This is to certify that the

thesis entitled

THE EFFECTS OF INFORMATION CONSISTENCY, TIMING OF CATEGORY SALIENCE, AND MOTIVATION ON THE RECALL OF RATEE BEHAVIOR

presented by

Miguel Angel Quinones

has been accepted towards fulfillment of the requirements for

M.A. degree in Psychology

Major professor

Date 11/26/90

THE EFFECTS OF INFORMATION CONSISTENCY, TIMING OF CATEGORY SALIENCE, AND MOTIVATION ON THE RECALL OF RATEE BEHAVIOR

By

Miguel Angel Quiñones

A THESIS

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

MASTER OF ARTS

Department of Psychology

1990

ABSTRACT

THE EFFECTS OF INFORMATION CONSISTENCY, TIMING OF CATEGORY SALIENCE, AND MOTIVATION ON THE RECALL OF RATEE BEHAVIOR

By

Miguel Angel Quiñones

This study explored the effects of a performance category (i.e. good performer) on recall in a performance appraisal situation. More specifically, the study investigated the effects of providing a category prior to and after the observation of a ratee on the recall of ratee behavior. The category used was that of good secretary. In addition, the effects of motivation (accountability) as a way of eliminating the biasing effects of providing a category were also investigated. Undergraduate psychology students (n=132) participating in groups of 2-7 (mode=6) were presented with a bogus performance appraisal form depicting a university secretary as being a good performer prior to or after viewing a videotape of this secretary containing instances of good and poor performance. Motivation was manipulated by telling half of the participants that they would have to justify their responses on a behavioral checklist to the rest of the group members. The hypotheses relating to accuracy of category consistent and inconsistent behaviors were not supported. Subjects were more accurate in recalling instances of poor behaviors. The failure to find the predicted effects and directions for future research are discussed.

ACKNOWLEDGEMENTS

I guess the first person I need to thank is myself for doing all the work and the typing required for finishing this project. However, I would not be myself if it wasn't for my beautiful wife Sherry. She has helped me become the person I am today. Without her love and constant support I would probably never have finished my thesis. My Chairperson, Dan Ilgen, has been instrumental in making this a quality project. Thanks a lot Dan for not letting me get away with mediocre work. My other committee members, Steve Kozlowski and J. Kevin Ford, have also provided me with invaluable guidance and support during my tenure here at Michigan State. I would also like to thank my colleagues, Stephen Gilliland and Jeffrey Schneider, for showing me that this was not an impossible task. They have provided excellent role models for me. Dawn Rosen deserves to be mentioned for helping me run the experiments. Finally I would like to thank my parents. The have always believed in me and my ability to succeed. This unconditional support has helped me overcome every obstacle I have encountered in my life, including this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Early research in performance appraisal focused on the rating instrument as a possible source of bias (i.e. halo, leniency, central tendency). This approach, as well as attempts to train raters to be more accurate, did little to reduce common rating errors (Landy & Farr, 1980). Researchers have since begun to look at performance appraisal as a complex and ongoing process in which the instrument and the rater are only a subset of the factors that can bias ratings (DeNisi, Cafferty, & Meglino, 1984). This more recent approach to performance appraisal has been labeled the process approach (Ilgen & Feldman, 1983).

A process perspective views the final rating as the result of an interaction between the organizational context within which the appraisal takes place, the rater's information processing system, and the behaviors of the ratee (Ilgen & Feldman, 1983). Inherent in this perspective is the recognition that a rater acquires information about a ratee over a period of time from different sources (DeNisi et al., 1984; Ilgen & Feldman, 1983). A primary goal of appraisal process research is to find key factors throughout the rating process that may bias the final rating. One such factor that has stimulated a number of research projects is the rater's information processing system.

It is suggested that raters process information about others who are to be evaluated by (1) attending to behavioral information about the ratee, (2) categorizing

this information in memory, (3) recalling it at the appropriate time, and (4) integrating the information for purposes of coming up with a final rating (DeNisi et al., 1984; Feldman, 1981). However, people are not perfect, unbiased perceivers of such information. Instead they rely, in part, on simplification heuristics that make categorization of incoming information easier (Cantor & Mischel, 1979; Fiske & Taylor, 1984). To the extent that incoming information is similar to a readily accessible category in memory, the information is more likely to be stored under that category (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Given the fact that people rely on categories to process information that they encounter, researchers have focused on studying the nature and the effects of categories on the information processing system (e.g. Cohen, 1981; Nathan & Lord, 1983).

Feldman (1981) noted that there are many situational factors that influence the salience of a particular category at any given time. Furthermore, research has shown that the use of categories influences the type of information to which people attend and recall (DeNisi et al., 1984; Srull & Wyer, 1989). Thus, if the performance appraisal process is ongoing and categories are made salient by a number of situational features, it is reasonable to conclude that these categories could be activated at any point in the appraisal process. However, most process models of performance appraisal view categories as influencing the type of information that is sought out and recalled after the category has been made salient (DeNisi et al., 1984). This condition will be referred to as the Filter Model.

In some rating situations, raters learn something about a ratee after he or she has observed that ratee for a period of time. That is to say, the rater is not presented with category information prior to the observation of performance. Rather, the

information about categories either evolves from the observations or is presented at some point after observations begin. Research has shown that under these conditions raters tend either to distort information previously encoded or to recall events that did not even occur (Srull & Wyer, 1989, Srull & Wyer, 1980). Other researchers have also noted differences in the kind of information that is recalled under conditions where category information follows exposure to relevant stimuli (Wyer, Bodenhausen, & Srull, 1984; Wyer & Srull, 1989). Information processing models describe the process involved when category information is received after observation and storage as one in which the activation of a category does not act as a filter of incoming information but rather as an organizer of information already available in memory (Wyer & Gordon, 1984; Wyer & Srull, 1989). This condition will be referred to as the Organizer Model.

In addition to the rater's information processing system, it has also been suggested that the rater's motivation and ability to provide accurate ratings will influence the performance appraisal process (DeCotiis & Petit, 1978). Ability and motivation are believed to interact to affect rating accuracy. Raters who are highly motivated to give accurate ratings may not be able to do so if they do not have the ability to do so and vice versa. Ability to give accurate ratings could be influenced by the fact that raters may have, in memory, biased information about ratees as a result of selective attention, encoding, or recall of information due to the presence of a category prior to or following observation. Thus, similar amounts of motivation could lead to differential accuracy in ratings due to the fact that biased information is available to raters as a result of categories being made accessible at different points in the performance appraisal process thus influencing the ability of the raters to give accurate ratings.

For example, a rater who believes that a particular ratee is usually a good performer may only pay attention to information that confirms this expectation (i.e. filter model). When the time comes to rate this particular individual, no amount of motivation will enable the rater to recall information that was never encoded in the first place such as instances of poor performance. However, if this rater learned about a ratee's past performance after he or she has had a chance to observe the ratee for a period of time (i.e. organizer model), the rater might bias the information that is recalled even though they have attended to and encoded all behavioral instances equally. In this case, high motivation at least has the possibility of overcoming the biasing effect of this performance category on the recall of the information since it is assumed that the information was encoded in an unbiased fashion.

This study will look at the effects of making a category salient prior to the presentation of behavioral information on the recall of this information and compare that effect to that observed when an identical category is made salient immediately after the presentation of behavioral information (i.e. will compare the effects of the filter model versus those of the organizer model). Furthermore, the influence of the type of information recalled on the ability of the rater to give accurate ratings will be explored under varying levels of rater motivation to accurately recall the observed behavioral information. It is expected that under similar motivational conditions, the rater's ability to give accurate ratings will be influenced by the time (before or after observation) at which a category was made salient.

A review of the literature concerning the effects of making a category accessible prior to and following observation will first be presented. Second, the results presented will be interpreted with respect to the specific problem of performance appraisals.

Third, a discussion of the effects of motivation and ability to give accurate ratings will be presented. Finally a framework for the current research will be presented along with a discussion of the specific hypotheses.

Filter Model

Many researchers have been interested in studying the effects of providing a category prior to observing a target person on information processing. The primary way that this effect has been studied has been to determine whether behaviors that are consistent or inconsistent with the category are better remembered (see Srull & Wyer, 1989). This situation has sometimes been referred to as impression testing because individuals use the behavioral information that is being presented about a target person to test the category or impression that they already possess about the person.

As has already been stated, impression testing occurs when a judgement about a person has already been made or an impression has already been formed and new information is presented about that person. Pre-existing judgments or impressions of people are believed to serve as anchors or standards for evaluating new information about a focal person by categorizing them as belonging to a particular group of people (i.e. poor performer); these anchors are sometimes called stereotypes or prototypes (Cantor & Mischel, 1979; Fiske & Taylor, 1984). Research has shown that such categories (stereotypes/ prototypes) affect subsequent processing of information about the focal person.

In a study by O'Sullivan and Durso (1984) people were given category labels about a person along with two pieces of information highly congruent with the category.

The participants in the research were then presented with biographical information that varied in its congruency with the initial stereotype label. The results showed that participants had higher recall for category consistent items. Other researchers have concluded that the effect of providing a category prior to the observation of a person is to increase recall for category consistent behaviors (Higgins & Bargh, 1987).

Other researchers have studied the filter effects but found different results. Hastie (1980) presented participants with a category about a focal person. The participants were then asked to read a series of descriptive statements about that person while forming an impression of him or her. The statements contained information that was both consistent and inconsistent with the category that was originally presented. The results indicated that participants had higher accuracy of recall for category inconsistent behaviors. The authors concluded that this effect was due to participants forming more associative bonds between the inconsistent items because they could not be associated with the general category.

To explain the inconsistencies between Hastie's (1980) data and the data of those who found consistent information better recalled, Higgins and Bargh (1987) pointed out that studies such as Hastie's (1980) ask the participants to form their own impression regardless of the category given to them prior to observation. This manipulation puts the subjects in a mode of forming an impression along with that of testing one. Support for the position that impression formation is a separate process comes from studies which ask subjects to either form an impression or remember as much as they can about a target person (Lichtenstien & Srull, 1987; Srull, 1981; Wyer & Gordon, 1982). These studies find that when subjects are asked to remember as much as they can, the impression formation effect disappears. Higgins and Bargh (1987)

concluded that participants in impression formation experiments are not relying on the original category presented to them prior to observation. Contrastingly different processes than these are at work in an impression testing paradigm which is the reason why they obtain different results.

Returning to the impression testing situation, there are several mechanisms that have been proposed to account for the filter effect of impression testing in which a recall advantage for consistent information is found. One proposed mechanism is that of discounting (Higgins & Bargh, 1987). Discounting occurs when people have a category and interpret inconsistent information in a way that is consistent with the category (Bodenhausen, 1988). For example individuals may attribute the inconsistent behaviors as being caused by factors outside of the person and thus not view these behaviors as being inconsistent with the category assigned to the person (Crocker, Hannah, & Weber, 1983). Thus the category acts as a filter which blocks out inconsistent behaviors and does not allow them to be processed into memory. Another possible mechanism which causes the filter effect is the way an individual "chunks" the behavioral information being presented to them into discrete behavioral events. Individuals may group information in a way that is consistent with their overall impression or category. This conclusion is evidenced by the lack of congruence in results found for studies using different modes of presenting the behavioral information.

All of the research that has been reviewed presents a rater with an extended amount of information about the person who is to be rated, evaluated, or simply described. There is some evidence that the mode through which the detailed information about the stimulus person is presented has an effect on the results of recall experiments. In particular, when the richer sources of information such as videotapes

are used and compared to less rich sources (such as paper and pencil descriptions of the same information) consistent information is remembered better (Cohen, 1981; Rothbart, Evans, & Fulero, 1979). Cohen (1981) argued that richer sources produce these effects because participants must reduce the cognitive load created by the variety of information. According to Cohen, they do this by chunking the behavioral episodes in a manner that favors consistent information.

In conclusion, it appears that providing a category prior to observation serves to affect recall by increasing accuracy for category consistent items. Several mechanisms for this effect have been proposed all of which are presumed to work by affecting the initial encoding of the information in the form of a filter. The effect has been most acute when rich formats such as a videotapes are used which allow the participants to reduce cognitive load by using the existing category.

Organizer Model

Up to now, the effects of a category have been discussed in terms of its effect on information gathering. However, categories can be made salient after the observation of information. For example, after working with an individual for a period of time, one might find out that a colleague was fired from his or her previous job for stealing. One might begin to recall instances in which the colleague acted in a suspicious manner. Thus, activating a category after observation affects the way an individual recalls previously stored information.

The effects of providing a categorization schema after the observation of behavior have been less conclusive than those based on categories provided prior to

observation. However, similar to the filter effects, the organizer effects have been studied by assessing their impact on recall of category consistent and inconsistent behaviors. Several studies have found consistent information better recalled, others inconsistent information, while still others have found no difference in recall advantage for either form of information. All studies have investigated these hypotheses by presenting participants with information about the ratee before the introduction of a categorization schema (Cohen, 1981; Pyszczynski, LaPrelle, & Greenberg, 1987; Snyder & Uranowitz, 1978). Some researchers have not explicitly stated that they were testing the organizer effects but also introduced the category after observation (Phillips & Lord, 1982). The latter research has produced results similar to the explicit tests (Phillips & Lord, 1982). There are, however, other studies which have failed to find the predicted effect for the organizer model (Belleza & Bower, 1981; Bodenhausen, 1988; Clark & Woll, 1981; Rothbart et al., 1979; Wyer, Srull, Gordon, & Hartwick, 1982). Before attempting to resolve the inconsistencies in results of the organizer effect studies, a brief summary of the most important ones will be presented. It should also be added that most studies use the organizer model as a control condition in studying the filter model. This is done in order to rule out the possibility that the category in the filter model serves as a recall cue. Thus if, a recall advantage for either consistent or inconsistent information is found in the filter model and not in the organizer model then the hypothesis that category effects are due to recall cues can be ruled out (e.g. Rothbart et al., 1979).

In a study by Rothbart et al. (1979), subjects were presented with 50 behaviors attributed to a fictitious person. The behaviors were coded as friendly (n=17), intelligent (n=17), unfriendly (n=3), nonintelligent (n=3), unrelated (n=10). Their

design was a 2 x 2 x 2 x 2 factorial design with 2 between-subjects and 2 within-subjects factors. The between-subjects factors were the type of behaviors the subjects were led to expect of the fictitious person (friendly vs. intelligent) and the time at which they received the categories of friendly or intelligent (before vs. after observing the behaviors). The within-subjects factors were type of behavior (friendly behaviors vs. intelligent behaviors) and type of instance (expectancy confirming vs. expectancy disconfirming). The behaviors were presented in the form of written descriptions on randomly ordered slides. The authors expected to find a difference in recall of confirming vs disconfirming behaviors as a result of providing the category after observation. Results showed no advantage in recall for expectancy inconsistent behaviors as was expected in the organizer model condition; the organizer model was not supported.

In another study, Bodenhausen (1988) examined the effects of a category label on recall and guilt judgments of a mock jury when presented with information about a defendant during a mock trial. In this study, the author was also testing the hypothesis of biased processing observed in the filter model and incorporated the organizer model in order to rule out the possibility of the organizer effect acting only as a recall cue. Results showed no difference in recall for category consistent or inconsistent behaviors in the organizer model condition.

Cohen (1981) presented subjects with behavioral information on videotape. The subjects in the study were led to believe that a woman depicted in the tape was either a librarian or a waitress. This information was provided to half of the participants prior to viewing the tape and to the other half after they had viewed the tape. Results showed a significantly higher accuracy in recall for consistent information in both conditions.

The author concluded that both processing as well as weak recall effect were responsible for the differences in accuracy observed in the results. Since no difference was found when the category label of librarian or waitress was presented before or after observation, participants were not only relying on this category to selectively process incoming information but also to aid them in recalling observed behaviors.

Similarly, in a study by Pyszczynski, LaPrelle, and Greenberg (1987), subjects were presented with behavioral information about a person either before or after they were allowed to read a general characterization (category) of that person. The behavioral incidents were presented in the form of a recorded audio message. The results showed a higher recall for category-congruent items in the condition where the category was introduced after the behavioral incidents were presented. There was no effect for the filter model. In their discussion, the authors hypothesized that the reason for an organizer effect was that the audio messages made it difficult for participants to form associations among behavioral incidents. Since the incidents were encoded independently, the category served as a retrieval cue for category-consistent items. This is in contrast to studies which use a written format. The written format allows subjects to memorize the items as a list. It was suggested that when participants are asked to recall the behaviors, they were merely recalling a memorized list and did not need to use of the category (Pyszczynski et al., 1987). This is a possible explanation as to why the failure of Rothbart et al. (1979) to find an organizer effect.

In a study by Phillips and Lord (1982), subjects were presented with behavioral information of a four-person problem-solving group on a videotape. After watching the tape, they were presented with a bogus performance category about the group leader (i.e. good leader) and were then asked to fill out a questionnaire containing effective

and ineffective leadership behaviors. The subjects were presented with a list of behaviors and asked to mark off those behaviors they recalled observing the leader of the group perform. A higher recall for consistent items was found. That is, subjects who were told that the group leader had performed well recalled more behaviors associated with good leadership than poor. The results also indicated a higher false alarm rate for consistent behaviors. This means that subjects reported having observed more consistent than inconsistent behaviors that were not present in the videotape.

The Phillips and Lord (1982) study suggests that the organizer effect occurs because participants use the category provided to them after observation and use it to interpret the information previously observed. In recalling this information, they "fill in" behavioral instances that are consistent with this category but were never observed. This is the reason why Phillips and Lord (1982) found a large false alarm rate for category consistent behaviors. This implies that one must take into account false alarm rates when developing an accuracy estimate due to people's tendency to distort their recall toward consistent items. When the distortion was taken into account in this study by correcting for false positives, inconsistent behaviors were more accurately recalled than consistent. Since other studies which found consistent information more accurately recalled than inconsistent did not correct for false alarms, their results (after correction) may be consistent with Phillips & Lord (1982). This possibility will be examined in the section entitled "Contradictions".

In conclusion, there are studies that have failed to find an organizer effect (eg. Bodenhausen, 1988) while other studies have found an organizer effect (e.g. Phillips & Lord, 1982). Of these, some have found higher accuracy for items consistent with the category given after observation (eg. Pyszczynski et al., 1987). Others have found higher

accuracy for category inconsistent behaviors (eg. Phillips & Lord, 1982). Thus there appears to be evidence for an organizer effect but whether it acts by increasing accuracy for category consistent or inconsistent behaviors is not clear. A possible explanation for this is the correction for false alarms. There are two issues that need to be resolved. One issue is why studies found differences in strength of organizer effect. A second issue is why some found higher accuracy in recall for consistent information whereas others found exactly the opposite effect. Each of these issues will be dealt with separately.

Strength of Organizer Effects

The studies that failed to find a difference in recall accuracy when a category is presented after observation were those that used a written format rather than a video or audio tape to present the behavioral information. This finding is consistent with Cohen's (1981) idea that a richer context is necessary for people to begin using simplifying mechanisms such as processing biases and retrieval cues to deal with incoming information. Thus it appears that there is an effect associated with providing subjects with a category after the presentation of behavioral information but this effect is present only when a richer more realistic format is used to present the behavioral information.

Contradictions

The second issue deals with the seemingly contradictory results which have been obtained. Some studies found higher accuracy in recall for category congruent items while others found higher accuracy in recall for category incongruent items. A closer

look at the dependent measures used may be helpful in resolving this inconsistency. In the Cohen (1981) study which found higher recall for category-consistent information, accuracy was measured using a two-alternative forced-choice instrument which contained only items that were present in the videotape. Even though the authors accounted for errors that were made by the participants, these errors consisted only of participants choosing items that did not go with the appropriate occupational category. The responses coded as errors did not include items which were stereotypical but which did not occur on the videotape. This is critical because Phillips and Lord (1982) have suggested that people are more likely to choose category-consistent behaviors and thus Cohen's results may be a reflection of guessing on the part of the subjects. Correcting Cohen's data for false positives may have cleared up the discrepancy between the studies.

In the Pyszczynski et al. (1987) study, accuracy was measured by counting the number of items that the subject wrote down which actually occurred in the audio tape. However, there is no mention in the results of taking into account the number of items written down which were not presented in the audio tape. Given the fact that accuracy is measured in different ways, it is difficult to directly compare the results of these studies as they relate to accuracy in recall. However, as some studies suggest, false alarm rates are an important issue in the study of processing biases. Thus the results of the Phillips and Lord (1982) appear to be the most appropriate because their measure accuracy takes into account the biases which can occur in the context of categorization effects. Therefore it seems that in the situation in which a category is made salient after the presentation of behavioral information, it is inconsistent information that is

more likely to be accurately recalled when the measure of accuracy corrects for false alarms.

Conclusion:

In conclusion, two issues need to be resolved in order to test the organizer model. First, the stimulus material should be as realistic as possible. This would include video or audio taped behavioral incidents. Second, the results should be compared with a similar condition in which the categorical information is presented before observation, and both measures must take false alarm rates into account. This would result in a direct comparison of the pre and post situation which controls for the effects of categorization (namely false alarms).

The information processing effects presented above assume that people have equal motivation to recall the information presented to them. However, it is reasonable to assume that not everyone is equally motivated to recall information and that these differences can influence the way people process information. This is an issue that will be dealt with in the following section.

Motivation

It has been recognized that the rater's motivation to give accurate ratings will affect the final ratings (DeCotiis & Petit, 1978). It has also been recognized that a rater's ability will limit the accuracy of ratings regardless of the level of motivation. As Taft (1971) has pointed out:

> if the judge is motivated to make accurate judgments about his subject and if he feels himself free to be objective, then he has a good chance of achieving his aim, provided of course that he has the requisite ability and can use the appropriate judgmental norms (p. 177).

As has been noted earlier, a person's ability to accurately recall information and hence, give accurate ratings, is affected by the processes used to encode the information. If selective processing took place (i.e. the filter model) then it is predicted that biased information will be present in memory. In this case, high motivation might only serve to increase the number of category consistent items recalled but should have no effect on inconsistent information because such information should not be properly encoded. However, in the situation where categories serve as retrieval cues (i.e. the organizer model), motivation should affect recall of both category consistent and inconsistent behaviors due to the fact that all behaviors should have been encoded equally during observation.

Thus it appears that motivation will have a differential impact on the recall of consistent and inconsistent information under the organizer condition but not on the filter condition. In the latter case, there will be no differential effect although recall should be elevated for both forms of information when low motivation is compared to high.

Motivation can be manipulated in a number of ways. Some studies have used monetary incentives. However, this manipulation assumes that all subjects will value a given amount of money similarly. In studies dealing with decision making, motivation has been successfully manipulated by varying accountability for the decisions (Tetlock, 1985). In the present study, deciding whether or not a behavior was observed requires

the participants to make a choice among several alternatives. Therefore, it was decided to manipulate motivation to be accurate by holding participants accountable to their decisions. As justification for the use of such a motivation manipulation, a review of the effects of using accountability to manipulate motivation in decision tasks follows.

The effects of accountability on decision making behavior has been studied by a number of researchers (see Tetlock, 1985). Accountability typically refers to the extent to which the decision maker's performance is open for review by others. Motivationally, it is suggested that the presence of accountability increases the probability of evaluation apprehension (Marcy, 1990). Accountability typically reduces the level of tolerance for mistakes and thus increases the demands for a correct decision (Beach & Mitchell, 1978).

Most studies have explored the effects of accountability on decision making strategy. More specifically, studies have focused on the type of information used by decision makers to make their decisions under conditions varying in accountability. Rozelle and Baxter (1981) manipulated accountability by telling subjects that they would be required to attend a later session to discuss their ratings with other students. The task was to rate students applying to graduate school by watching a videotape of the candidate during an interview. The results showed that high accountability subjects' descriptions of the candidate more closely matched the behaviors observed on the videotape than low accountability subjects. This increase in accuracy reflects an increase in motivation towards giving a "correct" answer.

In a study by Tetlock and Kim (1987), accountability was manipulated before and after participants were allowed to read responses by a test taker to 16 items from Jackson's Personality Research Form. The task was to predict the responses that the

same test taker would make to another set of questions. Accountability was manipulated by telling participants that they would be taking part in an interview describing the information used to base their predictions. Results showed that in the pre-exposure accountability condition, subjects reported more integratively complex impressions of test takers, made more accurate behavioral predictions, and reported more appropriate levels of confidence in their ratings than no accountability and post-exposure accountability subjects. The authors concluded that the accountability manipulation influenced the way in which subjects encoded and processed information resulting in a more complex integration of the information. The authors believed that this was because when participants are accountable, and know this prior to information presentation, they actively prepare for counter arguments to their expressed opinions. It can be argued that this active preparation for counter arguments reflects an increase in motivation toward producing correct responses.

McAllister, Mitchell & Beach (1979), studied the effects of accountability on decision strategy choice and anticipated time for performing a task. The task included rating the marketability of two groups of products and deciding which product had the best market potential. The results indicated that accountable subjects chose more cognitively complex strategies, anticipated needing more time to complete the task, and took more time in completing the actual task; all these were considered to be a reflection of higher motivation.

Researchers have concluded that when subjects expect to justify their decisions to others, they use more cognitively complex strategies which rely more in the information available to them than on information inferred through categorization (Tetlock, 1987). Furthermore, accountability seems to increase cognitive effort as

evidenced by the types of decision strategies employed by subjects in an accountability condition (Tetlock, 1983).

While accountability does not seem to have an effect on processing biases when introduced after the presentation of information, it does seem to increase the use of information that is available to the decision maker. Thus, in situations where participants are asked to recall information that they have observed at a previous time and are also informed that they will have to justify their responses to a group, one can expect that they will give a more accurate representation of the information that is available to them in memory. Furthermore, if a category is introduced after the observation of a target person, accountability should work to reduce the subjects' reliance on this category for making decisions and thus give a more accurate picture of the information that they posses in memory. However, if accountability is not present, subjects should be more likely to rely on an available category which requires less cognitive effort on their part.

Although motivation to perform decision tasks well has been assumed to increase with accountability, a more direct measure of motivation is required if one is to conclude that accountability has the effect of making participants think more deeply about the information that they posses in memory. One possible measure of motivation is the time that it takes for subjects to complete the task for which they will be held accountable. It is hypothesized that accountable subjects will take more time in completing the task due to the fact that they will have to come up with counter arguments to any criticism leveled against their decisions (McAllister et.al, 1979). It is also possible that accountable subjects will want to review the information that was presented to them more often than non-accountable subjects. Thus, if subjects are

asked whether they would like to review the information presented to them, accountable subjects should be more likely to say yes than non-accountable subjects. Accountable subjects may also be expected to rate themselves higher than non-accountable subjects in measures of effort and motivation toward the task.

Framework for the Current Research

The current research is based on the process view of performance which suggests that the rater's information processing system will have an impact on the accuracy of the final ratings (Feldman, 1981; Ilgen & Feldman, 1983). More specifically, the current research takes the approach of Feldman (1981) that categories and the time of their activation will affect the way information is encoded and retrieved. This study does not look at rating accuracy in terms of a rating scale but looks at accurate recall of behavioral events. This is based on the assumption that a necessary condition of being able to rate accurately is being able to accurately recall the events that lead to that rating.

Criteria

Accuracy. Since the present study deals with the effects of categories on processing of information, the literature reviewed suggested that a measure of accuracy that takes into account the systematic distortions produced by the presence of categories must be used. Lord (1985) argued that, when dealing with recall accuracy in these situations, one must take into account hit rates (correctly identified behaviors which were observed) as well as false alarm rates (behaviors which are reported to have been

observed but which were actually not exhibited). This can be done by using a questionnaire that contains an equal number of behavioral descriptions that are both exhibited and not exhibited in the stimulus materials on the ratee. Raters can then be asked to indicate whether each behavior on the list was or was not presented in the observed segment of behaviors, and, from these responses, accuracy indices can be constructed. First, the proportion of correctly and incorrectly identified behaviors can be calculated for both category consistent and category inconsistent behaviors using the following formulas:

$$\text{Proportion of consistent behaviors correctly Identified} = P_{Cc} = \frac{n_{Cc}}{t_{Cc}}$$

$$\text{Proportion of inconsistent behaviors correctly Identified} = P_{Ic} = \frac{n_{Ic}}{t_{Ic}}$$

where n_{Cc} is the total number of consistent behaviors correctly identified, n_{Ic} is the total number of inconsistent behaviors correctly identified, t_{Ic} and t_{Cc} are the total number of consistent and inconsistent behaviors actually exhibited by the ratee. In the present study t_{Ic} will be set equal to t_{Cc} in order avoid confounding consistency with set-size (Higgins & Bargh, 1987). Incorrectly identified behaviors can be calculated with the following formulas:

$$\text{Proportion of consistent behaviors incorrectly identified} = P_{Ci} = \frac{n_{Ci}}{t_{Ci}}$$

$$\text{Proportion of inconsistent behaviors incorrectly identified} = P_{Ii} = \frac{n_{Ii}}{t_{Ii}}$$

where n_{Ci} is the total number of consistent behaviors incorrectly identified, n_{Ii} is the total number of inconsistent behaviors incorrectly identified, t_{Ci} and t_{Ii} are the total number of consistent and inconsistent behaviors on the questionnaire which were not exhibited by the ratee. Similarly to the correct behaviors, t_{Ci} will be set equal to t_{Ii} in order to eliminate the set-size confound.

Finally, accuracy can be calculated for total, consistent, and inconsistent behaviors by subtracting the proportion of correctly identified behaviors from the proportion of incorrectly identified behaviors for each of these in the following manner:

Accuracy for Consistent Behaviors	$= A_C = P_{Cc} - P_{Ci}$
Accuracy for Inconsistent Behaviors	$= A_I = P_{Ic} - P_{Ii}$
Accuracy for Total behaviors	$= A_T = A_C + A_I$

Recognition vs. Recall. Memory is most frequently indexed by two forms of measures, recognition and recall. The present study uses a recognition measure of memory. The participants were presented with a list of items and required to mark the ones they remember as occurring in the stimulus material. It is called recognition because participants only have to recognize the items and not remember them. Recall, on the other hand, refers to measures that provide no cues as to what occurred; the participant is asked to write down every item which they remember occurring in the stimulus material.

Many studies have been carried out as an attempt to understand the similarities and differences between responses to recognition and recall measures (see Gillund &

Shiffrin, 1984 for a review). Researchers agree that both forms of memory tasks require the individual to utilize the information that they possess in memory. However, there is some disagreement as to whether or not a recognition task requires a complete search or merely the use of more familiar and easily accessed memory (Gillund & Shiffrin, 1984). This is important because if subjects rely on the familiarity of the object to be recognized, then their judgments will be biased toward those that are consistent with their impression (Wyer, Bodenhausen, & Srull, 1984). However, studies of these differences have been inconclusive and researchers have concluded that in general, both types of memory tasks require some search of the contents of memory (Gillund & Shiffrin, 1984).

This study uses recognition because a signal detection theory form of accuracy was selected to be consistent with the advice of Lord (1985). When one presents information to individuals and then ask them to make memory judgments, the number of correct items that they can remember is known by the researcher. However, when false alarms are assessed, it is impossible to know how many behaviors an individual is considering unless the total set of stimuli to be considered is controlled. With the recognition format, the total number of incidents, both consistent and inconsistent, is fixed by the list presented to the participants therefore ratios can be constructed.

Nature of Categories

The current research deals with performance appraisal situations and the effects of performance categories on recall of ratee behaviors. Performance categories are prototypes or schemas about a person's general level of performance on their job. Specifically, in this study raters were informed that the person they watched (or were to

watch) was either a good or a poor performer. DeNisi et al. (1984) have noted that job performance categories can be constructed to be salient to raters by studying past performance appraisals and adopting information from the appraisal forms. Thus, performance category in the present research was operationalized as the rating (i.e. good or bad) supposedly given to the target ratee (secretary) by another group of students.

Time of Activation of Categories

The time at which a category is made salient has been shown to influence the type of information that is recalled (e.g. Pyszczynski et al., 1987). More specifically, it has been found that in the filter model category consistent items are more accurately recalled while in the organizer model it is category inconsistent behaviors that are more accurately recalled. This variable was operationalized as the time at which the subjects got the see the bogus performance appraisal filled out by another group of students portraying the ratee (secretary) as a good worker. From the review of the literature presented above the following hypotheses were made regarding the time at which a category is made salient:

> Hypothesis 1a: Subjects for whom a category has been made salient prior to the observation of performance will more accurately recognize category-consistent behaviors than category-inconsistent behaviors (i.e. $A_c > A_i$).
>
> Hypothesis 1b: Subjects for whom a category has been made salient after the observation of performance will more accurately recognize category-inconsistent behaviors than category-consistent behaviors (i.e. $A_i > A_c$).

Another way to look at the combination of Hypotheses 1a and 1b is to say that an interaction is predicted between consistency and timing of activation of the performance category.

Cohen (1981) found that having a category present while observing behaviors helped subjects remember more total behaviors than not having one present. This result is consistent with the general conceptualization of categories as serving to organize incoming information and reduce cognitive chaos (Cohen, 1981). This is also consistent with the concept of advanced organizers studied in learning and training (Ausubel, 1968; Howell & Cooke, 1989). This literature has found that providing individuals with an organizing schema at a higher level of abstraction than that of the material that is to be learned aids in the recall of the material (Mayer & Bromage, 1980). Thus, the following hypothesis were made regarding total number of behaviors recognized:

> Hypothesis 2: Subjects for whom a category has been made salient prior to observation will more accurately recognize a greater number of behaviors than those for whom a category has been made salient after observation when the accuracy measure removes categorically consistent and inconsistent behaviors that are falsely reported as having been observed (i.e. Pre A_T > Post A_T).

Phillips and Lord (1982) found that introducing a category after observation resulted in a higher false alarm rate for category-consistent behaviors (i.e. organizer effect). The subjects were more likely to report remembering category-consistent behaviors that did not take place. Therefore, the following hypothesis were made regarding false alarm rates in recognition:

Hypothesis 3: Subjects for whom a category has been made salient after observation will have a higher false alarm rate for category-consistent behaviors than for category-inconsistent behaviors (i.e. $P_{Ci} > P_{Ii}$).

Motivation

In the present research, motivation was operationalized as the absence or presence of accountability which has been shown to influence motivation. This was manipulated by having one group of subjects believe that they would have to justify their responses in the checklist to the rest of the group members (accountability). The other group was not given this instruction. In order to test the motivational effect of the accountability manipulation, the time that each individual took in filling out the checklist was recorded. Thus if an individual took more time, it can be inferred that they were more motivated and put more effort into the task. The following general hypothesis was made regarding the effect of motivation on recognition:

Hypothesis 4: Subjects in the motivation condition will more accurately recognize more total behaviors than those in the no motivation condition (i.e. Mot. A_T > No Mot. A_T).

As discussed earlier, subjects presented with a category prior to observation (filter model) will engage in selective processing and thus are not able to recall as many category-inconsistent behaviors due to failure to attend to these behaviors. In contrast, subjects for which a category is made salient after observation (organizer model) should have most of the information encoded in a similar fashion therefore a motivation manipulation should have a greater impact. Therefore the following hypotheses were

made regarding motivation, timing of the introduction of a category, and consistency/inconsistency of behaviors observed:

Hypothesis 5: There will be a three way interaction between consistency, timing of category, and motivation. More specifically, of the subjects who receive the category information after observation, those who are not motivated will show a greater recall advantage over category inconsistent behaviors than those who are motivated (i.e. for motivated subjects $A_I = A_C$, for not motivated subjects $A_I > A_C$).

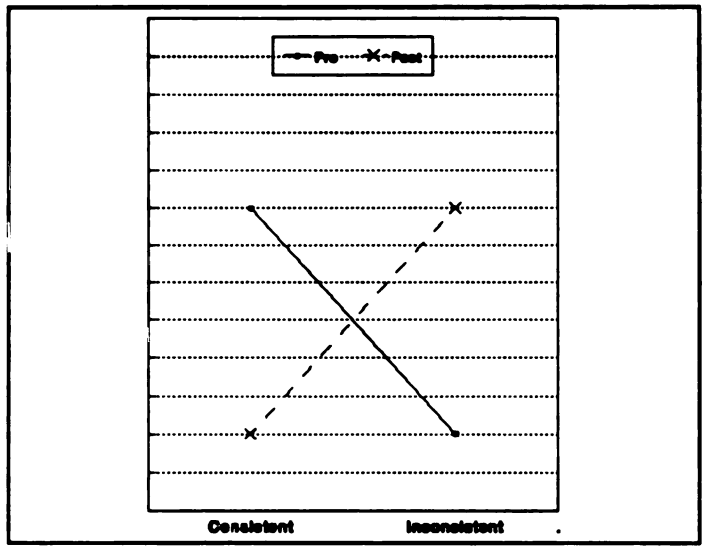
Of the subjects who receive the category information prior to observation, motivation should have no effect on the recall advantage of category consistent behaviors(i.e. $A_C > A_I$ for both motivation groups). This interaction is depicted in Figure 1.

Figure 1

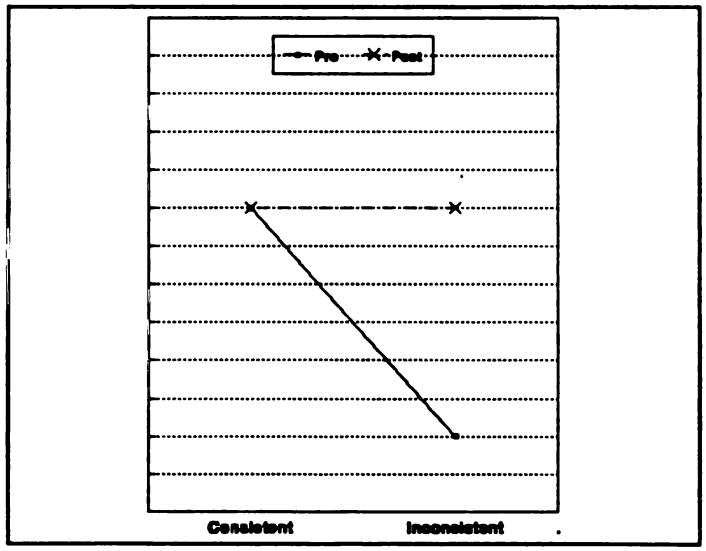
Expected Relationship Between Consistency, Category Timing,

and Motivation

NO MOTIVATION



MOTIVATION



METHOD

Subjects

Subjects were 132 undergraduate psychology students who received course credit for their participation in the study. A power analysis revealed that this number of subjects provided enough statistical power (.80) to detect a "medium" effect size with the design used (Cohen, 1988). Subjects participated in groups ranging from 2-7 individuals (mode=6). Each group was randomly assigned to one of four experimental group conditions.

Design

The current study was a 2 X 2 X 2 mixed design with 2 between-subjects factors and 1 within-subjects factor. The between-subjects factors were timing of category information presentation (before vs. after observation), and motivation/accountability (absent vs. present). The within subject factor was information consistency (category consistent vs. category inconsistent).

Stimulus Materials

The stimulus materials consisted of videotaped incidents of good and poor behavior exhibited by a secretary in a university department. The videotapes were part

of a larger set of videotapes and incidents developed for a study by Padgett and Ilgen (1989). The videotape contained 10 behavioral incidents out of which 5 were instances of good behavior and 5 were instances of poor behaviors. The behavioral incidents fell into one or more of four categories. These categories were (1) Job Knowledge and Skill, (2) Dealing with professors, (3) Working with Others, and (4) Organization of Work. The level of performance for each of the behavioral incidents was determined by having expert judges (secretaries) rate each of the incidents. Prior to making their ratings, the raters were asked to indicate which of the four dimensions were represented in the incident. They then rated the level of performance on the selected dimension(s) using seven point BARS developed specifically for clerical workers by Padgett and Ilgen (1989). The BARS scale anchors for dimension 1 (Job Knowledge and Skill) can be found in Figure 2. Appendix A contains the BARS scale anchors for the four dimensions. In order to determine the overall rating for the incident, the rating(s) for all the dimension(s) were averaged. The mean ratings for each incident on each dimension as well as the overall rating can be found in Table 1. Table 2 contains a description of each of these behavioral incidents. Only a subset of the behaviors was chosen for this study in order to have equal number of good and poor performance incidents. A behavioral incident was regarded as representing good performance if its overall rating was above 4.00 and poor behavior if its overall rating was below 4.00.

Figure 2

BARS for Dimension 1: Job Knowledge and Skill

Job Knowledge and Skill: The extent to which the secretary has acquired information relevant to doing his/her job (e.g. about university policies and procedures) as well as knowledge about the different technical aspects of his/her job and the extent to which he/she is able to execute them skillfully.

	Anchor	Description
	7	Secretary could be expected to correctly answer questions requiring specialized knowledge about university policies and procedures.
Above Average Performance		
	6	Secretary could be expected to have mastered the finer points of operating complicated office equipment.
	5	Secretary could be expected to correctly fix spelling and grammatical errors in letters and manuscripts he/she is asked to type.
	4	Secretary could be expected to follow correct style requirements
Average Performance		typing letters and manuscripts.
	3	Secretary could be expected to turn in manuscripts or letters that contain typographical errors.
	2	
Below Average Performance		
	1	Secretary could be expected to be unable to properly operate equipment used on the job (such as telephone, typewriter, or copier).

Table 1

Experts' Ratings of Behavioral Incidents Performed on the Videotape

Behavioral Incident	Job Knowledge	Dealing W/Profs	Working W/Others	Organization of Work	Overall for Incident
1	--	--	1.5[a]	1.5	1.50
2	--	1.5	--	1.3	1.40
3	5.4	5.9	--	--	5.64
4	2.1	2.9	--	--	2.50
5	6.2	--	5.9	--	6.05
6	--	3.4	--	2.0	2.70
7	--	--	1.4	--	1.40
8	--	--	6.4	--	6.40
9	6.2	6.1	--	--	6.16
10	--	5.0	4.9	--	4.93

[a] See Appendix A for a description of the seven point BARS scales used to rate each of the four dimensions.

Table 2

Description of Behavioral Incidents

On Videotape

Behavioral Incident Number	Description
1.	Gets angry at another secretary for not returning a borrowed stapler.
2.	Does not get a professor's presentation notes and overheads typed on time.
3.	Repeats a phone message from the wife of one of the professors back to her to be sure that it is correct.
4.	Does not know that a particular professor is in her department and is unable to tell a student where the professor's office is and what his office hours are.
5.	Demonstrates detailed knowledge about the various idiosyncrasies of the professors in the department (e.g. about correcting grammatical errors in their work or giving out their phone number).
6.	Does low priority work, such as filing, before typing exams, the latter of which should be given top priority.
7.	Refuses to change her plans to take a day off to accommodate the other secretaries.
8.	Brings up boxes from the mail room for another secretary.
9.	Allows a professor to use the copier even though she needs to use it. She instead goes down to the business library to use the copier.
10.	Does not let another secretary see confidential evaluation forms that she has been trusted to type up.

Procedure

Upon arrival to the experimental site, subjects were given a consent form to read and complete (see Appendix B). They were then asked to read a set of instructions concerning the present study (see Appendix C). These instructions stated that they were to watch a videotape which depicts a secretary working at a university. They were also told that their task was to observe the secretary for the purpose of rating her performance on a rating form as accurately as possible. A blank copy of the rating form was shown to all subjects prior to their watching of the videotapes (see Appendix D).

The category was manipulated by filling out one of the rating forms depicting the secretary was being "good" (i.e. circling fours and fives on the rating scales) (see Appendix E). Subjects were told that these ratings represented the mean ratings given to the secretary by another group of students. Category timing was manipulated by presenting subjects in the prior category with this rating form prior to viewing the videotape. The post category subjects were presented with this form after viewing the videotape and performing a filler task.

Accountability was manipulated by telling subjects that they would have to justify their responses to the other students in their group who were completing the task at that time. These groups ranged in size from 2-7 with a mode of 6.

After viewing of the videotape, the subjects were given a 15 minute distractor task (see Appendix F). This filler task was a baseball and a general knowledge quiz. All subjects were given fifteen minutes to work on this filler task and were told not to talk or work on anything else if they finished early. The purpose of this filler task was to allow enough time to pass between observation and recognition in order to place a

greater cognitive demand on the participants. After the filler task, subjects were given a checklist containing the dependent measure. After completing this checklist the participants were asked to rate the secretary on the four dimensions as well as an overall rating. The rating form used is the same one containing the category manipulation. The participants were then debriefed and told not to discuss the experiment with other potential subjects.

Dependent Measures

Measures Derived From Behavioral Checklist. The main dependent measure in this study is the checklist describing 20 behavioral incidents (see Appendix G). Half of them were actually represented on the videotape and the other half were not. Of each of these, 5 described instances of good behavior (consistent) and 5 described instances of poor behavior (inconsistent). The incidents which occurred on the videotape are those presented in Table 2. The behaviors which did not occur in the videotape were derived from Padgett and Ilgen (1989). Subjects were told to check off all incidents that they recalled observing in the videotape. The checklist contained some blanks where the subjects could write themselves notes that might help them in recognizing the behaviors. The checklist was used to compute proportions of consistent and inconsistent information that was correctly and incorrectly recognized. The following dependent measures were calculated from the responses to this checklist:

$$\text{Proportion of consistent behaviors correctly Identified} = P_{Cc} = \frac{n_{Cc}}{5}$$

Proportion of inconsistent behaviors correctly Identified $= P_{Ic} = \frac{n_{Ic}}{5}$

Proportion of consistent behaviors incorrectly identified $= P_{Ci} = \frac{n_{Ci}}{5}$

Proportion of inconsistent behaviors incorrectly identified $= P_{Ii} = \frac{n_{Ii}}{5}$

Accuracy for Consistent Behaviors $= A_C = P_{Cc} - P_{Ci}$

Accuracy for Inconsistent Behaviors $= A_I = P_{Ic} - P_{Ii}$

Accuracy for Total behaviors $= A_T = A_C + A_I$

Other Measures. Several other measures were obtained. These were the following:

TIME: Time in minutes spent working on the checklist. This measure was obtained by having the experimenter write down the time at which the participants in the group began working on the checklist. The participants in the group began working on the checklist at the same time. The participants were told to turn over the checklist as soon as they were finished. The experimenter then wrote down the time at which each of the participants finished the checklist. By subtracting the starting time from the ending time, a measure of time spent working on the checklist was obtained.

RATINGS: The participants were asked to rate the secretary on the four dimensions using the checklist in Appendix E. The checklist also contained a scale for an overall rating.

RESULTS

Descriptive Data

Table 3 contains the means, standard deviations, and intercorrelations for the variables in the study.

Manipulation Checks

In order to see if the accountability manipulation had the effect of increasing motivation in filling out the checklist, a t-test of the time spent working on this form was conducted. The results of this test can be found in Table 4. As the results indicate, the no accountability group spent an average of 2.7 (SD=1.3) minutes working on the checklist versus 3.8 (SD=2.2) minutes for the accountability group. These mean differences were statistically significant (t=-3.31, df (130), $p<.01$), and the manipulation accounted for 7.8% of the variance in time spent ($eta^2=.077$).

Analysis of Hypotheses

Overall Tests. A 2 X 2 X 2 repeated measures analysis of variance was conducted to assess the impact of timing of category and motivation on the recognition of category consistent and inconsistent behaviors (Hypotheses 1a, 1b, 5). The dependent variables were accuracy for consistent (A_c) and inconsistent (A_i) behaviors which made

up the within subjects effect of Consistency. The between subjects effects were Timing (pre vs post) and Motivation (absent or present).

The results of this analysis are presented in Table 5. The results indicated a nonsignificant interaction between consistency, timing, and motivation as was predicted by Hypothesis 5 ($F<1$). Hypotheses 1a and 1b predicted an interaction between consistency and timing. Results from Table 5 indicate a marginal effect ($F=3.11$, $df(1,128)$, $p<.10$). Table 6 presents the means associated with this interaction. An inspection of the means indicate that this effect is in the opposite direction than that predicted. The means as well as the significant main effect for consistency indicate that overall, subjects in both timing conditions recognized more category inconsistent behaviors. The interaction was due to the fact that in the pre condition, subjects recognized a much larger proportion of category inconsistent than consistent behaviors whereas those in the post condition recognized almost equal proportion of consistent and inconsistent behaviors (see Figure 3).

Overall Accuracy. The effects of category and motivation on overall accuracy were also tested. A 2 X 2 ANOVA with overall accuracy (A_T) as the dependent variable was conducted. Table 7 presents the results of this analysis. Hypothesis 2 predicted that subjects presented with the category prior to observation will more accurately recognize a greater number of behaviors than those receiving the category after observation. The results indicated a non-significant effect for timing ($F<1$). Hypothesis 4 predicted that subjects in the motivation condition will more accurately recognize a greater number of behaviors than those in the no motivation condition. The results indicate a marginally significant effect for motivation ($F=3.63$, df (1,130), $p<.10$).

Table 3

Means, Standard Deviations, and Intercorrelations

VARIABLE	MEAN	SD	1	2	3	4	5	6	7	8	9	10
1. P_{cc}[1]	.83	.17										
2. P_{ci}[1]	.05	.12	-.05									
3. P_{ic}[1]	.95	.10	.33*	-.07								
4. P_{ii}[1]	.09	.14	-.07	.15*	-.05							
5. A_c[2]	.78	.21	.83*	-.60*	.31*	-.15*						
6. A_i[2]	.86	.17	.25*	-.16*	.62*	-.81*	.30*					
7. A_T[3]	1.64	.31	.71*	-.50*	.56*	-.55*	.85*	.76*				
8. TIME	3.26	1.92	.14	.14	-.05	.02	.03	-.05	-.01			
9. RATING[4]	2.30	.57	.16*	.24*	-.01	-.17*	-.01	.13	.06	.06		
10. OVALL[5]	2.19	.67	.12	.20*	.02	-.21*	-.02	.17*	.08	.14	.85*	

* $p<.05$, $n=132$

[1] Possible range is 0 to 1.

[2] Possible range is -1 to +1.

[3] Possible range is -2 to +2.

[4] Possible range is 1 to 5. Corresponds to the average of the four dimension ratings.

This scale had a coefficient alpha of .60

[5] Possible range is 1 to 5. Corresponds to the one-item overall rating.

Table 4

Manipulation Check for Accountability

Accountability	Time (min)	SD	t	p
Absent	2.72	1.33		
Present	3.78	2.23	-3.31	.001

Table 5

ANOVA Summary Table of Consistency by Timing by Motivation

With Accuracy (A_c and A_i) as the Dependent Variable

SOURCE	df	MS	F	eta^2
Timing (T)	1	.05	.97	.00
Motivation (M)	1	.18	3.63*	.02
T x M	1	.00	.09	.00
Within (S/TM)	128	.05		
Consistency (C)[a]	1	.36	13.59**	.03
T x C	1	.08	3.11*	.01
M x C	1	.01	.54	.00
T x M x R	1	.02	.69	.00
Within (CS/TM)	128	.03		

*p<.10
**p<.05

[a]Consistency is a within subjects factor of accuracy for consistent (A_c) and inconsistent (A_i) behaviors

Table 6

Means of Consistency by Timing

	Consistent	Inconsistent
Pre	A_c=.752[a]	A_I=.861
Post	A_c=.814	A_I=.852

[a] The numbers indicate the proportion of behaviors correctly recalled after subtracting the proportion of behaviors incorrectly recalled (i.e. corrected for false alarms).

Figure 3

Interaction of Consistency by Timing

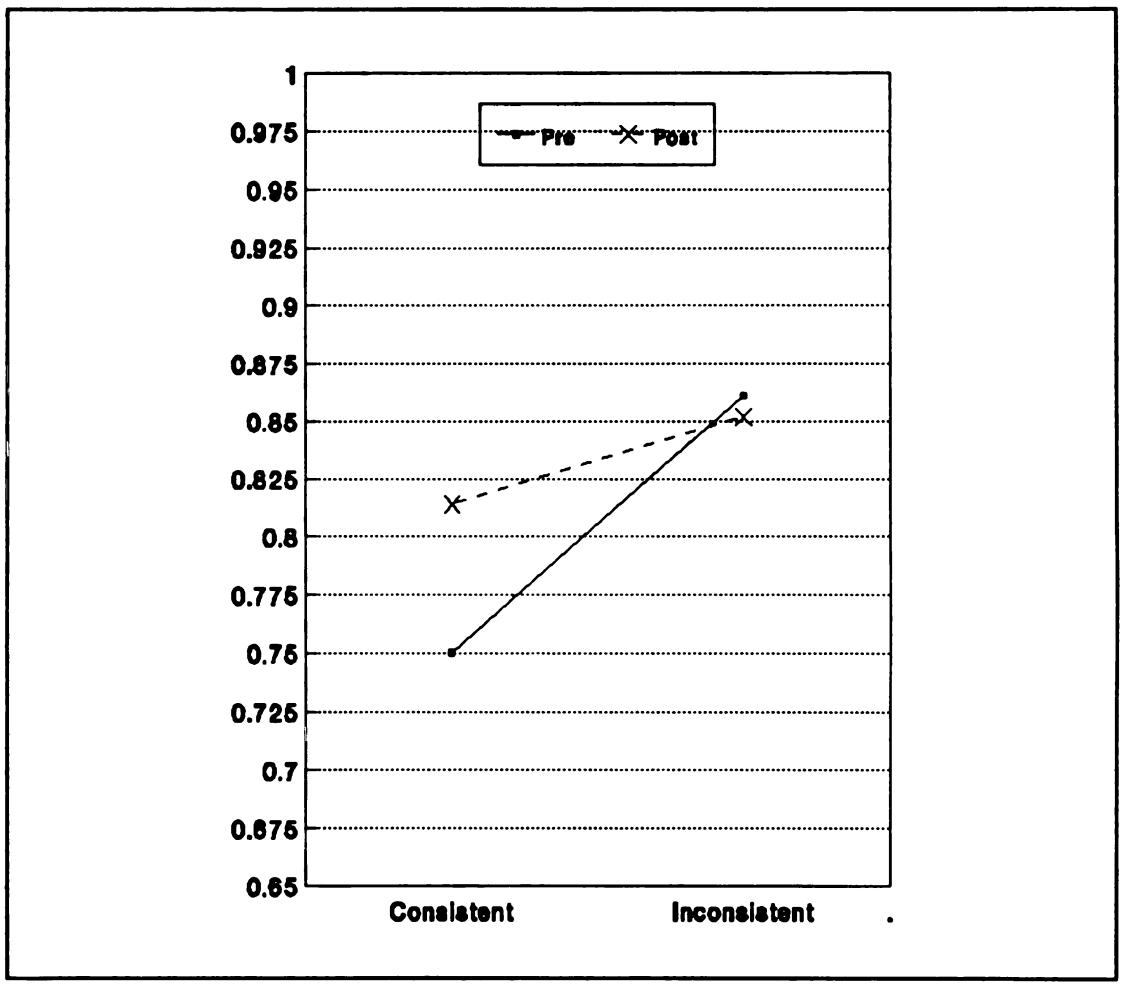


Table 7

ANOVA Summary Table of Total Accuracy (A_T) by Timing by Motivation

SOURCE	df	MS	F	eta^2
Timing (T)	1	.09	.97	.01
Motivation (M)	1	.35	3.63*	.03
T x M	1	.01	.09	.00
Within (S/TM)	128	.10		

* $p<.10$

An analysis of the cell means indicates that subjects in motivation condition actually had lower overall accuracy (m=1.59) than those in the no motivation condition (m=1.69). Thus the trend was in the opposite direction than predicted by Hypothesis 4.

False Alarm Rates. Hypothesis 3 stated that subjects who received category information prior to observation would have a greater false alarm rate for category-consistent behaviors than for category-inconsistent behaviors. To explore this effect an analysis of variance with the proportions of incorrectly identified behaviors (consistent and inconsistent) as the dependent variables was conducted. The independent variables were timing and motivation. The hypothesis predicted a significant interaction between consistency and timing. The results of this analysis are presented in Table 8. The results indicate a significant interaction between consistency and timing ($F=4.34$, df (1,130), $p<.05$). Table 9 presents the means associated with this interaction. The means indicate that subjects receiving the category prior to observation had nearly equal number of false alarms for consistent and inconsistent behaviors. Thus Hypothesis 3 was not supported. The interaction was due to the fact that subjects receiving the category after observation had a greater false alarm rate for inconsistent (m=.10) than consistent (m=.03) behaviors.

Table 8

ANOVA Summary Table of False Alarms (P_{Ci}, P_{Ii}) by Timing by Motivation

SOURCE	df	MS	F	eta^2
Timing (T)	1	.02	.48	.00
Motivation (M)	1	.01	.57	.00
T x M	1	.00	.05	.00
Within (S/TM)	128	.02		
Consistency (C)	1	.13	9.47*	.03
T x C	1	.06	4.24**	.01
M x C	1	.02	1.64	.00
T x M x C	1	.01	.94	
Within (CS/TM)	128	.01		

* $p<.01$
** $p<.05$

Table 9

Means of False Alarms by Timing

	Pre	Post
Consistent (P_{Ci})	.07	.03
Inconsistent (P_{Ii})	.08	.10

DISCUSSION

The purpose of this study was to test the effects of a category label on information recognition. More specifically, the purpose was to test the effects of providing a category label either prior to, or after, observing a ratee. The introduction of a category at these two points in time was hypothesized to differentially affect recognition of information about the ratee by either filtering information going into memory or organizing information previously stored in memory. These effects were studied by providing information about the ratee which was either consistent or inconsistent with the category label. Then, the effects of providing the category label at the two points in time was assessed by its impact on the recognition of category consistent and inconsistent behaviors.

The discussion of this study is organized in three sections. First a summary of the study results is presented as well as a summary and discussion of unexpected findings. Second, limitations of the study are presented and discussed in terms of their possible effects on the findings in this study. Finally, a discussion of future research is presented.

Summary of Results

Overall Tests. The main thesis presented in this study was that a category presented prior to observation affects the encoding of information by filtering out category inconsistent information (Hypothesis 1a). A category presented after observation was hypothesized to bias recognition by organizing the information stored in memory prior to receiving the category. This organization effect was hypothesized to result in a greater false alarm rate for category-consistent behaviors hence resulting in greater accuracy for category-inconsistent than consistent behaviors (Hypothesis 1b). These hypotheses predicted an interaction between consistency and timing. The results indicated a marginal interaction ($F=3.11$, df (1,128), $p<.10$). However, an examination of the means indicated that subjects receiving the category label prior to observation recognized a greater proportion of category-inconsistent behaviors whereas for those receiving the category label after observation this difference was not as great. This marginal interaction suggests that the category manipulation may have not been strong enough to have the hypothesized effect. This possibility will be discussed in more detail later.

Hypothesis 5 stated that under conditions of heightened motivation to recall accurately, those who received the category label after they have been presented with information about the ratee and presumably stored it in memory should be more able to counteract the bias created by the category as opposed to those receiving the category label prior to observation. The latter group should have biased information stored in memory due to the filtering effect of providing a category label prior to observation and motivation should have not effect in overcoming this bias.

This hypothesis predicted a three-way interaction between consistency, category timing, and motivation, with recognition accuracy for consistent and inconsistent information as the dependent variable. This interaction was not observed. The ability to detect this interaction depended on finding the interaction predicted by Hypotheses 1a and 1b. However, since this interaction was only marginally significant, it is not surprising that the three way interaction was not significant. There are several possible explanation for the failure to find an effect for the category manipulation.

The participants in the study were presented with a category label depicting the secretary as being a good performer. Furthermore, the videotape contained equal number of good and poor performance instances. However, the results indicate that all subjects had greater accuracy in recognizing category-inconsistent behaviors. In short, subjects were more accurate in recognizing instances of poor performance. The design of this study defined consistency and inconsistency in relation to the category presented to the participants. This category depicted the secretary as being a good worker. Therefore, consistent behaviors were those which were good behaviors and inconsistent were poor behaviors. The fact that this design confounds consistency with performance level is important in the failure to find some of the hypothesized results.

Defining consistency in relation to the "good secretary" category presented to the subjects makes the results of this study depend on the subjects accepting that category as their impression of the secretary at the point in time at which it is introduced. This category should then affect information acquisition in the pre group and information recall in the post group. If for some reason the category does not become the subjects' own impression, then it is very unlikely that the hypothesized effects for timing would be observed. There was some evidence to suggest that the category label of "good

secretary" did not have the effect of making the subjects observe or recognize the secretary with this category in mind.

If the category label was not accepted by the subjects, then it is possible that subjects formed their own impression and observed and recognized information according to that impression. One way to examine this possibility was to reanalize the data using the subjects' own impression of the secretary as the category. A surragate measure of the subjects' impression of the secretary was the overall rating from the rating form which the subjects completed at the end of the experiment. The mean rating given to the secretary was 2.19 (SD=.69) on a 5 point scale. This represents an average rating below the midpoint. Furthermore, 68% of the subjects rated the secretary 2 or below and 98% rated her 3 or below on this 5 point scale. These ratings were not affected by the time at which the category was introduced. This is evidenced by the results of an ANOVA of ratings by timing which was nonsignificant ($F<1$).

A further test of the hypothesis that the subjects formed their own impression involves analyzing some of the study hypotheses using the subjects' overall rating as a measure of their impression. Given the fact that there was no difference in ratings by timing, subjects were collapsed across the two levels of timing. Subjects who rated the secretary a 3 or above on their overall rating were put into a "good impression" group (n=41). Those which rated her 2 or below were put into a "poor impression" group (n=88). An analysis of variance was conducted with accuracy for good (A_c) and poor (A_i) behaviors as the dependent variables. The independent variable was the impression formed (good or poor). The results of this analysis can be found in Table 10. A main effect for Consistency was found ($F=18.95$, df (1,127), $p<.001$). This indicates that overall, the subjects had greater accuracy for poor behaviors. The results also showed

an interaction between Consistency and Impression ($F=4.80$, df (1,127) $p<.05$). Figure 4 shows the means associated with this interaction. These results indicate that subjects with a "good impression" had greater accuracy for poor behaviors. Subjects with a "poor impression" also had greater accuracy for poor behaviors than good ones but this difference was not nearly as large. This interaction indicates a trend toward greater accuracy for behaviors inconsistent with the subjects' impression. This result is consistent with Hastie's (1980) study who found that when subjects were asked to form an impression of a target person, inconsistent behaviors were better recalled. Thus, it appears that even though a category was presented to the subjects, they still formed their own impression which was different from the manipulation. A possible explanation for this finding can be found in the intructions given to the subjects. Even though they were presented with the category, the subjects were told to watch the videotape for the purpose of rating the secretary as accurately as possible. So, although the subjects were presented with a category label which represented another group of students' impression of the secretary, they might have felt that they needed to form their own impression in order to be as accurate as possible in their ratings. It is possible that a stronger manipulation of category might have overcome this ambiguity in the instructions. Another way to manipulate the category could be to let the subjects view a videotape of the secretary performing all good behaviors either prior to or after showing them the tape with good and poor performance instances.

Table 10

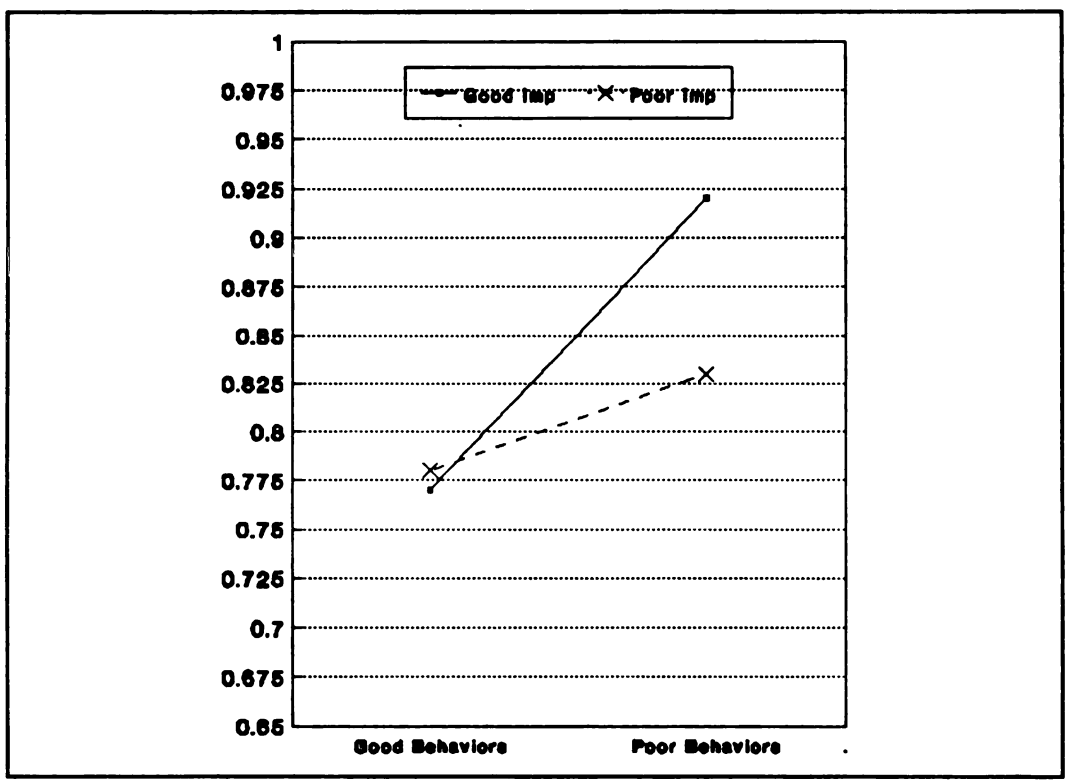
ANOVA Summary Table of Consistency by Impression

SOURCE	df	MS	F	eta^2
Impression (I)	1	.10	.16	.01
Within (S/I)	127	.05		
Consistency (C)	1	.50	18.95*	.05
I x C	1	.13	4.80**	.01
Within (CS/I)	127	.03		

* $p<.001$
** $p<.05$

Figure 4

Interaction of Consistency by Impression



Overall Accuracy. Overall accuracy (A_T) was defined as the sum of the proportions of category-consistent and inconsistent behaviors after subtracting the proportions of category consistent and inconsistent behaviors incorrectly identified. Hypothesis 2 predicted that subjects receiving the category label prior to observation would have greater overall accuracy than those receiving the category label after observation. Contrary to Hypothesis 2, the results indicated a nonsignificant effect for timing of category. As stated earlier, a possible reason for the nonsignificant interaction was the failure of the category manipulation to make salient a "good secretary" impression to the subjects at two different points in time. If this were the case, then receiving the category label prior to observation may have formed their own impression in a manner similar to those receiving the category after observation. The data are consistent with this interpretation due to the absence of an effect for category timing on overall accuracy.

Hypothesis 4 predicted that subjects in the motivation condition would have greater overall accuracy than those in the no motivation condition. Although a marginal effect was found, it was in the opposite direction. Subjects in the no motivation condition tended to have greater accuracy than those in the motivation condition. There are at least two possible explanations for this. Given that accuracy is dependent on the proportion of behaviors correctly as well as incorrectly identified, accountability could have had an impact on either of these two measures. As a result, heightened motivation could have decreased accuracy by (1) increasing the number of false alarms or (2) decreasing the number of behaviors correctly identified. Each of these possible explanations was explored separately.

In order to test the possibility that motivation increased the number of false alarms, the effects of motivation on false alarms were investigated using an analysis of variance. The dependent variable was false alarms for consistent (P_{Ci}) and inconsistent (P_{Ii}) behaviors as the within subject factor of Consistency. The independent variable was motivation. No effect for motivation on false alarms was found.

The second possible explanation was that motivation caused subjects to record fewer correct behaviors. Such a response would have occured if the increased motivation led the subjects to be more cautious in an attempt to reduce their mistakes, and it would decrease overall accuracy if the number of false alarms stayed constant. An analysis of variance was conducted with the proportion of consistent (P_{Cc}) and inconsistent (P_{Ic}) behaviors correctly identified. A significant main effect for motivation was found ($F=3.93$, $p<.05$, $eta^2=.03$). The pattern of the means indicated that the high motivation group recorded fewer correct behaviors ($m=.87$) than the no motivation group ($m=.91$). This is consistent with the hypothesis that the motivation group had lower overall accuracy because they recorded fewer items. A possible explanation for this finding is that manipulating motivation with accountability might have had the effect of causing accountable subjects to mark only those behaviors of which they were absolutely sure because they knew they would have to justify their answers to the rest of the group members. It should be noted that these explanations are purely exploratory. Furthermore, the original interaction between accuracy and motivation was only marginally significant. A stronger test of this hypothesis would be to have subjects rate their certainty of having observed a particular behavior. One would expect a higher degree of certainty for the no accountability group on more total behaviors than the

accountability group. The latter group might show a high degree of certainty on a fewer number of behaviors.

False Alarms. Hypothesis 3 predicted that subjects receiving the category information after observation would have a greater false alarm rate for category-consistent than category-inconsistent behaviors. The results indicated a significant main effect for Consistency and a Consistency X Timing interaction when false alarms are used as the within subjects factor of Consistency. The main effect indicated that all subjects had a greater false alarm rate for poor behaviors (inconsistent). The interaction indicated that subjects receiving the category after observation showed a greater difference between the number of consistent and inconsistent behaviors incorrectly identified with inconsistent behaviors having the larger false alarm rate (see Figure 5). Given the fact that consistency was defined in terms of the "good secretary" category presented, the hypothesis was not supported because poor performance was inconsistent with this category. However, since the data seemed to suggest that subjects formed a negative impression of the secretary on their own, then poor behaviors were consistent with this impression. Therefore, the results would not have been the opposite predicted if one assumes that the category held by the subjects was poor rather than good performance.

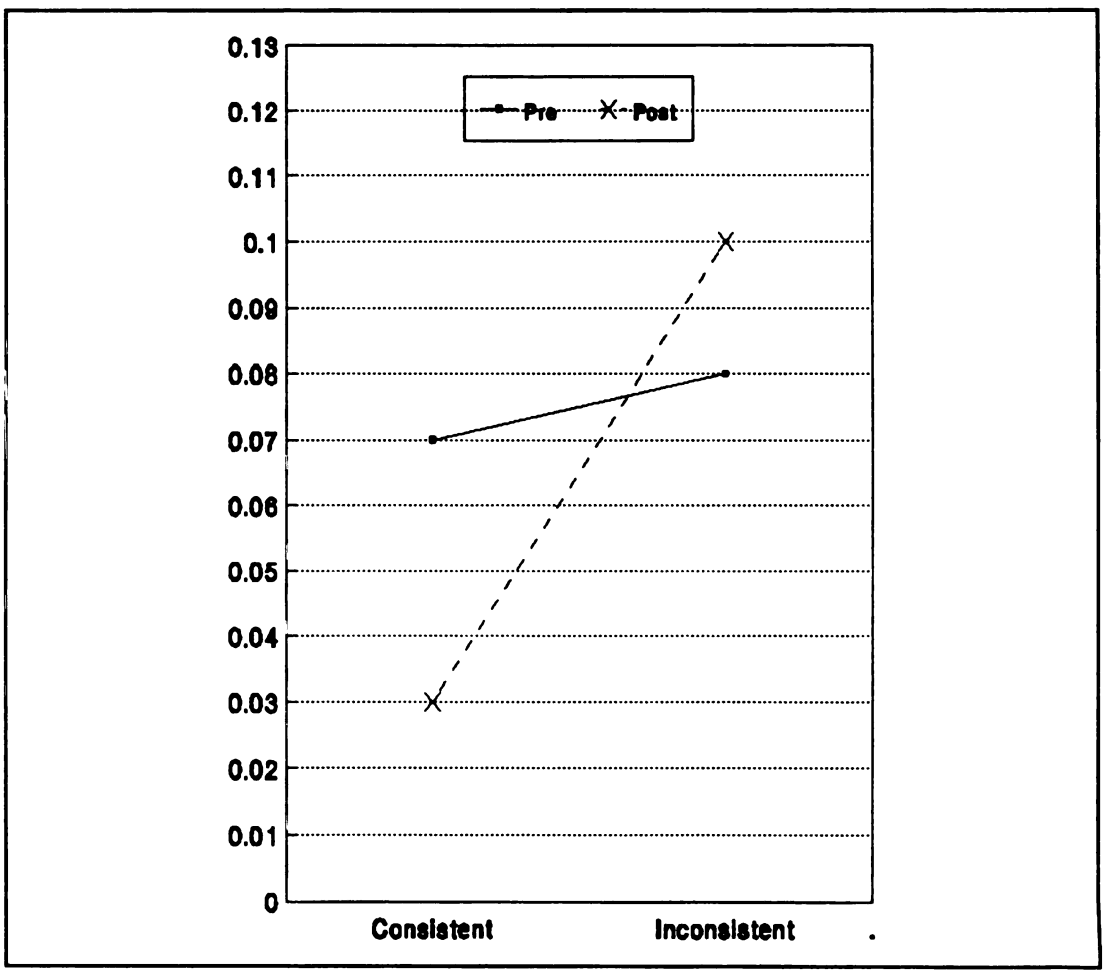
Further evidence for the conclusion that the results must be interpreted using the subjects' own impression is the correlation between overall ratings and proportion of consistent and inconsistent behaviors incorrectly identified. The correlation between subjects' overall ratings and the proportion of good performance instances incorrectly identified was .20 ($n=129$, $p<.05$), and it was -.21 ($n=129$, $p<.05$) with poor

performance instances incorrectly identified. That is, the more positive the subjects' impression were, the higher the false alarm rate for good behaviors. Conversely, more positive impressions were related to lower false alarm rates for poor behaviors. These results are consistent with those of Phillips and Lord (1982) who found that subjects had greater false alarms for behaviors consistent with their overall impression.

An issue which needs to be resolved is the fact that category timing had an effect on false alarms as well as accuracy as evidenced by the Consistency x Timing interaction for Hypotheses 1a, 1b, and 3. As has already been stated, the failure to make the "good secretary" category the one used by subjects to process and recognize information led to the conclusion that the subjects were forming their own impression and the category manipultation was not strong enough to counter this impression. Evidence for this was found in the fact that when the data were analyzed using the subjects' impression as the category, they were consistent with past literature. However, the interaction between false alarms and timing seems to suggest that the category might have had some effect. But the differences in false alarms, which amounted to less than one half of a behavior difference, may not have been enough to affect accuracy. This finding points to the possibility of restriction of range limiting the findings in the present study. This will be discussed in the following section.

Figure 5

False Alarms by Timing



Limitations of the Study

There are several limitations to this study that could pose a potential threat to its internal validity. The first has to do with the category manipulation. In general, the subjects did not accept the manipulation that described the secretary's performance as good. If, rather than using "good secretary" as the category for evaluating the results of the study, the subjects' reported impression is used, the results are more similar to those predicted. Thus, it appears that subjects might have paid attention to this manipulation but then felt free to form their own impression. Furthermore, since subjects were told that they would be rating the secretary, this could have had the effect of asking subjects to form their own impression.

The second limitation has to do with restriction of range on the dependent variables. A filler task was used to take the subjects' mind off the tape before filling out the checklist, but it appears that the amount of time taken up by this task may have not been enough. Most subjects recognized an average of 4 behaviors (out of 5) for each of the two performance types (good and poor). Furthermore, most subjects falsely identified less than 1 behavior of each type. It appears that either a longer delay between observation and recognition and/or the inclusion of category irrelevant information might have helped to make the subjects to rely on a category which is given to them in order to reduce the cognitive load. This restriction of range is important because most of the hypotheses in the present study predicted differences across conditions due to penalties imposed by subtracting false alarms. However, when false alarms account for less that one half of a behavior difference across groups, it is not surprising that no difference in accuracy was found.

A final limitation which has already been discussed has to do with the motivation manipulation. It appears that instead of motivating the subjects to recognize as much as possible, the manipulation served to make subjects more cautious when answering the checklist. This effect severely limited the possibility of finding an effect for the main hypothesis presented in this study because it relied on this manipulation increasing accuracy and not decreasing it.

Future Research

There are several ways of improving the present study which future research can address. The first issue has to do with the cognitive demands placed on the subjects. A longer delay between observation and recognition should make the raters rely more on a performance category. Also the inclusion of category-irrelevant behaviors should aid in making the task more cognitively demanding.

Another issue is that of motivating individuals to accurately recognize behaviors. As has already been discussed, accountability in this case appeared to serve to make the raters mark only those behaviors which they thought they could defend. To explore this more, it would have helped to have subjects give confidence ratings along with their recall responses. One would expect that accountability subjects would show a high degree of confidence on the behaviors that they identify as having been seen when compared to the confidence ratings of the no accountability raters. If this were the case, then it would suggest that making managers accountable for justifying their ratings in terms of the behaviors observed could lead to ratings based on only a few behavioral instances. This is only a problem if the behaviors that are recalled are not representative of the total behaviors exhibited by the ratee. So for example, if a

subordinate exhibits 40 instances of good performance and 10 instances of poor performance and the manager is only certain of 4 instances of good behavior and 1 instance of poor behavior, then the ratings of the subordinate would probably not be biased. However, if the manager is sure of 10 instances of poor behavior and only 5 of good behavior bias should be more likely.

Future studies should address the question of what happens when the category which is introduced is opposite to the impression that participants form on their own. If the participants in the present study formed a poor impression, then those receiving the category after observation were receiving a category which was opposite to the one which they formed on their own. It is possible that in this situation, characteristics of the source, such as source credibility, become important. One may be more apt to change his or her impression if the source is highly credible. This effect has been found in research dealing with attitude change (e.g. see review by Chaiken & Stangor, 1987). However, in the present study, characteristics of the source of the category were kept constant. It is interesting to think of what would be the effects of substituting a subject's impression with another from a more or less credible source. Would the effects on recall occur in relation to their original impression or with the newly acquired impression? The issues raised above should help in eliminating some of the weaknesses in the present study and gaining further understanding of the effects of a category on the recall of behavior.

LIST OF REFERENCES

LIST OF REFERENCES

Ausubel, D.P. (1968). Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston.

Bargh, J.A., & Thein, R.D. (1985). Individual construct accessibility, person memory, and the recall-judgement link: The case of information overload. Journal of Personality and Social Psychology, 49, 1129-1146.

Beach, L.R., & Mitchell, T.R. (1978). A contingency model for the selection of decision strategies. Academy of Management Review, 3, 439-449.

Belleza, F.S., & Bower, G.H. (1981). Person stereotypes and memory for people. Journal of Personality and Social Psychology, 41, 856-865.

Bodenhausen, G.V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. Journal of Personality and Social Psychology, 55, 726-737.

Cantor, N. & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol 12). New York: Academic Press.

Chaiken, S. & Stagnor, C. (1987). Attitudes and attitude change. Annual Review of Psychology, 38, 575-630.

Clark, L.R., & Woll, S.B. (1981). Stereotype biases: A reconstructive analysis of the role in reconstructive memory. Journal of Personality and Social Psychology, 41, 1064-1072.

Cohen, C.E. (1981). Person categories and social perception: Testing some boundaries of the processing effects of prior knowledge. Journal of Personality and Social Psychology, 40, 441-452.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd ed. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Crocker, J., Hannah, D.B., & Weber, R. (1983). Person memory and causal attributions. Journal of Personality and Social Psychology, 44, 55-66.

DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable hypothesis. Academy of Management Review, 21, 635-646.

DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.

Feldman, J.M. (1981). Beyond Attribution Theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Fiske, S.T., & Taylor, S.E. (1984). Social Cognition. Reading, MA: Addison-Wesley.

Gillund, G. & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In R. Hastie, T. Ostrom, E. Ebbensen, R. Wyer, D. Hamilton, & D. Carlston (Eds.), Person memory: Cognitive basis of social perception (pp. 155-178), Hillsdale, N.J.: Erlbaum.

Higgins, E.T., & Bargh, J.A. (1987). Social cognition and social perception. Annual Review of Psychology, 38, 369-425.

Howell, W.C. & Cooke, N.J. (1989). Training the human information processor: A review of cognitive models. In Goldstein, I.L. & Associates (Eds.), Training and Development in Organizations (pp. 121-182), San Francisco, CA: Jossey-Bass.

Ilgen, D.R. & Feldman, J.M. (1983). Performance Appraisal: A process focus. Research in Organizational Behavior, 5, 141-197.

Landy, F.J. & Farr, J.L. (1980). Performance Rating. Psychological Bulletin, 87, 72-107.

Lichtenstein, M., & Srull, T.K. (1987). Processing objectives as a determinant of the relationship between recall and judgement. Journal of Experimental Social Psychology, 23, 93-118.

Lord, R.G. (1985). Accuracy in behavioral measurement: An alternative definition based on rater's cognitive schema and signal detection theory. Journal of Applied Psychology, 70, 66-71.

Marcy, L.M. (1990). Accountability of the decision maker: A closer examination of it's role in decision making. Unpublished Master's Thesis. Michigan State University, East Lansing, MI.

Mayer, R.E. & Bromage, B.K. (1980). Different recall protocols for technical texts due to advance organizers. Journal of Educational Psychology, 72, 209-225.

McAllister, D.W., Mitchell, T.R., & Beach, L.R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. Organizational Behavior and Human Performance, 24, 228-244.

Nathan, B.R. & Lord, R.G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68, 102-114.

O'Sullivan, C.S., & Durso, F.T. (1984). Effect of schema-incongruent information on memory for stereotypical attributes. Journal of Personality and Social Psychology, 47, 55-70.

Padgett, M.Y. & Ilgen, D.R. (1989). The impact of ratee performance characteristics on rater cognitive processes and alternative measures of rater accuracy. Organizational Behavior and Human Decision Processes, 44, 232-260.

Phillips, J.S., & Lord, R.G. (1982). Schematic information processing and perceptions of leadership in problem-solving groups. Journal of Applied Psychology, 67, 486-492.

Pyszczynski, T., LaPrelle, J., & Greenberg, J. (1987). Encoding and retrieval effects of general person characterizations on memory for incongruent and congruent information. Personality and Social Psychology Bulletin, 13, 556-567.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. Journal of Experimental Social Psychology, 15, 343-355.

Rozelle, R.M. & Baxter, J.C. (1981). Influence of role pressures on the perceiver: Judgments of videotaped interviews varying judge accountability and responsibility. Journal of Applied Psychology, 66, 437-441.

Snyder, M., & Uranowitz, S.W. (1978). Reconstructing the past: Some cognitive consequences of person perception. Journal of Personality and Social Psychology, 36, 941-950.

Srull, T.K. (1981). Person memory: Some tests of associative storage and retrieval model. Journal of Experimental Psychology: Human Learning and Memory, 7, 440-463.

Srull, T.K., & Wyer, R.S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. Journal of Personality and Social Psychology, 38, 841-856.

Srull, T.K., & Wyer, R.S. (1989). Person memory and judgement. Psychological Review, 96, 58-83.

Taft, R. (1971). The ability to judge people. In W.W. Ronan & E.P. Prien (Eds.), Perspectives on the Measurement of Human Performance. New York: Appleton-Century-Crofts.

Tetlock, P.E. (1983). Accountability and the complexity of thought. Journal of Personality and Social Psychology, 45, 74-83.

Tetlock, P.E. (1985). Accountability: The neglected social context of judgement and choice. In B.M. Staw & L. Cummings (Eds.), Research in Organizational Behavior, Vol. 7. Greenwich, CT: JAI Press.

Tetlock, P.E., & Kim, J.I. (1987). Accountability and judgement processes in a personality prediction task. Journal of Personality and Social Psychology, 52, 700-709.

Wyer, R.S., & Gordon, S.E. (1982). The recall of information about person and groups. Journal of Experimental Social Psychology, 18, 128-164.

Wyer, R.S., & Gordon, S.E. (1984). The cognitive representation of social information. In R.S. Wyer & T.K. Srull (Eds.), Handbook of social cognition (Vol. 2, pp. 73-150). Hillsdale, N.J.: Erlbaum.

Wyer, R.S., Bodenhausen, G.V., & Srull, T.K. (1984). The cognitive representation of persons and groups and its effect on recall and recognition memory. Journal of Experimental Social Psychology, 20, 445-469.

Wyer, R.S., Srull, T.K., Gordon, S.E., & Hartwick, J. (1982). Effects of processing objectives on the recall of prose material. Journal of Personality and Social Psychology, 43, 674-688.

APPENDICES

APPENDIX A

BARS for Dimension 1: Job Knowledge and Skill

Job Knowledge and Skill: The extent to which the secretary has acquired information relevant to doing his/her job (e.g. about university policies and procedures) as well as knowledge about the different technical aspects of his/her job and the extent to which he/she is able to execute them skillfully.

Anchor		Description
	7	Secretary could be expected to correctly answer questions requiring specialized knowledge about university policies and procedures.
Above Average Performance		
	6	Secretary could be expected to have mastered the finer points of operating complicated office equipment.
	5	Secretary could be expected to correctly fix spelling and grammatical errors in letters and manuscripts he/she is asked to type.
	4	Secretary could be expected to follow correct style requirements
Average Performance		typing letters and manuscripts.
	3	Secretary could be expected to turn in manuscripts or letters that contain typographical errors.
	2	
Below Average Performance		
	1	Secretary could be expected to be unable to properly operate equipment used on the job (such as telephone, typewriter, or copier).

BARS for Dimension 2: Dealing With Faculty, Students, etc.

Dealing With Faculty, Students, etc.: The extent to which the secretary interacts in a positive manner with professors, students, callers, or other people that he/she comes into contact with on the job (e.g.positive attitude, willing to help out)

Anchor		Description
	7	Secretary could be expected to volunteer to take work home with him/her in order to get an emergency assignment done for a professor on time.
Above Average Performance		
	6	Secretary could be expected to offer help a professor with some of their special work that isn't part of his/her regular job duties.
	5	Secretary could be expected to be pleasant and helpful to people who call the office on the phone or who come in to ask questions.
	4	
Average Performance		
	3	Secretary could be expected to get upset and defensive when a professor asks him/her to correct his/her work.
	2	Secretary could be expected to gossip with other people in the office about administrators, professors or students that he/she doesn't like.
Below Average Performance		
	1	

BARS for Dimension 3: Working Cooperatively With Other Secretaries

Working Cooperatively With Other Secretaries: The extent to which the secretary works as a part of a team with the other secretaries (e.g. by sharing the workload with them, sharing important information with them, checking with them before taking time off, etc.).

Anchor		Description
	7	Secretary could be expected to offer to help other secretaries with their work if he/she was finished with his/her work.
Above Average Performance		
	6	Secretary could be expected to share with the other secretaries any new or special information he/she learns that might be helpful to them.
	5	Secretary could be expected to be willing to teach an inexperienced office worker basic information about office functioning.
	4	
Average Performance		
	3	Secretary could be expected to agree to help if another secretary asked him/her to help proofread a manuscript the other secretary had typed.
	2	Secretary could be expected to insult or ridicule another secretary who makes a mistake on the job.
Below Average Performance		
	1	

BARS for Dimension 4: Organization of Work

Organization of Work: The extent to which the secretary is able to organize his/her time and work to get things done efficiently, is able to set priorities concerning what work is most important, is able to handle multiple demands on his/her time and can meet deadlines.

Anchor		Description
	7	Secretary could be expected to develop an innovative procedure that is more efficient and saves work time.
Above Average Performance		
	6	Secretary could be expected to use time when there is no work to be done in the office to straighten up and reorganize some files rather than reading the paper or talking.
	5	Secretary could be expected to prioritize work assignments so that the most important things get done first.
	4	Secretary could be expected to have his/her work Performance station organized so that he/she can quickly and easily locate supplies and information that he/she needs.
Average Performance		
	3	Secretary could be expected to be sufficiently organized the he/she is able to get work assignments completed on time.
	2	Secretary could be expected to get upset and flustered if he/she was in the office alone when several people needed to speak with her and the phone was ringing frequently.
Below Average Performance		
	1	Secretary could be expected to miss important work assignment deadlines.

APPENDIX B

CONSENT FORM
Employee Rating Study

This study is designed to investigate how people rate employees. You will be asked to watch a short videotape of a secretary at work. You will then be asked a series of questions relating to the tape.

Your participation in this research is completely voluntary and you are free to decline to answer any questions or to terminate the session at any time. Termination prior to completion of the session will not affect your credit for participation. Your participation in this study will be confidential. Your data will be included in a summary report along with those of others, but will not be broken down in any way in which the responses can be identified individually as yours. If you have any questions regarding the study, you may ask the experimenter.

Subject's Statement

I agree to participate in the Employee Rating Study. I understand that I will be asked to fill out a questionnaire. Furthermore, I understand that I will be asked to fill out a questionnaire. Furthermore, I understand that my participation is voluntary, that I may discontinue at any time without any adverse consequences to me, that my answers will be anonymous, and that my responses will not be analyzed in any way that allows me to be identified.

Signature_____
Date _____

Printed Name_____

Experimenter: Miguel Quiñones
22 Baker
355-2171

APPENDIX C

Employee Rating Study

Instructions

Welcome to the Employee Rating Study. In this study you will be playing the role of a manager that has just begun a new job. One of your first tasks is to rate the performance of your new employees. Your employees are university secretaries working in the management department. Please take a moment to examine the **EMPLOYEE EVALUATION FORM** which you received from the experimenter. In this form there is a rating scale ranging from Unacceptable to Outstanding behavior. This scale is used to rate the employee on the four dimensions shown. These dimensions are (1) Job Knowledge and Skill, (2) Organizational Ability, (3) Dealing with Faculty/Students, and (4) Working Cooperatively with Others. There is also a section for an overall evaluation and some comments.

You will be watching a videotape of your employee during a normal work day. It is your task to watch this videotape for the purpose of rating the employee as **ACCURATELY** as possible. If you have any questions please ask the experimenter.

APPENDIX D

EMPLOYEE EVALUATION FORM

Employee Name_____ Job Title_____

Review Date_____

Use the following scale to rate the employee's performance on the following dimensions.

1 Unacceptable	2 Below Average	3 Average	4 Above Average	5 Outstanding

Job Knowledge and Skill	1 2 3 4 5
Organizational Ability	1 2 3 4 5
Dealing With Faculty/Students	1 2 3 4 5
Working Cooperatively With Others	1 2 3 4 5
Overall Rating	1 2 3 4 5

Comments

.....

.....

.....

.....

.....

.....

APPENDIX E

EMPLOYEE EVALUATION FORM

Employee Name Sue Johnson Job Title Secretary I

Review Date 1/15/89

Use the following scale to rate the employee's performance on the following dimensions.

	1 Unacceptable	2 Below Average	3 Average	4 Above Average	5 Outstanding
Job Knowledge and Skill					(5)
Organizational Ability				(4)	
Dealing With Faculty/Students					(5)
Working Cooperatively With Others				(4)	
Overall Rating					(5)

Comments

Sue is an excellent worker. She is very knowledgeable and gets along well with others. She is very efficient around the office. Sue is definitely an asset in our department.

APPENDIX F

Baseball Knowledge Quiz

Instructions: Read the following questions and circle the correct answer.

1. Who holds the record for the most home runs hit in a single season?

   a. Roger Maris
   b. Micky Mantle
   c. Reggie Jackson
   d. Jose Canseco

2. How many feet are there between home plate and first base?

   a. 79.5
   b. 100
   c. 60
   d. 90

3. The left field wall in Boston's Fenway Park is referred to as:

   a. the ivy border
   b. the wall of fame
   c. the brick barrier
   d. the green monster

4. The "Cy Young Award" is an award for the best:

   a. rookie
   b. batter
   c. pitcher
   d. fielder

5. A "texas Leaguer" refers to:

   a. a baseball player from Texas Minor Leagues
   b. an error made by an outfielder
   c. a hit made between an infielder and outfielder
   d. a rookie who is still "wet between the ears"

6. What is a squeeze play?

   a. a runner on third runs home while the batter bunts the ball
   b. two outfielders run into each other when chasing a ball
   c. a shortstop and second baseman convert a double play
   d. a pitcher pretends to throw the ball to home plate and the runner at first is tagged out by the first baseman who has the ball

7. What is the name of the league that is immediately below the Major Leagues?

   a. Texas League
   b. AAA League
   c. Independent League
   d. Minor League

8. When a "balk" is called during a game what occurs?

   a. the batter is given a strike
   b. play is halted to allow the umpires to confer on a decision
   c. runners are allowed to move to the next base
   d. the pitcher is thrown out of the game for throwing a "spit" ball

9. The "infield fly" rule is invoked only when:

   a. a popup goes foul
   b. there are no runners on base
   c. there are two outs already
   d. none of the above

10. If the batter walks with the bases loaded:

    a. the batter gets an RBI and is charged with a time at bat
    b. the batter does not get an RBI or a time at bat
    c. the batter gets the RBI but not the time at bat
    d. the batter does not get the RBI and is charged with a time at bat

11. How many innings does a starting pitcher have to work to be elegible for the win?

    a. one
    b. nine
    c. five
    d. seven

12. Who won last year's (1989) World Series

    a. Chicago Cubs
    b. San Francisco Giants
    c. Oakland Athletics
    d. Baltimore Orioles

13. A passed ball is:

    a. the same as a wild pitch
    b. charged as an error to the pitcher
    c. charged as an error to the catcher
    d. both a and b

14. Using standard scorekeeping notation, a 5-4-3 play starts with:

    a. the shortstop
    b. the second baseman
    c. the third baseman
    d. the first baseman

15. Which of the following is NOT a current major league baseball team?

    a. Seattle Mariners
    b. Texas Rangers
    c. Washington Capitals
    d. San Diego Padres

16. How many umpires work during a Major League regular season game?

a. four
b. six
c. three
d. two

17. What kind of wood are bats made of?

a. pine
b. ash
c. oak
d. hickory

18. A good fast ball pitcher like Dwight Gooden can throw the ball at approximately what speed?

a. 50 mph
b. 80 mph
c. 95 mph
d. 110 mph

19. A "knuckle ball" is a pitch that is thrown by holding the ball:

a. on the fingertips
b. with the index finger and thumb
c. in the palm
d. with the knuckles

20. The rubber of the pitcher's mound is how far from home plate?

a. 60'6"
b. 46'10"
c. 90'8"
d. 72'2"

21. How many teams are there in the Major Leagues?

a. 14
b. 22
c. 26
d. 32

22. The oldest baseball park in which Major League games are still played is:

a. Tiger Stadium
b. Boston Fenway Park
c. Chicago Wrigley Field
d. Chicago Cominski Park

23. If the home plate umpire hinders the catcher's attempt to throw the ball, it is:

a. tough luck
b. umpire interference
c. an error
d. obstruction

24. Home plate is:

a. in foul territory
b. pointed toward the pitcher
c. five sided
d. none of the above

25. The base path is defined as:

a. the imaginary line 3 feet on either side of the direct line between the bases
b. only the direct line between the bases
c. the space between the foul line and the direct line
d. the distance between first base and third base

General Knowledge Quiz

Please answer the following open-ended questions as best as you possibly can. Write your answers in the spaces provided.

1. Who invented the light bulb? _____

2. How many continents are there? _____

3. What is the freezing point of water (F)? _____

4. In what state is Mt. Rushmore? _____

5. What is the capitol of Alabama? _____

6. Name the largest ocean _____

7. What oceans border Australia? _____

8. Who is the president of France? _____

9. How many great lakes border Michigan? _____

10. What year did the US enter WWII? _____

11. What does PLO stand for? _____

12. How many senators are there in congress _____

13. Who ran for president against Lyndon Johnson in 1964 _____

APPENDIX G

Behavioral Incident Checklist
No Accountability Condition

SUBID_____

Employee Rating Study

Using the following checklist, place a mark by the items which you recall having seen the secretary perform. It may help you to try to think about the specific context in which the situation described took place (i.e. who was there, where did it take place, etc).

After you finish filling out this form please turn it over and place it in front of you.

_____ Makes sure she takes accurate phone messages

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Develops more efficient office procedures

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not do work in the proper priority order

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Is not knowledgeable of the phone system and makes mistakes using it.

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does favors for other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Provides a student with detailed information about university hiring procedures in order to answer their question

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Gets angry at other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not get a professor's work typed on time

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Tells a student that she is too busy to answer their question and that they should ask one of the other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Gets the pages out of order when stapling some tests for a professor

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not have knowledge of all of the professors and the TA's in the department

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Demonstrates detailed knowledge about the various idiosyncrasies of the professors in the department

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Refuses to do work for a professor that she feels was given to her on short notice

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ She is friendly and courteous to professors when they ask her if they can use the office equipment

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Does not let another secretary see confidential information that she has been trusted with

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Volunteers to do some work for a professor that is not part of her regular job description

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Refuses to change her plans to accommodate the other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Develops a method to keep track of all the assignments that she is working on and when they are to be finished

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Does very sloppy and inefficient filing

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Compliments another secretary on how organized she is

. . . . . .
. . . . . .
. . . . . .
. . . . . .

PLEASE TURN THE QUESTIONNAIRE OVER WHEN YOU ARE FINISHED

Behavioral Incident Checklist
Accountability Condition

SUBID_____

Employee Rating Study

Using the following checklist, place a mark by the items which you recall having seen the secretary perform. It may help you to try to think about the specific context in which the situation described took place (i.e. who was there, where did it take place, etc).

After you finish filling out this form please turn it over and place it in front of you. After everyone in the group is finished you will be asked to state your answers to the rest of the group members and justify that you indeed saw the behaviors that you marked.

_____ Makes sure she takes accurate phone messages

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Develops more efficient office procedures

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not do work in the proper priority order

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Is not knowledgeable of the phone system and makes mistakes using it.

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does favors for other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Provides a student with detailed information about university hiring procedures in order to answer their question

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Gets angry at other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not get a professor's work typed on time

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Tells a student that she is too busy to answer their question and that they should ask one of the other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Gets the pages out of order when stapling some tests for a professor

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not have knowledge of all of the professors and the TA's in the department

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Demonstrates detailed knowledge about the various idiosyncrasies of the professors in the department

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Refuses to do work for a professor that she feels was given to her on short notice

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ She is friendly and courteous to professors when they ask her if they can use the office equipment

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Does not let another secretary see confidential information that she has been trusted with

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Volunteers to do some work for a professor that is not part of her regular job description

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Refuses to change her plans to accommodate the other secretaries

. . . . . .
. . . . . .
. . . . . .
. . . . . .

_____ Develops a method to keep track of all the assignments that she is working on and when they are to be finished

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Does very sloppy and inefficient filing

. . . . . .
. . . . . .
. . . . . .
. . . . . .

____ Compliments another secretary on how organized she is

. . . . . .
. . . . . .
. . . . . .
. . . . . .

PLEASE TURN THE QUESTIONNAIRE OVER WHEN YOU ARE FINISHED