INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Performance cue effects in work behavior ratings: Memory or response bias?

Willis, Cynthia Emrich, M.A.

Rice University, 1991
RICE UNIVERSITY

PERFORMANCE CUE EFFECTS IN WORK BEHAVIOR RATINGS:
MEMORY OR RESPONSE BIAS?

by

CYNTHIA EMRICH WILLIS

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
MASTER OF ARTS

APPROVED, THESIS COMMITTEE

Richard F. Martell, Assistant Professor Chair (Psychology)

David J. Schneider, Professor (Psychology)

Michael J. Watkins, Professor (Psychology)

Houston, Texas
April, 1991
Performance Cue Effects in Work Behavior Ratings:

Memory or Response Bias?

by

Cynthia Emrich Willis

Abstract

This study examined whether a memory or response bias mediated the effects of performance cues on observers' recollections of a work group's behavior. Fifty-nine subjects observed a film of a group at work. Then, immediately or one week later, subjects rated the group's performance using behavioral, evaluative, and objective outcome rating instruments. Prior to observing the group, subjects were given performance cues that led them to believe the group had either performed well or poorly in the task. Results identified a systematic response bias and not a memory bias as the cognitive process mediating the effects of performance cues in work behavior ratings: Subjects adopted a more liberal decision criterion when attributing effective (ineffective) behaviors in response to positive (negative) performance cues. Unlike the behavioral ratings, recollections of specific group outcomes were immune to the biasing effect of performance cues.
Acknowledgements

To Rich, my esteemed advisor . . . Thank you! Your constant support and encouragement mean a lot. Also, your successful campaign to secure a research fellowship could not have come at a better time.

To Dave and Mike, my committee members, thank you for adding your own special touches to my days as a graduate student. I hope that you will someday be able to look back on the time you spent serving on my committee with fondness.

To Marnie and Mae, my partners-in-crime. Thanks for everything and Good Luck!

To Mike and Matthew, my nearest and dearest. With your love and support, I am one step closer. I thank Mike for listening to me “gripe, complain, and carry-on” and also for reading manuscripts that no person outside of psychology should be expected to read. I thank Matthew for asking countless questions beginning with “Why?” because they make me pause long enough to step back and get a look at the “big picture.”
# Table of Contents

List of Tables

List of Figures

Introduction

- The Performance Cue Effect
- 3
- Moderators of the Performance Cue Effect
- 4
- Generalizability of the Performance Cue Effect
- 8
- Potential Mediators of the Performance Cue Effect
- 8
- Goals of this Research
- 13
- Experimental Hypotheses
- 15

Method

- Procedure
- 16
- Independent Variables
- 16
- Dependent Variables
- 17

Results

- Manipulation Check
- 18
- Behavioral Ratings (Hit Rates)
- 19
- Interpretational Limits of Hit Rates
- 21
- Introduction to Memory Sensitivity (Pr) and Response Bias (Br)
- 23
- Memory Sensitivity (Pr)
- 25
- Response Bias (Br)
- 25
Table of Contents, Continued

Further Investigation of Response Bias (Br) as Mediator of the Performance Cue Effect 27

Effective Behaviors 29
Ineffective Behaviors 32

Evaluative Ratings of Group Process 34
Outcome Ratings 35

Discussion 38
References 45

Appendices

Introduction to Videotape 47
Behavioral Rating Instrument 48
Evaluative Rating Instrument 54
Objective Outcome Rating Instrument 55
List of Tables

Mean Ratings of Work Behavior: Hit Rates and False Alarm Rates
Mean Ratings of Memory Sensitivity, and Response Bias (Br)
Mean Ratings of Group Process and Recollections of Group Outcome
List of Figures

Mean Hit Rate as a Function of Performance Cues

Mean Decision Criterion ($Br$) as a Function of Performance Cues

Mediation Model for Ratings of Effective Behavior.

Mediation Model for Ratings of Ineffective Behavior
Introduction

Performance Cue Effects in Work Behavior Ratings: Memory or Response Bias?

The accurate and unbiased appraisal of work performance is critical to both industrial/organizational psychologists and organizations. The degree to which a performance appraisal system meets these objectives can mean the difference between sound promotion, termination, and pay decisions and poor ones; motivated, satisfied, and committed employees and employees disgruntled and perhaps confused as to the organization's expectations regarding their work performance. With so much at stake, researchers have spent the past several decades working to improve the quality of performance appraisal systems. The bulk of this work has focused on the development of new rating formats with the goal of identifying those immune to the numerous rater biases and errors that have plagued already existing methods.

This approach has resulted in a number of carefully developed scales designed to elicit ratings based on raters' observations of work behaviors. Researchers believed that these behaviorally-based ratings would circumvent raters' tendencies to rely on traits, implicit theories, or overall impressions. A study conducted by Murphy, Martin, and Garcia (1982) highlighted the fallibility of this underlying assumption. In their study, behaviorally-based ratings and evaluative ratings were collected under two conditions of memory demand. In the condition of little or no demand, the behaviorally-based scale produced ratings that were moderately correlated with those derived from evaluative ratings based on trait judgments;
however, as the demands on memory increased to only a modest level (i.e., a 24-hour delay), the correlation between the two types of ratings increased significantly leading the researchers to conclude that "these [behavioral] observations [were] in fact disguised measures of traitlike performance judgments" (p. 566).

Subsequent research on the accuracy of the various behaviorally-based rating scales has provided further evidence of the intrusion of rater bias and error. One bias, in particular, has broad implications for the appraisal of individuals and groups in work organizations and has been labeled "the performance cue effect." Originally, this effect was used to describe instances in which individuals led to believe that a group has performed well (versus poorly) evaluate what went on in the group (group process) and the performance of the group (group outcome) more favorably despite identical or equivalent group performance. Observers' reports of specific and observable work behaviors were thought to be immune to the biasing effects of performance cues given the relative ease of "merely" reporting what one observes versus having to observe and then form some type of evaluative judgment. This view was recently corrected by Martell and Guzzo (in press) who found that performance cues biased subjects' recollections of a work group's behavior such that individuals led to believe that a group had performed well (versus poorly) attributed more effective and fewer ineffective behaviors to an identically performing work group. From this study and numerous others, it appears that performance cues bias observers' reports of specific observable behaviors as well as their evaluations of group process and outcome.
The Performance Cue Effect

The performance cue effect was first demonstrated in a study conducted by Staw in 1975. Staw set out to test the assumption that the relationship demonstrated in correlational research between perceptions of various aspects of task performance and performance ratings represented the following causal flow:

perceptual judgments -------------> performance ratings
(e.g., task behavior)

Based on work in the area of attribution theory, Staw posited a reversal in this causal order arguing that "individuals utilize knowledge of performance as a cue by which they ascribe characteristics to an individual, group, or organizational unit" (p. 416). To test this theory, Staw randomly assigned subjects to three-person groups and gave each newly-formed group the task of estimating company sales and earnings per share for an obscure electronics company described in a company report. Immediately following participation in the task, each group was given bogus feedback regarding its performance in the task relative to other groups. Groups were either told that they had performed "quite well" (in the top 20% of all groups) or "quite poorly" (in the bottom 20% of all groups) and then group members completed a questionnaire regarding "what went on in the group" (individual contribution and group process) and the group's performance in the task (group outcome and manipulation check). In line with Staw's prediction, groups told they had performed well rated their group process (e.g., cohesiveness, communication, cooperation, openness to change, ability, etc.) more favorably than did groups told they had performed poorly despite equivalent performances in the two feedback conditions. Also, self-ratings of individual contribution (e.g., cooperation,
effort, ability, motivation, openness to change, etc.) revealed the same 
pattern of bias. Staw concluded that "mere" knowledge of performance was 
a "relatively potent independent variable" (p. 425) leading individuals to 
attribute traits consistent with the stereotype "cued" by knowledge of 
performance. In this case -- stereotypes of high versus low performing 
groups.

Staw's research has important implications for both theory 
development and performance appraisals conducted in work organizations. 
Prior to Staw's research, researchers, for the most part, relied on self-
report questionnaires to capture the behaviors, traits, and characteristics 
assumed to represent the "antecedents" of performance. According to 
Staw's findings, this approach to theory development is more likely to 
provide information about raters' naive theories of how individuals and 
groups function in organizations rather than the actual causes of 
performance in a given work situation. The implications for performance 
 appraisals in organizations are equally important. The potential for 
knowledge of performance, in some form, to bias reports of the 
characteristics, behaviors, and traits of individuals and groups is 
enormous in any given work organization.

Moderators of the Performance Cue Effect

Numerous studies have since been conducted on the performance 
cue effect in an effort to increase the field's understanding of the extent to 
which performance cues might bias performance ratings in organizations. 
The early work inspired by Staw's research can best be characterized as a 
directed effort to identify variables that might limit the generalizability of 
this effect to other research settings and, especially, actual work 
organizations.
Downey, Chacko, and McElroy (1979) questioned the generalizability of the Staw (1975) study on the basis that the groups in his study were newly formed, lacking any "group history" to draw upon when confronted with bogus performance feedback. The authors believed that group history (a characteristic of most groups in organizations) would provide sufficient time and opportunity for group members to observe and assess their group's performance, thereby, significantly reducing their reliance on performance outcomes in making causal attributions about their group and group members. To test this hypothesis, Downey et al. enlisted 26 three-person groups that had been working on a variety of financial problem-solving tasks for approximately 12 weeks. Half of the groups remained intact (group history condition) whereas individuals in the remaining groups were "scrambled" to form groups with no prior group history. Replicating Staw's experimental procedure, groups engaged in a financial problem-solving task and then received either good or poor bogus performance feedback. Questionnaire responses collected immediately following the feedback revealed a significant effect of tone of performance feedback such that members of groups receiving good performance feedback evaluated both individual contribution and group process more favorably than did members of groups receiving poor performance feedback despite equivalent performance on the experimental task. More importantly, the results revealed "little impact [of group history] on individual perceptions of intragroup processes" (p. 297). Specifically, the impact of group history did not moderate the effect of performance feedback and, as a main effect, was limited to group members' evaluations of individual and group motivation in solving the financial puzzle.
Binning and Lord (1980), in a similar attempt to identify boundary variables for the performance cue effect, explored the possibility that performance cues were utilized not only in instances of inadequate information, but also as a tool to simplify complex information. The implication here was that performance cues might operate in highly familiar groups as well as groups with no prior history thereby significantly increasing the potential for performance cues to operate in work organizations. In Binning and Lord's study, newly-formed groups performed a series of three tasks in separate experimental sessions. Each task required approximately one hour to complete with a 48-hour interval between task sessions. The sessions were set up in this manner to create sufficient time and opportunity for group members to work together on a variety of group tasks, providing a moderate degree of familiarity. Bogus feedback was given to groups immediately following the completion of the task in each of the three experimental sessions. The feedback across tasks varied in strength and consistency, creating four feedback conditions. In two conditions, the performance feedback was consistently good or poor across the three tasks. The remaining two conditions consisted of either "good, poor, good" or "poor, good, poor" feedback, in that order. Binning and Lord reasoned that a plausible information simplification strategy was that of merely "averaging" the performance feedback. Such a strategy would result in a "consistent linear trend in process ratings across conditions as feedback becomes more favorable" (p. 121). To explore this possibility, the authors modified Staw's original questionnaire so that the items referred to individual contribution and group process spanning "all three tasks." It was then administered to subjects following feedback on the third task. Group members' responses on the questionnaire were consistent with the
"averaging" hypothesis with the exception that poor performance feedback resulted in larger average adjustments than did good performance feedback. Demonstration of this bias in groups with a moderate degree of group history provided support for the notion that individuals rely on performance cues not only in instances of inadequate information but also in situations in which complex information must be simplified in order to evaluate individual contribution and group process.

Another potential moderator with implications for performance ratings made in organizations is the length of temporal delay between the observation of work behavior and the formal appraisal of an individual's work performance. It is clear that memory, including memory for observed behaviors, decays over time. This decay forces raters to rely on traits and impressions rather than specific behavioral information in assessing an employee's work performance. As one type of generalized impression, performance cues should bias delayed ratings more than immediate ratings. Rush, Phillips, and Lord (1981) examined this possibility by collecting subjects' ratings of a group leader either immediately or 48 hours after they observed the group's performance. As in previous performance cue research, bogus performance feedback was administered to subjects immediately after they viewed the videotaped vignette. The authors hypothesized that "... over time, the percentage of variance associated with performance [cues] would increase" (p. 445). Analysis of subjects' responses on a 20-item behavior rating instrument provided no evidence of an increased impact of performance cues on delayed (versus) immediate performance ratings. To explain the equivalence of the immediate and delayed ratings, Rush et al. suggested that "subjects [in both immediate and delayed rating conditions] may have
integrated information concerning the leader's behavior into a simplified, more easily remembered form, utilizing an already existing leadership [category]" (p. 446).

**Generalizability of the Performance Cue Effect**

From studies conducted up to this point, researchers were able to conclude that the effects of performance information on ratings of "what goes on in groups" generalized to ratings:

1) made by both observers of groups and group members,
2) of both individuals and groups,
3) based on varying amounts of information (i.e., inadequate to, perhaps, more than adequate),
4) made by members of intact groups with a moderate degree of group history as well as newly-formed groups, and
5) made in both immediate and delay conditions.

Thereafter, the emphasis of research shifted to identifying possible mediators in an effort to one, better understand the mechanisms responsible for the performance cue effect; and two, to devise strategies or interventions that would be successful in attenuating it.

**Potential Mediators of the Performance Cue Effect**

Based on Staw's original paper (1975), researchers assumed that performance cues operated by leading subjects to rely upon stereotypes, categories, or naive theories that were consistent with the cues in attributing traits and characteristics to an individual or group. A study by Phillips and Lord (1982) was conducted to more directly test this assumption. In their study, subjects observed a videotape of a four-person problem-solving group. One group member was designated as the leader and was coached to exhibit 5 behaviors of each of the following behavior
types: prototypic of an effective leader, prototypic of an ineffective leader, and nonprototypic (i.e., behavior that one would expect equally from an effective versus ineffective leader). After viewing the videotape, subjects were either told that that the group had performed well or poorly. Subjects then completed a 30-item questionnaire that consisted of 10 behavior descriptions from each behavior type. Within each behavior type, half actually occurred and half did not and the subjects' task was to discriminate between those behaviors that were present versus absent. The authors' experimental hypotheses were designed to test the categorization process advanced earlier. Accordingly, their first hypothesis was that subjects would have difficulty correctly distinguishing between prototypical behaviors, both effective and ineffective, that were present versus absent in the videotape. In their second hypothesis, they predicted that subjects would show better performance in distinguishing between present versus absent nonprototypic behaviors. Their third experimental hypothesis was that tone of performance cues (good versus poor) would lead to the reporting of behaviors consistent with the cued category (prototypically effective or ineffective, respectively), but have no effect on the reporting of nonprototypic behaviors. The results from the appropriate analyses provided evidence for the three experimental hypothesis and were, therefore, consistent with the notion that performance cues influenced behavioral ratings of prototypical behaviors (but not nonprototypical ones) by cuing a category prototype that led subjects to reconstruct the stimulus behavior in a manner consistent with the prototype.

The study conducted by Phillips and Lord (1982) served to test more explicitly the widely-held notion that performance cues influence performance ratings by leading subjects to access cue-consistent categories
or stereotypes. However, the question still remained as to the precise mediating processes responsible for the performance cue effect on performance ratings.

The first study to provide a direct test of possible mediating processes was one conducted by Martell and Guzzo (in press) The goals of this study were twofold: one, to determine whether specific observable stimulus behaviors were subject to biasing effects of performance cues; and two, to identify the cognitive processes underlying any effects of performance cues on subjects' recollections of group behaviors. From past research, it was clear that performance cues systematically biased subjects evaluations of group process and group outcome. A logical extension of these findings to behavioral ratings led the authors to predict that observers who received positive (versus negative) performance cues would attribute more effective and fewer ineffective behaviors to an identically performing work group. Based on the literature in social cognition, two very different processes were posited as potentially underlying these hypothesized effects of performance cues on observers' recollections of group behavior. The first was a memory bias whereby performance cues influence recollections by activating observers' stereotypes or naive theories, making more accessible behaviors that are consistent with performance cues (i.e., effective behaviors given positive cues, and vice versa). An alternative process was a probabilistic response bias. Here, performance cues activate observers' stereotypes that, in turn, make consistent behaviors seem more "plausible." These consistent behaviors seem more plausible because observers have altered their decision criteria to reflect the accessed stereotype. That is, when judging the occurrence of a consistent behavior, observers employ a more
lenient or liberal criterion than when judging the occurrence of an inconsistent behavior.

In order to investigate these two possibilities, Martell and Guzzo had subjects observe a videotape of a work group attempting to get a box and group members across a pool of water. Immediately after subjects viewed the videotape, they received either good, poor, or no bogus feedback regarding the group's performance in the task and then immediately or one-week later completed a 40-item questionnaire containing descriptions of both effective and ineffective behaviors. Within each type of behavior, half of the behaviors were present in the videotaped vignette and half were absent. The subjects' task in completing the behavioral ratings was to discriminate between behaviors that were present versus absent. From subjects' responses, measures of memory sensitivity (Pr) and response bias (Br) were calculated in accordance with the method outlined by Snodgrass and Corwin (1988). These measures served as dependent variables in a 3 (Tone of Performance Cues: good, poor, none) X 2 (Time of Rating: immediate versus one-week delay) analysis of variance. Consistent with Rush, Phillips, and Lord (1981), ratings made in the delay condition were no more biased than those made in the immediate rating condition. However, contrary to current belief, the effect of performance cues extended beyond evaluative ratings of group process and outcome to the ratings of specific observable behaviors. Further investigation of this bias revealed that a systematic response bias and not a memory bias mediated the effects of performance cues on behavioral ratings of the work group. That is, performance cues operated by influencing subjects' decision criteria in such a way that they were more lenient in judging the occurrence of behaviors that were consistent with the administered performance cues.
Alternatively, subjects applied a more stringent criterion when judging the occurrence of inconsistent behaviors. These results suggest that performance cues operate by activating observers' stereotypes or naive theories. However, this activation did not result in improved or enhanced memory for cue-consistent behaviors, but in a systematic adjustment of subjects' decision criteria in favor of cue-consistent behaviors.

A limitation of the Martell and Guzzo study was that, like most research in this area, the performance cues were administered only after subjects had observed the work group. For both theoretical and practical reasons, it is critical to extend the scope of research to the study of pre-observation performance cues.

The current trend in theoretical development is to identify underlying cognitive processes mediating the effects of performance cues. Two processes were investigated in Martell and Guzzo's study. From past research in social cognition, it is clear that a response bias operates at time of rating. However, every step in the rating process (i.e., attention, perception, encoding, storage, and retrieval) has implications for memory. For example, it is necessary for one to attend to a behavior in order for that behavior to be a candidate for further processing. In addition, the fact that two raters pay equal attention to a given behavior does not guarantee similar perceptions. Alternatively, two raters may, in fact, have similar perceptions, however, one takes greater pains to encode the behavior for future reference. The point of this line of reasoning is that administering performance cues post-observation precludes differential processing at any of these steps (i.e., attention, perception, and encoding) necessarily disadvantaging the mediation of performance cues by a memory bias. This is consistent with Fiske and Taylor's (1991) review of the research in
schematic processing. Specifically, that the effects of schemas, impressions, and the like most strongly affect memory when present from the beginning of observation.

The more practical goal of current research in this area is to develop interventions to circumvent the operation of performance cues in work organizations. Performance cues have been shown to distort both evaluative and behavioral ratings. This distortion has serious implications for theory development and for performance appraisal in organizations. Theories are developed in an attempt to explain phenomena and predict future events. If theories reflect subjects' beliefs about the way individuals and groups function in organizations rather than the way they actually function, theories will fail to achieve these designated purposes. For this same reason, performance appraisals based on raters' systematically distorted judgments will prove inadequate. In organizations, performance appraisals are often used for training and development with special attention given to identifying deficits in order to aid employees in improving their work performance. From past research, it is clear that raters who are influenced by positive performance cues or information will overattribute effective behaviors and characteristics and underattribute ineffective ones. This underattribution of ineffective behaviors and characteristics will result in a failure to identify employee deficits to the detriment of both employees and the organization.

Goals of This Research

The first goal of the present study was to provide a fair test of both a memory and response bias as potential mediators of the performance cue effect. This was accomplished by providing performance cues regarding a work group's performance prior to subjects observing the group. In the
study, subjects came to the experimental session and were told that they would observe a work group as it attempted to get a box and group members across a pool of water and then evaluate the group's performance either immediately after observing the group or one week later. Subjects did not receive more specific information regarding when they would be required to evaluate the group to preclude the possibility that they would adopt systematically different information processing strategies in response to the time of rating manipulation. Prior to observing the group, the experimenter administered either positive (top 25% of competing groups) or negative (bottom 25% of competing groups) bogus performance cues both orally and in writing.

A second goal of the study was to determine whether ratings of objective group outcomes (e.g., "How many group members made it across the pool?") might be immune to the biasing effects of performance cues. Different types of rating formats can be viewed as falling somewhere along a continuum in terms of their level of objectivity. For example, evaluative ratings are generally considered to be one of the most highly subjective types of performance ratings as they require raters to not only observe an employee's behavior but to then infer underlying traits, attributes, and characteristics from these observations. Behavioral ratings fall somewhere in the middle involving some degree of subjective judgment in interpretation or perception of the observed behaviors. Objective outcome ratings represent the most highly objective ratings possible in performance ratings. In this case, observers need only attend closely enough to an individual's or group's performance so that they are able to say, for example, "I counted four group members successfully crossing the pool of water." Obviously, there is very little room for interpretation in this
example. With this continuum of objectivity in mind, subjects completed evaluative ratings of what went on the group (group process), behavioral ratings, and objective outcome ratings. Including all three formats in the study made it possible to assess the relative influence of performance cues.

**Experimental Hypotheses**

1) It was predicted that subjects given positive performance cues would attribute more effective and fewer ineffective behaviors to the observed group. In contrast, subjects given negative performance cues were expected to exhibit the opposite pattern of responding, attributing fewer effective and more ineffective behaviors to the observed group.

2) A systematic response bias was expected to mediate the effects of performance cues on behavioral ratings such that subjects given positive cues would report more effective and fewer ineffective behaviors than would subjects given negative cues.

3) A systematic memory bias was expected to mediate the effects of performance cues on behavioral ratings such that subjects would show improved memory accuracy for behaviors that were consistent with the performance cues (e.g., effective behaviors if given positive cues).

4) Performance cues were expected to influence evaluative ratings of group process such that subjects who received positive performance cues would rate the group's process more favorably than would subjects given negative performance cues.

5) Ratings of specific group outcomes were expected to be unaffected by knowledge of group performance.
Method

Procedure

Fifty-nine subjects viewed the same 14-minute film of a work group and then rated the observed group using evaluative, objective outcome, and behavioral rating instruments. The film depicted a group of five men attempting to build a bridge of planks and ropes (see Appendix A) with the goal of transporting themselves and a box across a pool of water. Prior to observing the group, subjects were given positive or negative cues concerning the group's performance. A manipulation check was administered to all subjects immediately after they viewed the film.

Independent Variables

Performance Cues. Prior to observing the group, subjects were told that the group had performed either quite well (positive cues) or quite poorly (negative cues) in comparison to other groups in the contest.

These cues were communicated (orally and in writing) by telling subjects that the observed group:

"... was rated by a panel of experts specially trained in the observation of group performance. These experts rated the group on several aspects of group performance, including such things as the time it took to cross the bridge, the number of participants to reach the other side, and the time to get the box across. Compared with other groups performing this task, the group you observed was judged to be in the [bottom/top] quarter of all groups. That is, the performance of this group was [quite poor/quite good] in comparison with other groups."

Time of Rating. Subjects in the immediate condition completed their ratings immediately after viewing the film while subjects in the delayed condition completed theirs one week later. The immediate rating condition
required one session lasting approximately 40 minutes. Two 20-minute sessions were required for the delayed condition. Subjects were not informed of the timing of their ratings until after they viewed the film.

Behavior Type. Behavior ratings were collected using a 40-item questionnaire. Of the 40 items, 22 described effective behaviors and 18 described ineffective behaviors.

Dependent Variables

Behavioral Ratings. Subjects' recollections of the group's behavior were assessed with a 40-item questionnaire (see Appendix B). Using a 6-point recognition memory scale with endpoints labeled (1) "very certain the behavior did not occur" to (6) "very certain the behavior did occur," subjects judged whether the 20 behaviors depicted in the film (11 effective and 9 ineffective) did or did not occur. In addition, 11 effective and 9 ineffective behaviors not depicted in the film were included. Behavior descriptions included in the questionnaire were designed to be specific and concrete. Some sample items are: "While attempting to cross the pool, a group member carefully explained to the others how to best position their feet on the bridge to avoid falling in the water." (effective, did occur); "Several group members stated that they enjoyed working in the group." (effective, did not occur); "One group member yelled in a loud, unpleasant voice at another that he was balancing on the makeshift bridge in the wrong way." (ineffective, did occur); "One group member shoved another out of the way when reaching for the box." (ineffective, did not occur). An independent set of judges confirmed the classification of ineffective and effective items (Martell & Guzzo, in press).

Evaluative Ratings of Group Process. Subjects rated various aspects of the observed group's process using a 10-point rating scale, endpoints
labeled -- depending on the item -- (1) "very low" or "not at all" and (10) "very high" or "to a great extent." The rating instrument (see Appendix C) required subjects to indicate their views on seven group process issues: quality of communication, effort, coordination, team spirit, quantity of communication, enjoyment of working together, and commitment to the task. Ratings were averaged to form a single mean composite indicating the evaluation of group process (coefficient alpha = .91).

**Outcome Ratings.** Subjects completed a three-item scale composed of specific group outcomes that included: number of group members to successfully cross the pool, number of times a group member fell into the water, and number of times the box was dropped into the water (see Appendix D). This scale required subjects to simply place a check next to the correct response. Values ranged from 0 to 5 for the three objective outcome items. The three items could not be combined to represent a single factor or dimension as indicated by a coefficient alpha = .11; therefore, a separate analysis of variance was conducted for each item.

Rating packets were constructed so that the order of the objective outcome and evaluative rating instruments were counterbalanced. Behavior ratings were always completed last. Initial analyses revealed no interaction of Rating Order with the independent variables of interest. Therefore, subsequent analyses were conducted collapsing across this factor.

**Results**

**Manipulation Check**

To verify the effectiveness of the performance cue manipulation, subjects evaluated the group's performance "in terms of transporting the box and people across the pool" using a seven-point rating scale, endpoints
ranging from (1) "extremely poor" to (7) "excellent." As expected, performance cues influenced evaluations of the group's performance such that subjects who received positive cues evaluated the group more favorably ($M = 5.11$) than did subjects receiving negative cues ($M = 2.90$), $F (1,58) = 103.41, p < .001$.

**Behavioral Ratings (Hit Rates)**

It was predicted that subjects given positive (negative) cues would attribute more (fewer) effective behaviors and fewer (more) ineffective behaviors to the observed group. To test this prediction, overall hit rates were calculated for each subject. A "hit" was indicated when a subject assigned a rating of 4, 5, or 6 to an effective or ineffective behavior depicted in the film. Ratings of 4, 5, or 6 assigned to a behavior that did occur indicate some degree of confidence on subjects' parts that they did, in fact, observe the group engaging in the described behavior. Scoring of this nature reduced the original 6-point scale to a dichotomous yes/no recognition memory task. Hit rates represent the conditional probability of responding "yes" to a previously observed (i.e., "old") behavior with values ranging from 0 (no behaviors reported) to 1.0 (all behaviors reported). The mean values for hit rates for effective and ineffective behaviors appear in Table 1.

-----------------------------------

Insert Table 1 about here

-----------------------------------

A 2 (Performance Cues: positive versus negative) X 2 (Time of Rating: immediate versus delay) X 2 (Behavior Type: effective versus ineffective) repeated measures analysis of variance was conducted using overall hit rates as the dependent variable. The only effect associated with the time of
Table 1
Mean Ratings of Work Behavior: Hit Rates and False Alarm Rates

<table>
<thead>
<tr>
<th></th>
<th>Effective Work Behavior</th>
<th>Ineffective Work Behavior</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit Rate^a</td>
<td>False Alarm^b Rate</td>
<td>Hit Rate^a</td>
</tr>
<tr>
<td>Immediate Rating</td>
<td></td>
<td></td>
<td>False Alarm^b Rate</td>
</tr>
<tr>
<td>Positive Cue</td>
<td>.734 (.sd=.14)</td>
<td>.253 (.sd=.13)</td>
<td>.627 (.sd=.23)</td>
</tr>
<tr>
<td>(n = 13)</td>
<td></td>
<td></td>
<td>(.sd=.11)</td>
</tr>
<tr>
<td>Negative Cue</td>
<td>.495 (.sd=.14)</td>
<td>.099 (.sd=.10)</td>
<td>.706 (.sd=.16)</td>
</tr>
<tr>
<td>(n = 16)</td>
<td></td>
<td></td>
<td>(.sd=.14)</td>
</tr>
</tbody>
</table>

Delayed Rating

|                      |                         |                           |                      |
| Positive Cue         | .764 (.sd=.12)          | .325 (.sd=.19)            | .637 (.sd=.22)       |
| (n = 15)             |                         |                           | (.sd=.15)            |
| Negative Cue         | .647 (.sd=.19)          | .203 (.sd=.10)            | .763 (.sd=.17)       |
| (n = 15)             |                         |                           | (.sd=.14)            |

a. Hit rate ranges from 0 (responding "no" to all old behaviors) to 1.0 (responding "yes" to all old behaviors).

b. False alarm rate ranges from 0 (responding "no" to all new behaviors) to 1.0 (responding "yes" to all new behaviors).
rating factor was a main effect indicating that subjects attributed more behaviors, overall, in the delayed rating condition (M = .70) than in the immediate rating condition (M = .64), F = 5.03, p < .03. Of greater importance for the primary hypothesis was the significant interaction of Performance Cues and Behavior Type, F(1,55) = 14.78, p < .0003. Follow-up tests of simple main effects revealed that when rating the occurrence of effective behaviors, subjects receiving positive cues reported more behaviors (M = .75,) than did subjects given negative cues (M = .57), F(1,55) = 19.25, p < .0001. However, the opposite pattern was evident in the ratings of ineffective behaviors. Here, subjects receiving positive cues attributed fewer behaviors to the group (M = .63) than did subjects who received negative performance cues (M = .73), F(1,55) = 4.12, p < .05. The nature of this interaction is perhaps most clearly depicted in Figure 1.

Insert Figure 1 about here

Interpretational Limits of Hit Rates

Despite having viewed the same film, subjects receiving positive performance cues attributed more effective and fewer ineffective behaviors to the group. The opposite pattern emerged for subjects given negative performance cues. At this point, it is important to note that an analysis of hit rates yields limited information about the nature of the performance cue effect demonstrated in behavioral ratings. Specifically, hit rates provide a gauge of subjects' performance in correctly identifying behaviors depicted in the film as "old" behaviors. With the exception of extreme hit rate values (i.e., 0 or 1.00), hit rates provide no information regarding subjects' memory accuracy in distinguishing between "old" and "new" behaviors.
Figure 1. Mean Hit Rate as a Function of Performance Cues
(i.e., behaviors that were and were not depicted in the film, respectively). Furthermore, hit rates alone do not provide any insight as to possible underlying processes responsible for the biasing effect that performance cues exerted on recollections of work behaviors.

Introduction to Memory Sensitivity (Pr) and Response Bias (Br)

Two possible mediating processes were investigated in the present study; one was a memory bias, and the other a response bias. It is possible that performance cues operated by enabling subjects to more accurately recognize behaviors consistent with expectations induced by performance cues -- a memory bias. Alternatively, it is possible that performance cues influenced raters' decision criteria in such a way that they were more likely to assign behaviors to the group that were consistent with performance cues (i.e., effective behaviors when given positive performance cues, and vice versa), regardless of whether the behaviors were actually observed -- a probabilistic response bias. Measures of memory sensitivity and response bias outlined by Snodgrass and Corwin (1988) were computed to identify the mediating process(es) responsible for the bias occurring in behavioral ratings.

There are several indices of memory sensitivity and response bias; however, these indices have been shown to differ in the degree to which they are theoretically uncorrelated (Snodgrass & Corwin, 1988). This theoretical independence allows one to detect differences in subjects' response bias equally well across varying degrees of memory sensitivity. In practice, the two measures can be correlated. The important point is that the theory underlying these measures of memory sensitivity and response bias does not dictate that they be correlated. This issue of correlated measures is of critical importance to the primary goals of this research. If the indices
employed are theoretically correlated, then it impossible to determine whether one process operates to the exclusion of the other, or to assess the relative contributions of the two processes if both are shown to be operating in a given situation. By systematically manipulating subjects’ memory and decision criteria, Snodgrass and Corwin concluded that one set of indices, in particular, were theoretically uncorrelated and were, therefore, suitable for research aimed at investigating mediating processes. These measures are $Pr$ (memory sensitivity) and $Br$ (response bias).

$Pr$ ranges from -1 indicating no memory to +1 indicating perfect memory and is computed using the following formula:

$$\text{Hit Rate} - \text{False Alarm Rate}$$

By requiring subjects to distinguish between behaviors that did and did not occur, $Pr$ provides a gauge of subjects’ ability to accurately attribute behavior to the work group.

$Br$ ranges from 0 to 1 with .5 representing a neutral decision criterion (i.e., no bias). Values greater than .5 represent a liberal decision criterion (i.e., a bias to say a behavior did occur), whereas values less than .5 represent a conservative decision criterion (i.e., a bias to say a behavior did not occur). $Br$ is computed using the following formula:

$$\frac{\text{False Alarm Rate}}{1 - (\text{Hit Rate} - \text{False Alarm Rate})}$$

To compute $Pr$ and $Br$, the behavioral rating instrument was first reduced to a yes/no recognition memory task. This was done by treating both effective and ineffective behaviors depicted in the film and rated a 4, 5, or 6 (indicating some degree of confidence) as “hits,” while effective and ineffective behaviors not depicted in the film but mistakenly rated as a 4, 5,
or 6 were treated as "false alarms." The resulting hit and false alarm rates were then transformed into measures of memory sensitivity (Pr) and response bias (Br).

**Memory Sensitivity (Pr)**

A 2 (Performance Cues) X 2 (Time of Rating) X 2 (Behavior Type) repeated measures analysis of variance of memory sensitivity (Pr) scores revealed no significant effects (p's > .30), indicating that a systematic memory bias was not responsible for the bias occurring in subjects' attributions of both effective and ineffective behaviors. Mean Pr ratings are presented in Table 2.

```
Insert Table 2 about here
```

**Response Bias (Br).**

The same 2 X 2 X 2 repeated measures design was used to investigate a systematic response bias (Br) as the underlying process mediating the effect of performance cues on subjects' behavioral ratings. If Br mediates the effect of performance cues on behavioral ratings of the work group, similar effects should emerge from the analysis of Br and hit rates. Examination of the results did, in fact, reveal a nearly identical pattern of results, with subjects adopting a more conservative decision criterion in the immediate, $M = .40$, than in the delayed rating condition, $M = .50$, $F(1,55) = 7.05$, $p < .01$. Of greater interest was the significant interaction of performance cues and behavior type, $F(1,55) = 16.12$, $p < .0002$, indicating that subjects' decision criteria for judging the occurrence of the same effective or ineffective behavior varied systematically depending on the performance cues they received before observing the group's performance.
Table 2

Mean Ratings of Memory Sensitivity (Pr), and Response Bias (Br)

<table>
<thead>
<tr>
<th></th>
<th>Effective Work Behavior</th>
<th>Ineffective Work Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Memory&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Response&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Bias</td>
</tr>
<tr>
<td>Immediate Rating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Cue</td>
<td>.481</td>
<td>.506</td>
</tr>
<tr>
<td>(n = 13)</td>
<td>(sd=.14)</td>
<td>(sd=.26)</td>
</tr>
<tr>
<td>Negative Cue</td>
<td>.396</td>
<td>.165</td>
</tr>
<tr>
<td>(n = 16)</td>
<td>(sd=.16)</td>
<td>(sd=.10)</td>
</tr>
<tr>
<td>Delayed Rating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Cue</td>
<td>.439</td>
<td>.548</td>
</tr>
<tr>
<td>(n = 15)</td>
<td>(sd=.18)</td>
<td>(sd=.23)</td>
</tr>
<tr>
<td>Negative Cue</td>
<td>.444</td>
<td>.400</td>
</tr>
<tr>
<td>(n = 15)</td>
<td>(sd=.20)</td>
<td>(sd=.21)</td>
</tr>
</tbody>
</table>

a. Pr values range from -1.0 (indicating no memory) to +1.0 (indicating perfect memory).
b. Br values range from 0 to +1.0:
   Br>.50 indicates a liberal decision criterion (bias to say "old")
   Br=.50 indicates a neutral decision criterion (no bias)
   Br<.50 indicates a conservative decision criterion (bias to say "new")
Follow-up tests of simple main effects were conducted to gain more information about this significant interaction of performance cues and behavior type. These analyses revealed that when rating the occurrence of effective behaviors, subjects receiving positive cues adopted a more liberal decision criterion than did subjects given negative cues (means equal to .53 and .28, respectively; F[1,55] = 19.22, p < .0001). This pattern of responding was reversed in ratings of ineffective behaviors such that subjects receiving positive cues adopted a more conservative decision criterion than did subjects given negative performance cues (means equal to .42 and .55, respectively; F[1,55] = 5.10, p < .03.). The nature of this interaction is more clearly depicted in Figure 2.

-----------------------------
Insert Figure 2 about here
-----------------------------

Further Investigation of Br as Mediator of the Performance Cue Effect

The parallel results from the analysis of hit rates and response bias (Br) provided evidence that the effects of performance cues on ratings of work group behavior were mediated by Br. That is, performance cues biased behavioral ratings by altering subjects’ decision criteria such that subjects adopted a more liberal decision criterion when judging the occurrence of behavior consistent with administered performance cues. What is not clear, at this point, is the extent to which Br mediates the effect of performance cues. A method outlined by Baron and Kenny (1986) was followed separately for effective and ineffective behaviors to provide a more appropriate test of the hypothesis that Br mediates the effect of performance
Figure 2. Mean Decision Criterion (Br) as a Function of Performance Cues
cues on ratings of a work group's behavior. Baron and Kenny argue that
... an ANOVA provides a limited test of a mediational hypothesis.
To test for mediation, one should estimate the three following
regression equations: first, regressing the mediator [Br] on the
independent variable [performance cues]; second, regressing the
dependent variable [hit rates] on the independent variable
[performance cues]; and third, regressing the dependent variable [hit
rates] on both the independent variable [performance cues] and on
the mediator [Br]. (p. 1177)

Effective Behaviors. The regression equations outlined by Baron and
Kenny represent three distinct components of the test of the mediational
hypothesis under consideration. The first equation tests the potential for Br
to mediate the effect of performance cues on ratings of effective work group
behavior as measured using hit rates. Obviously, an independent variable
must influence the proposed mediator for a mediational model to be a
tenable one. The results from this regression confirmed that the regression
of Br on performance cues was, in fact, significant, \( p < .0001 \). The second
equation tests the linkage of performance cues and hit rates. This linkage
serves two purposes in Baron and Kenny's procedure: first, confirmation
that performance cues significantly impact hit rates -- a prerequisite for
investigating a mediational hypothesis; and second, an indication of the
"upper limit" of the influence of performance cues. As expected, the
regression of hit rates on performance cues was significant, \( p < .0001 \). The
linkage revealed in this regression serves as a benchmark in the sense that
the influence of performance cues on hit rates must be less in the third
equation (i.e., controlling for Br) than in the second equation to provide
support for a mediational hypothesis. If the influence of performance cues
is reduced but still significant after controlling for $Br$ then "partial" mediation is indicated. Partial mediation is interpreted to mean that performance cues influence hit rates directly and also indirectly via $Br$. Alternatively, "complete" mediation is present when, after controlling for $Br$, the influence of performance cues is reduced to a level of nonsignificance. Complete mediation can be interpreted to mean that performance cues exert only indirect influence (i.e., via $Br$) on hit rates. "Complete" mediation was indicated by the results of the third regression equation. Specifically, the effects of performance feedback were not significant after controlling for $Br$, $p > .20$.

The results following from the methods prescribed by Baron and Kenny (1986) confirmed those found earlier using an analysis of variance framework. Based on the results from the regression equations, the model depicted in Figure 3 was constructed.

Insert Figure 3 about here

This model represents a type of causal model that is referred to as an "overidentified" model. This "overidentified" model is conceptually synonymous with the "complete" mediational model described earlier. However, an "overidentified" model comes out of a different approach (Pedhazur, 1982) to identifying the nature of the causal relationships among a set of variable and can, therefore, provide additional and independent support for the notion that performance cues influence reporting of work group behaviors (operationalized here as hit rates) only indirectly by altering subjects' decision criteria. To test an "overidentified" model, a goodness-of-fit index ($Q$) was calculated. This index ranges in
Figure 3. Meditational Model for Ratings of Effective Behavior.
The path coefficients are regression beta-weights which, in the case of an overidentified model with one proposed mediator, are identical to the bivariate correlations (Pedhazur, 1982, p. 598).

* p < .05, ** p < .01
value from 0 to 1, reflecting the proportion of variance the model explains relative to the variance it could potentially explain. A value of 0 reflects a total lack of fit between the proposed model and the data whereas a value of 1 represents a perfect fit of the model to the data. The Q statistic obtained for the model depicted in Figure 3 was .97 and indicated an extremely good fit of the proposed "overidentified" model, $W = 1.65$, $p > .20$.

Ineffective Behaviors. The same analytical strategy was followed to further investigate the nature of the relationship among performance cues, $Br$, and hit rates for ineffective behaviors as in the earlier section describing the analysis of effective behaviors. Consistent with the results for effective behaviors, the separate regressions of hit rates and $Br$ on performance cues were each significant, $p < .05$ and $p < .03$, respectively. When performance cues and $Br$ were entered simultaneously into the equation predicting hit rates, the independent contribution of performance cues was reduced to a level of nonsignificance, $p > .80$, providing evidence for a "complete" mediational model. The overidentified model depicted in Figure 4 was tested by calculating the goodness-of-fit index ($Q$).

-----------------------------------------------
Insert Figure 4 about here
-----------------------------------------------

The resulting $Q$ value was equal to .99 and revealed an extremely good fit of the proposed model to the data, $W = .03$, $p > .80$.

Results from the separate analyses of effective and ineffective behaviors yielded an essentially identical pattern of results. In both sets of analyses, performance cues were found to bias the recollection of the work group's behavior (hit rates) only indirectly by causing subjects to alter their decision criteria ($Br$). Specifically, subjects who received positive
Figure 4. Mediation Model for Ratings of Ineffective Behavior.
The path coefficients are regression beta-weights which, in the case of an overidentified model with one proposed mediator, are identical to the bivariate correlations (Pedhazur, 1982, p. 598).
* p < .05, ** p < .01
performance cues adopted a more liberal decision criterion when judging the occurrence of effective behaviors and a more conservative decision criterion when asked to judge the occurrence of ineffective behaviors. The opposite pattern of responding was demonstrated by subjects given negative performance cues. In this condition, subjects given negative performance cues adopted a more conservative decision criterion when judging the occurrence of effective behaviors and a more liberal decision criterion when judging the occurrence of ineffective behaviors.

**Evaluative Ratings of Group Process**

Results from the analysis of hit rates, memory sensitivity (Pr), and response bias (Br) clearly demonstrate the nature of the bias that performance cues introduce into ratings of work behavior. The second goal of this study was to identify the types of rating formats that might be immune to the biasing influence of performance cues. A second step towards this goal was to first demonstrate that the influence of performance cues extend to evaluative ratings of group process. To accomplish this, a 2 (Performance Cues) X 2 (Time of Rating) analysis of variance was conducted on a composite of the seven items composing the evaluative rating instrument. The only significant effect to emerge from this analysis was a main effect for performance cues with subjects given positive performance cues evaluating the group's process more favorably than subjects given negative performance cues (means = 7.62 and 5.29, respectively; $F[1,58] = 63.38, p < .0001$). This result indicates that, like recollections of work behaviors, evaluations of group process are biased by performance cues administered to subjects prior to their observing the group. As was the case with the ratings of both effective and ineffective behaviors, the effects of performance cues were no more pronounced in the
delayed rating condition. Mean ratings of group process are included in Table 3.

-------------------------------
Insert Table 3 about here
-------------------------------

**Outcome Ratings**

Analyses of both the behavioral and evaluative ratings indicate that the effects of performance cues were not limited to a single rating format. Anticipating these results based on the research conducted by Martell and Guzzo (in press), objective group outcomes were generated and incorporated into a three item rating scale. It was predicted that recollections of specific group outcomes would be unaffected by performance cues. Separate analyses conducted in a 2 (Performance Cues) X 2 (Time of Rating) framework revealed no significant effects with all p values > .30. These results suggest that subjects' recollection of specific group outcomes (i.e., Outcome 1 - number of group members to successfully cross the pool, Outcome 2 - number of group members to fall into the pool, and Outcome 3 - number of times the box was dropped) was not influenced by performance cues or time of rating. Mean ratings for the three items are included in Table 3. Further investigation of group outcomes was conducted by testing the differences between percentage of correct responses in the positive versus negative cue conditions. Tests of these differences in the positive versus negative performance cue conditions revealed no significant differences (all p values > .30) for Outcome 1 (64% v. 71%), Outcome 2 (89% v. 87%), or Outcome 3 (100% v. 97%). Finally, inspection of incorrect responses produced no systematic patterns. For
Table 3

Mean Ratings of Group Process and Recollections of Group Outcomes

<table>
<thead>
<tr>
<th></th>
<th>Group a Process</th>
<th>Group b Outcome 1</th>
<th>Group c Outcome 2</th>
<th>Group d Outcome 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immediate Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Cue (n = 13)</td>
<td>7.63 (sd=1.16)</td>
<td>3.77 (sd=.44)</td>
<td>1.92 (sd=.28)</td>
<td>0.00 (sd=.00)</td>
</tr>
<tr>
<td>Negative Cue (n = 16)</td>
<td>5.25 (sd=1.35)</td>
<td>3.56 (sd=.51)</td>
<td>2.00 (sd=.00)</td>
<td>0.00 (sd=.00)</td>
</tr>
<tr>
<td><strong>Delayed Rating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Cue (n = 15)</td>
<td>7.61 (sd=.97)</td>
<td>3.73 (sd=.59)</td>
<td>1.87 (sd=.35)</td>
<td>0.00 (sd=.00)</td>
</tr>
<tr>
<td>Negative Cue (n = 15)</td>
<td>5.33 (sd=.93)</td>
<td>3.80 (sd=.41)</td>
<td>1.87 (sd=.52)</td>
<td>0.13 (sd=.52)</td>
</tr>
</tbody>
</table>

a. the higher the mean, the more favorable the rating
b. number of group members to cross the pool (correct = 4), possible values range from 0 to 5
c. number of group members to fall in the pool (correct = 2), possible values range from 0 to 5
d. number of times the box was dropped (correct = 0), possible values range from 0 to 5
example, subjects receiving positive cues were just as likely to underestimate the number of group members to successfully cross the pool as subjects receiving negative cues. Thus, as predicted, ratings of specific group outcomes were immune to the biasing effects of performance cues.
Discussion

Consistent with previous research, performance cues systematically biased individuals' ratings of group process and work group behavior such that individuals provided with positive cues rated the group's process more favorably than did subjects given negative cues. Furthermore, subjects given positive cues attributed more effective and fewer ineffective behaviors to an identically performing work group than did subjects who were given negative cues. Investigation of two potential mediating processes identified a systematic response bias and not a memory bias as the underlying process responsible for the biasing effects of performance cues on behavioral ratings of a work group. The failure to find any evidence of a memory bias occurred despite the fact that performance cues were available to subjects at time of encoding and suggests that interventions designed to improve raters' memory may not prove useful in eliminating performance cue bias in work behavior ratings.

The bias introduced by performance cues is undesirable for two reasons. One, judgments influenced by performance cues reflect a systematically distorted view of the way individuals and groups function in work organizations. Theories that are based on such judgments will reflect individuals' naive theories rather than actual organizational functioning. Two, from a practical viewpoint, these judgments will provide inadequate and, perhaps, misleading appraisals of an individual's or group's performance. Beyond the more obvious impact on pay, promotion, and termination decisions there is the impact on employee training and development. It is likely that even effective employees evidence some deficit(s) in their work performance. The results from this study and others suggest that raters will fail to report these deficits when appraising
an "effective" individual's performance. This means that effective employees will remain oblivious to deficits in their performance thereby greatly reducing the probability that they (or the organization) will work toward reducing these deficits.

There are two general strategies that might prove successful in eliminating the response bias invoked by knowledge of performance: training raters to recognize and avoid bias in their ratings and motivating raters to make accurate and unbiased ratings. Within each major category, there are a number of variations that have at their core a common philosophy regarding the nature of the bias.

A rater training program developed by Bernardin and Pence (1980) provides a good example of the type of program that might prove useful in attenuating the biasing effects of performance cues. The authors empirically tested the ratings produced by their program (RAT) and another more traditional rating program against true scores and found that the RAT ratings were more accurate than those produced by the traditional program. The RAT program is based on the underlying premise that "rater training should concentrate on enhancing the accuracy of ratings through discussion of the multidimensionality of work performance, the importance of fair and critical evaluation, and the development of stereotypes of effective and ineffective workers" (p. 61). Adapting the RAT program to the problem at hand would educate raters about the multidimensionality of the way in which groups work together to accomplish a task by stressing the fact that a group demonstrating, for example, a high level of communication is not necessarily at the same level in terms of cooperation or effort. The multidimensionality of individual and group performance is precisely the type of information that is ignored when
raters allow their naive theories or category prototypes to dictate their ratings. Stressing the importance of fair, unbiased, and accurate ratings (the second component), would serve to remind raters of their obligation to perform their duty to the best of their ability. The third component of the RAT program might prove especially useful in combating the performance cue effect. Having raters generate and define dimensions of performance for a particular task and then discuss examples of high, medium, and low behaviors for each dimension would help to develop and make explicit a shared set of stereotypes of effective, average, and ineffective workers. This is an important goal of a comprehensive performance appraisal system recently developed and implemented by J. M. Feldman (personal communication, February 15, 1991).

An additional component could be added to Bernardin and Pence’s training to further reduce the bias introduced by knowledge of performance. In this component, raters would systematically (dimension-by-dimension) compare known individuals or groups at varying levels of effectiveness to the previously-developed stereotypes. This exercise would serve to conceptually link the first and third components of Bernardin and Pence’s training and make more salient and concrete the point that performance is multidimensional with the bulk of individuals and groups showing some degree of variability across dimensions.

It might be possible to approach the problem of performance cues or expectations from a motivational perspective. Tetlock and his colleagues have conducted a rigorous and systematic investigation of “accountability” - - the need to explain and justify one’s judgments to others -- as a factor with the potential to motivate individuals to “put more effort or cognitive work into making judgments and decisions” (Tetlock, 1983; p. 74). The results
from this earliest study suggest that individuals will engage in more rigorous information processing when expecting to justify their judgments to individuals with unknown views. From the work conducted by Tetlock and colleagues (e.g., Tetlock & Kim, 1987), it seems clear that accountability has the potential for improving the accuracy of performance ratings in instances in which raters hold prior expectations of individual or group performance in a task if administered correctly. Specifically, raters should not be forewarned regarding the opinions of those to whom they expect to justify their opinions. This would be tantamount to replacing their preconceived notions with new (not necessarily improved) ones, one of the findings from Tetlock's original study (1983).

Another alternative to the problem at hand is to discontinue the use of bias-prone rating formats altogether in favor of those that produce accurate and error-free ratings of work performance. In this study, ratings of objective group outcomes were the only type of rating that proved immune to the biasing effects of performance cues. The notion of appraising work performance based on measures of objective task outcomes is not original to this research. In fact, this was one of the first rating formats to enjoy wide-spread use in organizations. However, ratings based on objective outcomes are enjoying a resurgence in popularity in the form of Management By Objectives (MBO) -- a participative strategy used not only to evaluate employee performance but to motivate employees by establishing clear, mutually agreed upon objectives at the outset of an appraisal period.

One advantage of basing performance appraisal on objective outcomes is their tendency to produce highly accurate ratings. Also, if an employee and manager can agree on reasonable outcomes, the employee's role ambiguity is reduced and the manager is clear as to the appropriate
information that he/she should rely upon in evaluating the employee's performance. However, there are at least two disadvantages to confining raters to relying solely on objective outcomes in forming their judgments. First, the accurate reporting of work behavior has the potential to fairly represent a group's performance in a task. This is true of objective task outcomes only to the degree that achieving these outcomes is potentially within the group's control. Many times, environmental factors facilitate or hinder groups to the extent that outcomes provide poor indicators of a group's actual performance in a task. Second, many jobs in organizations are highly "cognitive" in the sense that much of what an employee does is "hidden" from observation. In these types of jobs, the outcomes that are observable are very likely open to interpretation. The outcomes extracted from the videotaped group for use in this study were directly observable and offered no latitude in interpretation. For example, 98% of subjects correctly reported that the box was never dropped into the pool. It seems doubtful that this level of accuracy and agreement would be achieved in a situation in which raters were required to recollect "How many times did the manager reprimand a subordinate in front of other employees during the last quarter?"

Nonetheless, the results from this study suggest that objective outcome ratings hold promise for producing performance ratings with an exceptional degree of accuracy. It is this author's opinion that they should be utilized to the extent that they can meet the following requirements:

1) fairly represent an individual's or group's performance (i.e., reflect attributes and characteristics of the employee rather than the work environment),

2) be generated for a wide range of jobs, and
3) provide enough information to serve a number of potential functions of performance appraisal systems (e.g., coaching, motivating, training, etc.).

One limitation of the present study should be noted in light of its potential implications for the development of interventions designed to reduce the biasing effects of performance information on work behavior ratings. The behavioral rating task employed in the present study was essentially a recognition memory test. No measure of recall was collected. From Alba and Hasher's review of the research in schema theory (1983), it is clear that recall of information can be influenced by the preconceived notions, expectations and perspectives that individuals bring to a particular situation. A possible explanation for the failure to provide evidence for a memory bias in recognition tests of memory as opposed to recall is the fact that in recognition, the researcher provides all subjects with the same cues by virtue of the fact that each subject completes an identical behavioral rating instrument. These "identical cues" could potentially lead subjects across manipulation conditions to recollect enough of the same behaviors as to mask any memory bias that might have been detected had subjects been allowed to "provide their own cues" as is the case in recall measures of memory (D. LeCompte, personal communication, March 6, 1991).

This author’s recommendation for future research is twofold. First, a recall test of memory should be utilized in order to definitively rule out the possibility that a memory bias mediates the effect of performance cues on recollections of work behavior. One of the long-term objectives of research in performance appraisal is to improve the quality of performance appraisal. To achieve this objective, future research should focus on pinpointing the mediating processes responsible for the bias that continues
to plague performance ratings in both laboratory and field settings. Without precise knowledge of these mediating processes and their "method of operation," industrial/organizational psychologists will continue to design and develop interventions. However, these interventions will continue to enjoy limited success in organizations. The type of research conducted here is one step in the right direction.
References


Appendix A

Introduction to Videotape

This is a study investigating how people perceive and evaluate the way a group functions. The group you are about to observe is one of many groups participating in an actual contest. The contest required participating groups to get a box and as many people as possible across a pool of water in 14 minutes. As you might imagine, some of the groups performed better or worse than others in the contest.

Each group had a limited number of tools to use in accomplishing its task. These included:

1) Three boards; 10, 8, and 6 feet in length
2) A 25-foot rope
3) Gloves for all members
4) A sturdy support bar 15 feet above the middle of the pool with a ten foot rope hanging from the center of the bar.

The group you are about to observe consists of five members; one dressed in green, the others are numbered 3, 28, 38, and 40. The other people around the edge of the pool are safety monitors or nonparticipating observers. When this group was filmed, other competing groups were working on the same task. You might be able to hear these other groups in the background.

Your task in this study is to observe the group as they attempt to accomplish the goal of transporting the box and group members across the pool. Watch the videotape carefully. Observe what is done, who does it, and when it is done. Do not take notes. The video is 14 minutes long.

You will be asked to respond to questions about the group's activities. Some of the participants in this study will be asked to respond to these questions immediately following the video. Other participants will be asked to respond to the questions one week from now. After watching the video, you will either be asked to respond to the questions immediately or in one week.

Do you have any questions? Is your task clear?
Appendix B
Behavioral Rating Instrument

On the following pages you will find a list of specific behaviors. Using the scale below, please rate whether the behavior(s) occurred in the group that you observed.

The ratings are to be made by placing the number (1, 2, 3, 4, 5, or 6) which most closely represents your opinion on the line next to each behavior:

1. I am very certain the behavior(s) did not occur.
2. I am fairly certain the behavior(s) did not occur.
3. I am undecided, but think that the behavior(s) did not occur.
4. I am undecided, but think that the behavior(s) did occur.
5. I am fairly certain the behavior(s) did occur.
6. I am very certain the behavior(s) did occur.

Please respond to each question. The success of this research requires that you answer every question. Do not leave any blanks.

You may begin by turning to the next page.
1. I am very certain the behavior(s) did not occur.
2. I am fairly certain the behavior(s) did not occur.
3. I am undecided, but think that the behavior(s) did not occur.
4. I am undecided, but think that the behavior(s) did occur.
5. I am fairly certain the behavior(s) did occur.
6. I am very certain the behavior(s) did occur.

1. The group members often talked amongst themselves about hustling and performing the task well.

2. At the beginning of the task, one group member wanted to be sure that everyone knew what to do and therefore asked the group out loud: "Do we know what we're doing?" Most of the group members simply ignored his question.

3. When a group member voiced a suggestion, other members responded positively, saying "that's a good idea."

4. One group member shoved another out of the way when reaching for the box.

5. One group member yelled in a loud, unpleasant voice at another that he was balancing on the makeshift bridge in the wrong way.

6. Group members sometimes asked each other how they were doing to make sure that everyone was feeling okay about the group.

7. When a group member was having difficulty holding down the board upon which another group member was going to walk, other members quickly came to his assistance.

8. Several group members stated that they enjoyed working in the group.

9. While attempting to cross the pool, a group member carefully explained to the others how to best position their feet on the bridge to avoid falling into the water.
Appendix B (cont.)

1. I am very certain the behavior(s) did not occur.
2. I am fairly certain the behavior(s) did not occur.
3. I am undecided, but think that the behavior(s) did not occur.
4. I am undecided, but think that the behavior(s) did occur.
5. I am fairly certain the behavior(s) did occur.
6. I am very certain the behavior(s) did occur.

10. When a group member was slow in tying the rope necessary to build the bridge another member yelled at him in an unfriendly tone of voice, telling him to go faster.

11. The group determined who would be its leader by holding a quick election.

12. Several times group members complained that others in the group were not trying hard enough.

13. The group loudly encouraged each person attempting to cross the pool.

14. Often during the exercise several group members stayed to the side and did not actively contribute unless specifically asked to do so.

15. When something went wrong while constructing the bridge, one of the group members was sometimes made the scapegoat.

16. The group discussed and analyzed suggested tactics for building the bridge and crossing the water.

17. Several group members worked together to reach the rope suspended over the center of the pool.

18. Several group members each took turns working on the more difficult work.
Appendix B (cont.)

1. I am very certain the behavior(s) did not occur.
2. I am fairly certain the behavior(s) did not occur.
3. I am undecided, but think that the behavior(s) did not occur.
4. I am undecided, but think that the behavior(s) did occur.
5. I am fairly certain the behavior(s) did occur.
6. I am very certain the behavior(s) did occur.

19. At one point during the discussion the group leader made everyone stop talking in order to give one group member a chance to speak.

20. At times the group members stood by the pool and prepared to protect and assist in the event of a fall from the bridge.

21. Group members expressed the belief that they probably had one of the best groups.

22. One group member complained out loud that they were wasting a lot of time trying to tie a knot in the rope.

23. At times, two of the group members monopolized the work while the others passively watched.

24. When members of the group reached the rope over the center of the pool by walking out on a board over the water, another member on the wall steadied them.

25. The group polled its members to see who knew how to tie knots, who had the best balance, and who was the strongest.

26. Instructions given by one group member to another were often given incompletely.
Appendix B (cont.)

1. I am very certain the behavior(s) did not occur.
2. I am fairly certain the behavior(s) did not occur.
3. I am undecided, but think that the behavior(s) did not occur.
4. I am undecided, but think that the behavior(s) did occur.
5. I am fairly certain the behavior(s) did occur.
6. I am very certain the behavior(s) did occur.

27. Group members sometimes ignored the need to work quickly on their task.

28. The group member standing alone on the opposite side of the pool offered no ideas or suggestions from his perspective.

29. One group member called another “stupid” for not agreeing with his suggestion.

30. Group members repeatedly expressed that they were working together so well that they would have no problem finishing within the allotted time.

31. When one of the members had difficulty attempting to cross the bridge, several of his teammates encouraged him with cheers.

32. When a member had difficulty tying a knot, another member replied “that’s ok” and encouraged him to finish.

33. At the beginning of the exercise, each member volunteered to perform a specific task in order to build the bridge in an orderly fashion.

34. Frequently, members praised each other’s performance.

35. A member who was moving the plank across the water offered a suggestion which everyone ignored and members told him to hurry.
Appendix B (cont.)

1. I am very certain the behavior(s) did not occur.
2. I am fairly certain the behavior(s) did not occur.
3. I am undecided, but think that the behavior(s) did not occur.
4. I am undecided, but think that the behavior(s) did occur.
5. I am fairly certain the behavior(s) did occur.
6. I am very certain the behavior(s) did occur.

36. One member yelled to another crossing the bridge "Don't spend any time out there" and the bridge-crosser replied in an annoyed tone of voice "I can get across."

37. While one of the men was carefully crossing the bridge, one member yelled at him to hurry up and started shaking the bridge to make him move faster.

38. When a member appeared to be working too slowly in moving the plank across the water, another member pushed him out of the way and grabbed the board.

39. The group paused during its work to discuss its progress.

40. When it became clear that there was little time left, the group gave up and stopped working.
Appendix C

Evaluative Rating Instrument

For the following questions, circle the number which best describes your view of the group. Please be sure to answer all the questions.

1. How would you rate the quality of communication in this group?
   Very Low 1 2 3 4 5 6 7 8 9 10 Very High

2. How would you rate the effort put into the task by group members?
   Very Low 1 2 3 4 5 6 7 8 9 10 Very High

3. To what extent were the actions of group members coordinated?
   Not at All 1 2 3 4 5 6 7 8 9 10 To a great Extent

4. How would you rate the team spirit of the group?
   Very Low 1 2 3 4 5 6 7 8 9 10 Very High

5. How would you rate the quantity of communication in this group?
   Very Low 1 2 3 4 5 6 7 8 9 10 Very High

6. To what extent did people in the group appear to enjoy working together?
   Not at All 1 2 3 4 5 6 7 8 9 10 To a great Extent

7. To what extent was the group committed to accomplishing its task?
   Not at All 1 2 3 4 5 6 7 8 9 10 To a great Extent
Appendix D
Objective Outcome Rating Instrument

Here are some multiple choice questions asking you about the group you observed. Answer each question by simply putting an X next to the appropriate response. Please be sure to answer every question.

1. Of the five people in the group, how many successfully crossed over the pool before the time elapsed?
   5
   4
   3
   2
   1
   0

2. How many times did a group member accidentally fall into the water while attempting to cross the pool?
   5
   4
   3
   2
   1
   0

3. How many times was the box dropped into the water?
   5
   4
   3
   2
   1
   0