2.4	1.6	2.9	0.0	0.0

(table continues)

Case	Job Performance Dimensions					
	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
62	24.2	24.2	32.3	0.0	1.0	18.1
63	7.7	1.7	2.3	19.1	9.6	59.6
64	24.3	28.5	30.0	0.0	8.7	8.5
65	12.9	19.9	31.8	19.6	2.5	13.2
66	23.3	41.7	19.8	0.5	14.4	0.0
67	37.4	11.5	31.2	7.5	7.7	4.7
68	6.0	8.7	6.3	24.6	20.5	33.8
69	18.4	28.4	8.8	31.5	1.0	5.4
70	61.4	2.8	12.8	11.3	11.6	0.0
71	24.5	18.5	27.2	9.6	12.4	7.7
72	12.7	33.6	18.5	10.0	18.1	7.1
73	55.6	28.6	11.6	3.5	0.0	0.5
74	26.9	23.8	23.6	0.0	6.5	19.3
75	33.3	28.1	18.4	16.6	1.0	2.8
76	17.0	19.1	14.3	16.1	19.1	14.6
77	52.9	22.9	6.2	15.7	1.2	1.1
78	13.5	0.0	5.5	69.5	1.0	10.9
79	25.0	20.8	33.6	0.0	18.2	2.4
80	14.2	9.5	29.5	35.0	1.5	10.1
81	16.7	21.1	18.9	14.3	15.7	13.2
82	19.5	11.8	19.0	5.9	11.4	32.4
83	37.6	26.8	17.9	2.6	7.7	7.4
84	18.1	14.1	20.7	7.1	15.5	24.6
85	26.2	13.2	18.7	26.4	9.5	6.1
86	36.2	13.1	41.5	2.3	3.1	3.7
87	2.8	4.8	4.2	69.5	17.5	1.3
88	29.7	31.7	10.7	5.3	16.3	6.2
89	34.7	23.0	13.9	10.6	10.5	7.1
90	31.4	24.7	17.5	18.3	1.0	7.4
91	29.9	40.7	30.4	0.0	1.0	0.0

(table continues)

Case	Job Performance Dimensions					
	Job Knowledge	Initiative	Attitude	Arrests	Absences	Complaints
92	22.6	47.4	1.3	28.8	0.0	0.0
93	68.1	3.8	27.7	0.0	0.0	0.0
94	20.2	48.2	9.6	20.2	1.0	1.2
95	10.4	15.7	39.1	4.5	17.1	13.1
96	17.2	16.2	28.2	13.3	14.8	10.2
97	31.0	29.4	5.1	17.2	15.6	2.0
98	33.6	8.1	10.6	22.8	6.9	17.9
99	2.7	31.1	50.9	10.9	4.4	0.0
100	13.8	23.4	10.4	32.0	10.5	9.8

*Values in table are presented in terms of relative weights.

Individual Rater Multiple R^2 for
Policy Capturing Analysis

Case	R^2	Case	R^2
1	.547	31	.855
2	.576	32	.807
3	.858	33	.735
4	.904	34	.687
5	.883	35	.830
6	.587	36	.853
7	.879	37	.824
8	.732	38	.847
9	.893	39	.861
10	.914	40	.799
11	.843	41	.886
12	.454	42	.833
13	.464	43	.778
14	.837	44	.805
15	.552	45	.725
16	.823	46	.682
17	.603	47	.857
18	.845	48	.719
19	.880	49	.315
20	.778	50	.747
21	.873	51	.366
22	.945	52	.775
23	.756	53	.799
24	.940	54	.711
25	.755	55	.711
26	.922	56	.543
27	.836	57	.804
28	.869	58	.653
29	.794	59	.950
30	.830	60	.803

(table continues)

Case	R^2	Case	R^2
61	.884	81	.926
62	.640	82	.703
63	.676	83	.677
64	.856	84	.549
65	.505	85	.872
66	.623	86	.768
67	.743	87	.722
68	.871	88	.498
69	.753	89	.805
70	.845	90	.663
71	.864	91	.913
72	.682	92	.755
73	.724	93	.954
74	.800	94	.753
75	.637	95	.861
76	.873	96	.771
77	.717	97	.771
78	.303	98	.896
79	.947	99	.903
80	.692	100	.806

Sum of the Relative and Subjective
Weights by Type of Data

Case	<u>Sum of Relative Weights</u>		<u>Sum of Subjective Weights</u>	
	Supervisor Ratings	Personnel Data	Supervisor Ratings	Personnel Data
1	28	72	40	60
2	16	75	40	60
3	64	37	35	65
4	77	24	65	35
5	98	2	80	20
6	66	34	70	30
7	94	6	70	30
8	78	22	77	23
9	92	8	60	40
10	70	30	70	30
11	88	12	60	40
12	99	1	60	40
13	50	50	60	40
14	32	69	50	50
15	22	81	50	50
16	59	41	65	35
17	64	36	50	50
18	57	44	58	42
19	76	24	51	49
20	81	19	45	55
21	92	8	60	40
22	77	23	70	30
23	71	30	45	55
24	87	13	81	19
25	83	18	80	20
26	92	7	60	40
27	95	5	75	25
28	97	3	85	15

(table continues)

Case	<u>Sum of Relative Weights</u>		<u>Sum of Subjective Weights</u>	
	Supervisor Ratings	Personnel Data	Supervisor Ratings	Personnel Data
29	98	2	85	15
30	59	40	75	25
31	87	14	60	40
32	67	33	67	33
33	41	59	30	70
34	61	40	80	30
35	56	44	45	55
36	79	21	65	35
37	82	19	80	20
38	92	8	60	40
39	79	21	60	40
40	97	3	65	35
41	67	33	75	25
42	52	48	60	40
43	80	20	75	25
44	89	11	60	40
45	49	52	60	40
46	94	6	55	45
47	100	0	80	20
48	59	41	70	30
49	34	66	65	35
50	100	0	70	30
51	93	12	70	30
52	78	22	50	50
53	64	36	60	40
54	56	44	51	49
55	17	83	30	70
56	66	34	50	50
57	31	69	50	50
58	70	30	45	55
59	12	88	20	80

(table continues)

Case	<u>Sum of Relative Weights</u>		<u>Sum of Subjective Weights</u>	
	Supervisor Ratings	Personnel Data	Supervisor Ratings	Personnel Data
60	56	44	65	35
61	96	3	78	22
62	81	19	85	15
63	12	88	28	72
64	83	17	60	40
65	65	35	55	45
66	85	15	65	35
67	80	20	75	25
68	21	79	35	65
69	56	38	65	35
70	77	23	55	45
71	70	30	60	20
72	65	35	50	50
73	96	4	65	35
74	74	26	70	30
75	80	20	60	40
76	50	50	55	45
77	82	18	70	30
78	19	81	75	25
79	79	21	65	35
80	53	47	45	55
81	57	43	60	40
82	50	50	60	40
83	82	18	60	40
84	53	47	65	35
85	58	42	58	42
86	91	9	75	25
87	12	88	40	60
88	72	28	36	64
89	72	28	75	25
90	74	27	51	49

(table continues)

Case	<u>Sum of Relative Weights</u>		<u>Sum of Subjective Weights</u>	
	Supervisor Ratings	Personnel Data	Supervisor Ratings	Personnel Data
91	101	1	75	25
92	71	29	60	40
93	100	0	85	15
94	78	22	70	30
95	65	35	60	40
96	62	38	50	50
97	66	35	45	55
98	52	48	50	50
99	85	15	70	30
100	48	52	66	34

Spearman Rank-order Correlation Between
the Statistical and Subjective Weights

Case	Rho	Significance	Case	Rho	Significance
1	.062	.454	31	.926	.005
2	.493	.161	32	.348	.250
3	-.754	.042	33	-.185	.363
4	.377	.231	34	.353	.247
5	.717	.055	35	-.679	.070
6	-.088	.433	36	.883	.010
7	.812	.025	37	.618	.096
8	.600	.105	38	.278	.298
9	.698	.062	39	.698	.062
10	.580	.114	40	.893	.009
11	.795	.030	41	.883	.010
12	.515	.149	42	.334	.259
13	-.353	.247	43	.928	.004
14	-.174	.371	44	.883	.010
15	.828	.021	45	.277	.298
16	.577	.116	46	.618	.096
17	.828	.021	47	.938	.003
18	.638	.087	48	.821	.023
19	.131	.403	49	.371	.235
20	-.828	.021	50	.984	.001
21	.621	.095	51	.812	.025
22	.926	.005	52	.377	.231
23	-.185	.363	53	.883	.010
24	.880	.011	54	.414	.208
25	.618	.096	55	.655	.080
26	.941	.003	56	.638	.087
27	.812	.025	57	.494	.160
28	.928	.004	58	-.088	.434
29	.741	.047	59	.899	.008
30	.667	.075	60	.577	.116

(table continues)

Spearman Rank-order Correlation Between
the Statistical and Subjective Weights

Case	Rho	Significance	Case	Rho	Significance
1	.062	.454	31	.926	.005
2	.493	.161	32	.348	.250
3	-.754	.042	33	-.185	.363
4	.377	.231	34	.353	.247
5	.717	.055	35	-.679	.070
6	-.088	.433	36	.883	.010
7	.812	.025	37	.618	.096
8	.600	.105	38	.278	.298
9	.698	.062	39	.698	.062
10	.580	.114	40	.893	.009
11	.795	.030	41	.883	.010
12	.515	.149	42	.334	.259
13	-.353	.247	43	.928	.004
14	-.174	.371	44	.883	.010
15	.828	.021	45	.277	.298
16	.577	.116	46	.618	.096
17	.828	.021	47	.938	.003
18	.638	.087	48	.821	.023
19	.131	.403	49	.371	.235
20	-.828	.021	50	.984	.001
21	.621	.095	51	.812	.025
22	.926	.005	52	.377	.231
23	-.185	.363	53	.883	.010
24	.880	.011	54	.414	.208
25	.618	.096	55	.655	.080
26	.941	.003	56	.638	.087
27	.812	.025	57	.494	.160
28	.928	.004	58	-.088	.434
29	.741	.047	59	.899	.008
30	.667	.075	60	.577	.116

(table continues)

Case	Rho	Significance	Case	Rho	Significance
61	.319	.269	81	.500	.157
62	.892	.009	82	.370	.235
63	.971	.001	83	.406	.213
64	.500	.157	84	-.377	.231
65	.359	.243	85	.598	.106
66	.765	.039	86	.754	.042
67	.717	.055	87	.395	.220
68	.371	.235	88	-.679	.070
69	.353	.247	89	.971	.001
70	.883	.010	90	.828	.021
71	.725	.052	91	.752	.043
72	-.294	.286	92	.579	.115
73	.525	.143	93	1.000	.001
74	.926	.005	94	.735	.048
75	.441	.191	95	.406	.213
76	-.420	.204	96	.206	.348
77	.609	.100	97	-.118	.413
78	-.239	.325	98	.828	.021
79	.667	.075	99	.899	.008
80	-.152	.388	100	.706	.059

Appendix C

Appendix C

Coding for Rating Task Layout

<u>Card</u>	<u>Column</u>	<u>Variable</u>	<u>Response Codes</u>
1	1-3	Identification Number	001-100
1	4	Condition	1-5
1	5-6	Profile #	01-40
1	7	Dim 1-Job Knowledge	1-9
1	8	Dim 2-Initiative	1-9
1	9	Dim 3-Attitude	1-9
1	10	Dim 4-Arrests	1-9
1	11	Dim 5-Absences	1-9
1	12	Dim 6-Complaints	1-9
1	13	Overall Rating	1-9

Coding for Rating Strategy Questionnaire

<u>Card</u>	<u>Column</u>	<u>Variable</u>	<u>Response Codes</u>
1	1-3	Identification Number	001-100
1	4	Condition	1-5
1	5	Question 3	1-5
1	6	Question 4	1-4
1	7	Question 5	1-6
1	8	Question 6A	1-3
1	9	Question 6B	1-3
1	10	Question 6C	1-3
1	11-13	Q7-Weight-Job Knowledge	000-100
1	14-16	Q7-Weight-Initiative	000-100
1	17-19	Q7-Weight-Attitude	000-100
1	20-22	Q7-Weight-Arrests	000-100
1	23-25	Q7-Weight-Absences	000-100
1	26-28	Q7-Weight-Complaints	000-100
1	29	Q8-Job Know-Credibility	1-5
1	30	Q8-Job Know-Reliability	1-5
1	31	Q8-Job Know-Quality	1-5
1	32	Q8-Initiative-Credibility	1-5
1	33	Q8-Initiative-Reliability	1-5
1	34	Q8-Initiative-Quality	1-5
1	35	Q8-Attitude-Credibility	1-5
1	36	Q8-Attitude-Reliability	1-5
1	37	Q8-Attitude-Quality	1-5
1	38	Q8-Arrests-Credibility	1-5
1	39	Q8-Arrests-Reliability	1-5
1	40	Q8-Arrests-Quality	1-5
1	41	Q8-Absences-Credibility	1-5
1	42	Q8-Absences-Reliability	1-5
1	43	Q8-Absences-Quality	1-5
1	44	Q8-Complaints-Credibility	1-5
1	45	Q8-Complaints-Reliability	1-5
1	46	Q8-Complaints-Quality	1-5

Coding Scheme for Open-Ended Questions

<u>Card</u>	<u>Col</u>	<u>Variable</u>	<u>Response Codes</u>
1	1-3	Identification Number	001-005
1	4	Experimental Condition	1-5
1	5	1st Dimension looked at (1=job know	1-6,9=missing
1	6	2nd Dimension looked at 2=initiative	1-6,9=missing
1	7	3rd Dimension looked at 3=attitude	1-6,9=missing
1	8	4th Dimension looked at 4=arrests	1-6,9=missing
1	9	5th Dimension looked at 5=absences	1-6,9=missing
1	10	6th Dimension looked at 6=complaints)	1-6,9=missing
Note: If subject (s) lists order he/she looks at dimensions, write them in. If there is no listing, leave blank. If S only lists a few, write down those and leave the rest blank.			
1	11	Was there one dimension that the subject felt was most critical, i.e. he/she used that one almost to the exclusion of all others?	0=No, 1=Yes, 9=missing
1	12	If yes to above question, list the # of the dimension that was mentioned	1-6,9=missing
1	13	Did the s mention that he/she just took an average across ALL the ratings?	0=No, 1=Yes
1	14	Did the S use an averaging strategy for just a subset of the dimensions, e.g. averaged dimensions 1,2, & 3 only ?	0=No, 1=Yes
1	15	Did the S seem to look at a couple of the dimensions first in detail, and then check the others to see if they met some minimum criterion (eg. were either very high or very low)?	0=No, 1=Yes
1	16	Did the person state that he/she totally disregarded or did not use some of the dimensions?	0=No, 1=Yes
1	17	Did they disregard Dimension 1?	0=No, 1=Yes
1	18	Did they disregard Dimension 2?	0=No, 1=Yes
1	19	Did they disregard Dimension 3?	0=No, 1=Yes
1	20	Did they disregard Dimension 4?	0=No, 1=Yes
1	21	Did they disregard Dimension 5?	0=No, 1=Yes
1	22	Did they disregard Dimension 6?	0=No, 1=Yes

Coding Scheme for Open-Ended Questions

<u>Card</u>	<u>Col</u>	<u>Variable</u>	<u>Response Codes</u>
1	1-3	Identification Number	001-005
1	4	Experimental Condition	1-5
1	5	1st Dimension looked at (1=job know	1-6,9=missing
1	6	2nd Dimension looked at 2=initiative	1-6,9=missing
1	7	3rd Dimension looked at 3=attitude	1-6,9=missing
1	8	4th Dimension looked at 4=arrests	1-6,9=missing
1	9	5th Dimension looked at 5=absences	1-6,9=missing
1	10	6th Dimension looked at 6=complaints)	1-6,9=missing
Note: If subject (s) lists order he/she looks at dimensions, write them in. If there is no listing, leave blank. If S only lists a few, write down those and leave the rest blank.			
1	11	Was there one dimension that the subject felt was most critical, i.e. he/she used that one almost to the exclusion of all others?	0=No, 1=Yes, 9=missing
1	12	If yes to above question, list the # of the dimension that was mentioned	1-6,9=missing
1	13	Did the s mention that he/she just took an average across ALL the ratings?	0=No,1=Yes
1	14	Did the S use an averaging strategy for just a subset of the dimensions, e.g. averaged dimensions 1,2, & 3 only ?	0=No, 1=Yes
1	15	Did the S seem to look at a couple of the dimensions first in detail, and then check the others to see if they met some minimum criterion (eg. were either very high or very low)?	0=No,1=Yes
1	16	Did the person state that he/she totally disregarded or did not use some of the dimensions?	0=No, 1=Yes
1	17	Did they disregard Dimension 1?	0=No, 1=Yes
1	18	Did they disregard Dimension 2?	0=No, 1=Yes
1	19	Did they disregard Dimension 3?	0=No, 1=Yes
1	20	Did they disregard Dimension 4?	0=No, 1=Yes
1	21	Did they disregard Dimension 5?	0=No, 1=Yes
1	22	Did they disregard Dimension 6?	0=No, 1=Yes

<u>Card</u>	<u>Col</u>	<u>Variable</u>	<u>Codes</u>
1	23	From question 2, did the S's rating policy change over time?	0=No, 1=Yes
1	24	Did the task become easier over time? 1=Easier, 2=Harder, 3=Stayed the same	1-3,9=missing
1	25	Why did the task become easier? 1=better notion of policy 2=S eliminated some dimensions 3=S didn't have to refer to chart 4=Some combination of 1,2, or 3	1-4,9=missing

Footnote

¹ Hoffman used relative weights for comparison with the subjective weights. The formula for generating relative weights is as follows:

$$\text{relative weight} = \frac{B_{oi} C_{oi}}{R^2_{0,1,\dots,K}}$$

where B_{oi} = beta coefficient for the i th predictor
 C_{oi} = validity coefficient for the i th predictor
 R^2 = squared multiple correlation coefficient

Hoffman (1960) notes that relative weights are used rather than Beta weights because relative weights are comparable from one judge to the next, are theoretically capable of accounting for all of the predictable variance, and carry exact interpretation in terms of components of variance.

² Only a range of values for the modal R^2 were given by Zedeck and Cascio (1982). This value represents the most frequent category into which an individual's R^2 fell.

References

- Aiken, L.R. (1979). Psychological testing and assessment (3rd ed.). Boston: Allyn & Bacon.
- Alexander, E., & Wilkins, R. (1982). Performance rating validity: The relationship of objective and subjective measures of performance. Group and Organizational Studies, 7, 485-496.
- Anderson, B. (1977). Differences in teachers' judgment policies for varying numbers of verbal and numerical cues. Organizational Behavior and Human Performance, 19, 68-88.
- Balzer, W., Rohrbaugh, J., & Murphy, K. (1983). Reliability of actual and predicted judgments across time. Organizational Behavior and Human Performance, 32, 109-123.
- Bass, A., & Turner, J. (1973). Ethnic group differences in relationships among criteria of job performance. Journal of Applied Psychology, 57, 101-109.
- Beach, L.R., Mitchell, T.R., Deaton, M.D., & Prothero, J. (1978). Information relevance, content and source credibility in the revision of opinions. Organizational Behavior and Human Performance, 21, 1-16.
- Bernardin, H., Alvarez, K., & Cranny, C. (1976). A recomparison of BES to summated scales. Journal of Applied Psychology, 61, 564-570.
- Billings, R., & Marcus, S. (1983). Measures of compensatory and non-compensatory models of decision behavior: Process tracing versus policy capturing. Organizational Behavior and Human Performance, 31, 331-352.
- Birnbaum, M., & Stegner, S. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. Journal of Personality and Social Psychology, 37, 48-74.
- Birnbaum, M., Wong, R., & Wong, L. (1976). Combining information from sources that vary in credibility. Memory and Cognition, 4, 330-336.
- Borman, W., & Dunnette, M. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 60, 561-565.

- Brady, D., & Rappaport, L. (1973). Policy-capturing in the field: The nuclear safeguards problem. Organizational Behavior and Human Performance, 9, 253-266.
- Cascio, W.F. (1982). Applied Psychology in Personnel Management. Reston, Virginia: Reston Publishing.
- Cascio, W.F., & Valenzi, E. (1978). Relations among criteria of police performance. Journal of Applied Psychology, 63, 22-28.
- Christal, R.E. (1968). JAN: A technique for analyzing group judgment. Journal of Experimental Education, 36, 24-27.
- Cook, R., & Stewart, T. (1975). A comparison of seven methods for obtaining subjective descriptions of judgmental policy. Organizational Behavior and Human Performance, 13, 31-45.
- Cooper, W.H. (1981). Conceptual similarity as a source of illusory halo in job performance ratings. Journal of Applied Psychology, 66, 302-307.
- Cotton, B., Jacobs, R., & Grogan, J. (1983). Use of individually scaled versus normatively scaled predictor cues in policy-capturing research. Applied Psychological Measurement, 7, 159-171.
- Dawes, R. (1971). A case study of graduate admissions: Application of 3 principles of human decision making. American Psychologist, 26, 180-186.
- Dawes, R. (1979). The robust beauty of improper linear models in decision making. American Psychologist, 34, 571-582.
- DeNisi, A., Cafferty, T., & Meglino, B. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.
- Dickinson, T., & Zellinger, P. (1980). A comparison of the behaviorally anchored rating and mixed standard scale formats. Journal of Applied Psychology, 65, 147-154.
- Doherty, M., & Keeley, S. (1972). Use of subjective predictors in regression analysis for policy capturing. Journal of Applied Psychology, 56, 277-278.
- Ebert, R., & Kruse, T. (1978). Bootstrapping the security analyst. Journal of Applied Psychology, 63, 110-119.

- Einhorn, H. (1971). Use of nonlinear noncompensatory models as a function of task and amount of information. Organizational Behavior and Human Performance, 6, 1-27.
- Einhorn, H., Kleinmuntz, D., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. Psychological Review, 86, 463-485.
- Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Goldberg, L. (1971). Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. Organizational Behavior and Human Performance, 6, 438-479.
- Hobson, C.J., & Gibson, F.W. (1983). Policy capturing as an approach to understanding and improving performance appraisal: A review of the literature. Academy of Management Review, 8, 640-649.
- Hobson, C.J., Mendel, R.M., & Gibson, F.W. (1981). Clarifying performance appraisal criteria. Organizational Behavior and Human Performance, 28, 164-188.
- Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. Psychological Bulletin, 57, 116-131.
- Hogarth, R.M. (1980). Judgement and Choice. New York: John Wiley and Sons.
- Ilgen, D., & Feldman, J. (1983). Performance appraisal: A process focus. In Staw & Cummings (Eds.), Research in Organizational Behavior, Vol. 5. Greenwich, Connecticut: JAI Press.
- Kozlowski, S.W., Kirsch, M.P., & Chao, G.T. (in press, 1986). Job knowledge, rates familiarity, conceptual similarity, and halo error: An exploration. Journal of Applied Psychology.
- Landy, F., & Farr, J. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Lane, D., Murphy, K., & Marques, T. (1982). Measuring the importance of cues in policy capturing. Organizational Behavior and Human Performance, 30, 231-240.

- Einhorn, H. (1971). Use of nonlinear noncompensatory models as a function of task and amount of information. Organizational Behavior and Human Performance, 6, 1-27.
- Einhorn, H., Kleinmuntz, D., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. Psychological Review, 86, 465-485.
- Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Goldberg, L. (1971). Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. Organizational Behavior and Human Performance, 6, 458-479.
- Hobson, C.J., & Gibson, F.W. (1983). Policy capturing as an approach to understanding and improving performance appraisal: A review of the literature. Academy of Management Review, 8, 640-649.
- Hobson, C.J., Mendel, R.M., & Gibson, F.W. (1981). Clarifying performance appraisal criteria. Organizational Behavior and Human Performance, 28, 164-188.
- Hoffman, P.J. (1960). The paramorphic representation of clinical judgment. Psychological Bulletin, 57, 116-131.
- Hogarth, R.M. (1980). Judgement and Choice. New York: John Wiley and Sons.
- Ilgen, D., & Feldman, J. (1983). Performance appraisal: A process focus. In Staw & Cummings (Eds.), Research in Organizational Behavior, Vol. 5. Greenwich, Connecticut: JAI Press.
- Kozlowski, S.W., Kirsch, M.P., & Chao, G.T. (in press, 1986). Job knowledge, rater familiarity, conceptual similarity, and halo error: An exploration. Journal of Applied Psychology.
- Landy, F., & Farr, J. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Lane, D., Murphy, K., & Marques, T. (1982). Measuring the importance of cues in policy capturing. Organizational Behavior and Human Performance, 30, 231-240.

- Latham, G., & Wexley, K. (1977). Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 30, 255-268.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity to process information. Psychological Review, 63, 81-97.
- Muckler, F.A. (1982). Evaluating productivity. In M.D. Dunnette & E.A. Fleishman (Eds.), Human performance and productivity: Vol 1. Human capability assessment. Hillsdale, NJ: Erlbaum.
- Naylor, J., & Wherry, R. (1965). The use of simulated stimuli and the JAN technique to capture and cluster the policies of raters. Educational and Psychological Measurement, 25, 969-986.
- Nisbett, R., & Ross, L. (1980). Human Inference: Strategies and Shortcomings. Englewood Cliffs, New Jersey: Prentice-Hall.
- Ogilvie, J., & Schmitt, N. (1979). Situational influences on linear and nonlinear use of information. Organizational Behavior and Human Performance, 23, 292-306.
- O'Reilly, C. (1983). The use of information in organizational decision making: A model and some propositions. In Staw & Cummings (Eds.), Research in Organizational Behavior, Vol. 5, Greenwich, Connecticut: JAI Press.
- Payne, J., Braunstein, M., & Carroll, J. (1978). Exploring predecisional behavior: An alternative approach to decision research. Organizational Behavior and Human Performance, 22, 17-44.
- Rosenbaum, M., & Levin, I. (1969). Impression formation as a function of source credibility and the polarity of information. Journal of Personality and Social Psychology, 12, 34-37.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. Psychological Bulletin, 66, 178-200.
- Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 26, 79-101.
- Schmitt, N., & Levine, R. (1977). Statistical and subjective weights. Organizational Behavior and Human Performance, 20, 15-30.

- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.
- Smith, P. C. (1976). Behavior, results, and organizational effectiveness. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.
- Smith, P., & Kendall, L. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.
- Stumpf, S., & London, M. (1981). Capturing rater policies in evaluating candidates for promotion. Academy of Management Journal, 24, 752-766.
- Surber, C. (1981). Effects of information reliability in predicting task performance using ability and effort. Journal of Personality and Social Psychology, 40, 977-989.
- Svenson, O. (1979). Process descriptions of decision-making. Organizational Behavior and Human Performance, 23, 86-112.
- Taylor, R., & Wilsted, W. (1974). Capturing judgment policies: A field study of performance appraisal. Academy of Management Journal, 17, 440-449.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
- Weiss, D. (1979). The effects of systematic variations in information on judges' descriptions of personality. Journal of Personality and Social Psychology, 37, 2121-2136.
- Zedeck, S., & Cascio, W. (1982). Performance appraisal decisions as a function of rater training and purpose of appraisal. Journal of Applied Psychology, 67, 752-758.
- Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. Organizational Behavior and Human Performance, 18, 269-294.

MICHIGAN STATE UNIV. LIBRARIES



31293007029766

MICHIGAN STATE UNIV. LIBRARIES



31293007029766