INFORMATION TO USERS

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

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

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

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

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
RICE UNIVERSITY

EFFECTS OF PROCESSING METHOD, PERFORMANCE PATTERN, AND TIME PRESSURE ON PERFORMANCE RATINGS

by

Annette C. Spychalski

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY

Approved, Thesis Committee

Dr. Miguel Quiñones, Assistant Professor, Department of Psychology

Dr. Steven Currall, Assistant Professor Administrative Sciences and Psychology
Jones Graduate School of Administration

Dr. Robert Dipboye, Professor, Chair, Department of Psychology

Dr. David Schneider, Professor, Department of Psychology

Houston, Texas

May, 1997
ABSTRACT

EFFECTS OF PROCESSING METHOD, PERFORMANCE PATTERN, AND TIME PRESSURE ON PERFORMANCE RATINGS

by

Annette C. Spychalski

One hundred seventy-eight subjects observed videotaped incidents of a secretary performing on the job and rated her performance across four performance dimensions (and overall performance). Rating accuracy, behavior recognition accuracy, and rating level were measured under conditions of viewing improving or deteriorating performance, with behavior-based or impression-based processing methods, and with or without time pressure to complete the rating task. Results display robust main effects for both processing method and performance pattern on rating level. Raters viewing performance that improved over time gave the secretary lower ratings than raters viewing performance that deteriorated over time. Raters using behavior-based processing methods rated performance higher than raters using impression-based processing methods during the rating task. Processing method influenced rating accuracy for two of the performance dimensions; effects were mixed. Small relationships among measures of memory discrimination and response bias and rating accuracy and performance pattern were also discovered. Time pressure had no stable effect on any of the dependent variables. The results illustrate the pervasiveness of primacy effects in performance rating tasks. They also underscore the importance of standardizing rating procedures that are used to compare performance of different individuals. Additional research in applied settings is needed to capture the organizational influences on the performance rating process. Because of the complexities involved in performance rating systems in
organizations, it may be prudent to change the typical way in which supervisor input about employee performance is used in human resources decisions.
Acknowledgments

Sincere thanks are due to many who have supported me personally and professionally during my graduate career. To Mickey, an optimistic, patient, and encouraging advisor. I was his first doctoral candidate; I am sure that I am far from the last. To the rest of the committee---Steve Currall, Bob Dipboye, and Dave Schneider. Your many collective hours of toil have made this dissertation a great project.

To my new husband, Dwayne Ehlert: you make a terrific research assistant! You have been a supportive partner and a vociferous advocate of my professional development since the time we met. This dissertation is one of many challenges that I believe we can overcome together.

To my parents, Roman and Lucy: I recognize and appreciate the pains that you have taken to help me get where I am today. You have consistently demonstrated your values regarding education by doing what was necessary to secure the best education possible. That message has paid off for me and will continue to do so.

Finally, this project is written in fond remembrance of Edward W. Windsor (1924 to 1996).
Table of Contents

Introduction 1

Measures of Rating Quality 4

Cognitive Processes Involved in Rating Performance 6

Perceptual Biases Influencing the Rating Process 9

Effect of Rating Goal on the Rating Process 12

Use of Summary-based Versus Impression-based Performance Ratings 14

Effects of Time Pressure on Performance Ratings 18

Effects of Processing Method and Time Pressure on Behavior Recognition Accuracy for Early and Late Behavior 20

Effects of Processing Method on Rating Accuracy 23

Level of Performance Rating 25

Effects of Processing Method, Performance Pattern, and Time Pressure on Rating Level 29

Method 30

Participants 30

Design 30

Materials 30

Measures 35

Procedure 38

Results 42

Manipulation Check 42

General Variable Characteristics and Relationships 42
Tests of Hypotheses  50
Discussion  70
Alternative Explanations and Study Limitations  77
References  82
Appendix A
Critical Incidents  91
Appendix B
Experiment Materials  95
<table>
<thead>
<tr>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Predictions</td>
<td>31</td>
</tr>
<tr>
<td>Descriptive Statistics for Major Study Variables</td>
<td>43</td>
</tr>
<tr>
<td>Intercorrelations for Major Study Variables</td>
<td>45</td>
</tr>
<tr>
<td>Multiple Analysis of Variance for Memory Discrimination</td>
<td>52</td>
</tr>
<tr>
<td>Multiple Analysis of Variance for Response Bias</td>
<td>53</td>
</tr>
<tr>
<td>Analysis of Variance for Distance Accuracy</td>
<td>54</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Overall Performance</td>
<td>55</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Accuracy for Job Knowledge</td>
<td>56</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Accuracy for Organizational Ability</td>
<td>57</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Accuracy for Dealing With Faculty and Students</td>
<td>58</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Working Cooperatively with Other Secretaries</td>
<td>59</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Overall Performance</td>
<td>61</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Job Knowledge</td>
<td>62</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Organizational Ability</td>
<td>63</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Dealing With Faculty and Students</td>
<td>64</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Level for Working Cooperatively with Other Secretaries</td>
<td>65</td>
</tr>
<tr>
<td>Analysis of Variance for Rating Accuracy for Job Knowledge (Time Pressured Subjects Only)</td>
<td>68</td>
</tr>
</tbody>
</table>
List of Figures

Predicted Effect of Time Pressure and Processing Method on Behavior Recognition Accuracy for Early Behaviors 22

Predicted Effect of Time Pressure and Processing Method on Behavior Recognition Accuracy for Late Behaviors 24

Predicted Effect of Time Pressure and Processing Method on Rating Accuracy 26

Predicted Effect of Time Pressure, Processing Method, and Performance Pattern on Rating Level 32

Effect of Time Pressure, Processing Method, and Performance Pattern on Rating Level for Dealing with Faculty and Students 66

Effects of Processing Method and Time Pressure on Rating Accuracy for Job Knowledge (Time Pressured Subjects Only) 69
Effects of Processing Method, Performance Pattern, and Time Pressure on Performance Ratings

Employers rely on human ratings of employee performance in a multitude of work-related contexts. In fact, rating others' performance may be the most typical way to gather performance effectiveness information (Landy & Farr, 1980; Schmitt & Klimoski, 1991). Performance appraisal has been examined from a number of perspectives, including rating format, contextual influences on performance ratings, psychometric characteristics of rating distributions, and rating process variables. Researchers have discovered problems with reliance on supervisor-generated performance ratings in making human resources decisions: the ratings often have the potential for being inaccurate, unreliable, and biased estimates of performance (e.g., Dickinson, 1987; Landy & Farr, 1980; Longenecker, Sims, & Gioia, 1987; Nathan & Alexander, 1985). Because employers have relied on ratings of employee performance to make important business decisions, and because these decisions can only be as good as the ratings are sound, researchers have worked to find the conditions under which high-quality ratings of employee performance are made.

There are numerous criteria against which to evaluate performance ratings, such as reliability, freedom from rating error, practicality, and effect on ratee performance (Murphy & Cleveland, 1991; Murphy & Cleveland, 1995). Conclusions regarding the value of different rating techniques and rater training programs on performance ratings can vary depending on how the ratings are evaluated (Becker & Cardy, 1986). While there is no agreement among performance appraisal researchers regarding the most useful
criterion against which to evaluate performance ratings, direct and indirect measures of rating accuracy are among the most popular (Murphy & Cleveland, 1995). In general, accuracy describes the extent to which the target’s actual performance corresponds to the evaluation that the rater assigns to the target.

Another way to conceptualize accuracy is to find how precisely raters can recall what the performer did and did not do during the rating period (behavior recognition accuracy). Researchers have found that accuracy in describing what performers did during the rating period is associated with accuracy in rating performer behavior (Murphy, Garcia, Kerkar, Martin, & Balzer, 1982; Sanchez & De La Torre, 1996). However, all types of rating accuracy are not related to behavior recognition accuracy. Therefore, it seems important to understand raters’ skill in this area, and the way in which this skill is related to skill in creating accurate ratings.

The accuracy of performance ratings is critical: it represents the “upper limit” for the soundness of the human resource decisions that are grounded in the ratings. However, it is important to remember that the decisions are based not on the accuracy of the performance rating, but on the absolute level of the rating itself. Highly rated performers may be offered raises or promotions; those rated at marginal levels may be put on probation, sent to training, or terminated. Therefore, rating level will be examined in this study. This variable may also help clarify results found with respect to rating accuracy and behavior recognition accuracy.

This study will examine performance ratings along the dimensions of rating accuracy, behavior recognition accuracy, and rating level. These measures reflect
meaningful characteristics of performance that are of both theoretical and practical interest.

Researchers have discovered a number of factors that potentially influence the accuracy of performance ratings, such as time delay between performance observation and performance rating (Murphy & Balzer, 1986), performance consistency of the target (Youtz-_padgett & Ilgen, 1989), and purpose of rating (Murphy et al., 1989). One such factor is the information processing methods used to create the performance rating. Specifically, ratings assigned after an “objective” record of behavior is created are thought to be more accurate than those formed in a more intuitive manner (Murphy et al., 1982; Sanchez & De La Torre, 1996; Youtz- Padgett & Ilgen, 1989). This latter method is generally based on impressions that the rater creates of the performer during the performance period. In this study, a systematic test of the proposed advantage of the former method versus the latter method will be conducted.

This study examines how the quality of ratings are affected by cognitive processes that raters use in evaluating performance. It will review the cognitive and social cognition literature relevant to the task of observing and rating performance. Rater tendencies and biases, as well as ways in which situational factors affect the rating process, will be explored.

Another goal of this study is to examine performance rating processes under conditions that simulate actual rating circumstances in organizations. For example, it is often the case that ratees’ performance changes over time (e.g., Borman, 1991). Furthermore, raters are not typically encouraged to make performance rating a high-
priority task or to spend much time observing and rating performance. This study adds realistic complexity to the rating task by examining it when performance levels change over time and when there is not enough time to thoroughly consider all possible sources of performance information.

**Measures of Rating Quality**

As previously mentioned, there is a variety of ways to examine the quality of performance ratings. In this study, accuracy is the general focus. However, there are a number of ways in which to conceptualize this construct. Previous research has demonstrated that various types of rating accuracy are differentially related (Murphy et al., 1982; Sanchez & De La Torre, 1996; Sulsky & Balzer, 1988). Two types will be measured in this study: rating accuracy and behavior recognition accuracy. Each is described below.

**Rating accuracy.** Rating accuracy, generally considered the extent to which a rater assigns a numerical rating that reflects the ratee's actual level of performance, has been conceptualized and operationalized in numerous ways (e.g., Borman, 1977; Cronbach, 1955; Davison, 1985). There is no consensus among performance appraisal researchers regarding the best way to measure rating accuracy (Murphy & Cleveland, 1995). Distance accuracy will be used in this study; it is one of the few accuracy measures that can be applied to rating tasks involving a single performer (Sulsky & Balzer, 1988). It is based on the difference between a rater's rating and the corresponding true score estimate (e.g., based on subject matter expert ratings).
Behavior recognition accuracy. This measure estimates the extent to which raters have encoded and can retrieve what targets actually did during the rating period. In their study, Murphy and colleagues (1982) found evidence that raters who can accurately report the behaviors that targets performed also produce more accurate overall performance ratings. Similarly, Sanchez and De La Torre (1996) found a positive relationship between subjects’ ability to recognize behaviors that targets did and did not perform and their ability to accurately rate performance. Signal Detection Theory provides a mechanism to measure the extent to which raters attribute behaviors to performers accurately. This approach separately assesses raters’ response bias (i.e., the extent to which they incorrectly attribute behaviors that did not happen to a performer) and their memory discrimination (i.e., the extent to which they correctly attribute only the behaviors that did occur to a performer)(Lord & Maher, 1991). Response bias and memory discrimination are constructed with two basic behavior recognition measures: hit rate and false alarm rate. Hit rate refers to the proportion of times an individual correctly indicates that a behavior that a target performed did occur. False alarm rate refers to the proportion of times that the individual claims that a behavior did occur when it actually did not. Response bias is calculated by summing an individual’s hit rate and false alarm rate; memory discrimination is calculated by subtracting the latter proportion from the former. In this study, the ability of raters to accurately recognize what the performer did and did not do will be measured and compared with their ability to make accurate performance ratings.
Influences on accuracy. Performance appraisal research has identified a plethora of factors that can affect rating accuracy. Cognitive factors, interpersonal factors, and situational factors have the potential to enhance or detract from the accuracy of performance ratings (Murphy & Cleveland, 1991; 1995). Researchers have consistently found that the purpose of the appraisal can exert tremendous influence over behaviors that raters notice and process as they form their ratings. This affects the numerical evaluation that is assigned to those behaviors (Cardy & Keefe, 1994; DeNisi, Cafferty, & Meglino, 1984; Williams, DeNisi, Blencoe, & Cafferty, 1985), as well as rating accuracy (Murphy et al., 1992; Sanchez & De La Torre, 1996; Youtz-Padgett & Ilgen, 1989).

To better understand how accuracy is affected during the rating process, one must first understand the cognitive factors playing a role in performance evaluation. A description of the cognitive steps involved in creating performance ratings follows.

Cognitive Processes Involved in Rating Performance

Several cognitive tasks must be completed to create a performance rating: attending to/observing performance, organizing and storing performance information in memory, retrieving this information, and integrating it to form a performance judgment (DeNisi et al., 1984; Feldman, 1981; Ilgen, Barnes-Farrell, & McKellin, 1993). Cognitive and social cognition researchers generally agree that raters can introduce bias to performance ratings at any of these stages, and that bias introduced at any of the stages affects subsequent stages (DeNisi et al., 1984). Therefore, once bias has contaminated the rating process, it is almost impossible to remove.
**Performance observation and storage.** Rather than simply committing representations of specific behaviors into memory, raters often organize and interpret this information with respect to cognitive categories or schemas (Cardy, Bernardin, Abbott, Senderak, & Taylor, 1987; DeNisi et al., 1984; Fiske & Taylor, 1991; Foti & Lord, 1987; Sulsky & Day, 1994). Schemas are cognitive structures that contain knowledge of a concept or stimulus, including its characteristics and relationships among the characteristics (Fiske & Taylor, 1991; Foti & Lord, 1987). For example, we have schemas about particular types of people (e.g., politicians, athletes), events (e.g., going to a movie or to a wedding), and activities (e.g., entertainment, work). These schemas summarize specific behavioral information at a more general level, and allow us to fill in "gaps" (with reasonable guesses) where they exist (Feldman, 1981; Fiske & Taylor, 1991; Foti & Lord, 1987). Using such structures to summarize behavior is considered more cognitively efficient than maintaining an unintegrated "data base" of performance information. For this reason, summaries, not the behaviors on which they are based, are most frequently used in subsequent storage, recall, and performance information integration tasks (DeNisi et al., 1984; Feldman, 1981).

Therefore, the nature and quality of these summaries have tremendous ramifications for the rating process. To the extent that they are comprehensive and representative synopses of performer behavior, the final ratings can be accurate and sound. To the extent that they are biased over- or under-generalizations of performer behavior, the ratings will be necessarily inaccurate (Mount & Thompson, 1987; Nathan & Alexander, 1985).
Behavior retrieval and evaluation. It is virtually impossible to separate evaluative components from earlier stages of the performance rating process (Bernardin & Beatty, 1984; Ilgen et al., 1993; Nathan & Alexander, 1985). Hastie and Park (1986) found that raters make summary judgments as they observe targets’ behavior instead of after all the relevant behaviors have been observed. When new judgments about performance are needed, they are based on the previously created evaluations instead of on the previously observed behaviors. Hulland and Kleinmuntz (1994) point out that retrieval of a previously created summary impression prevents the formation of a new evaluation based on other (or additional) data, at least in the short term. Therefore, ratings made in this way may be insensitive to behaviors emitted toward the end of the rating period that suggest their modification.

After summary impressions have been formed, raters rely on them to a greater extent than the behaviors on which they were based (e.g., DeNisi et al., 1984; Mount & Thompson, 1987; Sulska & Day, 1994). This may be because memory for the summary impressions decays at a slower rate than memory for the behaviors themselves (Ilgen et al., 1993; Murphy & Balzer, 1986). Another reason that raters store and recall performance information in summary impression form is that it is cognitively easier than carefully weighting and combining behavioral information each time a summary impression is needed. When one does not have an impression of a target in memory for use, one must retrieve relevant behaviors, then perform a computational function on those behaviors before arriving at a summary impression (Lichtenstein & Srull, 1987). This operation takes time and effort (Fiske & Taylor, 1991). Therefore, it should be no
surprise that raters prefer to use summary impressions as a basis for performance ratings. Perceiving and processing performance information in terms of summary impressions can minimize information processing and storage demands (Behling, Gifford, & Tolliver, 1980; Fiske & Neuberg, 1990; Smith Adams, & Schorr, 1978). In particular, raters tend to use summary impressions under high cognitive load. Murphy and Balzer (1986) found that increasing memory demands on raters led them to form and store overall impressions of ratees rather than remember actual behaviors.

Summary impressions reflect both the rater’s knowledge of a ratee’s performance and other associated attitudes. Unfortunately, some of these attitudes may represent biases unrelated to performance per se. These initial summary impressions (biases included) affect subsequent judgments made about the performer (Balzer, 1986; Landy & Trumbo, 1980). Therefore, it is important to prevent these biases from affecting the ratings as much as possible.

Rater biases due to efforts to conserve cognitive energy represent one source of distortion in the rating process when raters rely on summary impressions of the ratee. Another source stems from more fundamental information perception tendencies. These biases affect raters relying on summary impressions, as well as those who try to maintain an accurate behavioral memory of the performer as a basis for performance ratings. They are described in the following section.

**Perceptual Biases Influencing the Rating Process**

Consciously or not, raters play an active role in determining which behaviors they notice, encode, store, retrieve, and use as a basis for performance ratings (DeNisi, et al.,
Decisions regarding which behaviors warrant attention and further processing have a significant impact on the performance ratings that raters construct.

Researchers have consistently found that human raters are prone to particular cognitive errors as they take in performance information. The two most typical errors are primacy and recency effects.

**Primacy effects.** A pernicious bias that influences many memory tasks is the "primacy effect". That is, stimuli that are encountered early in an interaction session are better remembered than other items in that session. In terms of person perception tasks, the primacy effect is what helps create a long-lasting and influential "first impression" of a person. Researchers have found that summary impressions of others are colored more heavily by behaviors encountered early in the observation period compared to those in the middle or at the end of the observation period. Consistent with this tendency, performance appraisal researchers have often found primacy effects when raters create performance ratings (e.g., Anderson, 1989; Freund, Kruglanski, & Shpitzajen, 1985; Kruglanski & Freund, 1983; Lichtenstein & Srull, 1987). In other words, performance ratings frequently reflect behavior early in the appraisal period too strongly.

**Recency effects.** A second consistent bias applicable to memory and perception tasks is the "recency effect". This effect occurs when items encountered at the end of an observation period are better remembered than other items. In other words, stimuli encountered most recently are easier to remember than those encountered earlier in time. Researchers studying person observation tasks have discovered that when observers are
asked to recall or recognize behaviors that a target performed, behaviors viewed late in
the observation period are more frequently recalled and more accurately recognized than
those in the beginning or middle of the rating period (e.g., Anderson, 1989; Lichtenstein
& Srull, 1987). For example, in Lichtenstein and Srull's (1987) study, subjects who were
asked to recall performer actions before evaluating their performance tended to base the
performance rating more heavily on the most recent behaviors than on those observed
earlier in the rating period. However, those asked to first evaluate the target produced
ratings that were related more strongly to behaviors encountered early than to those
encountered later.

Before leaving this topic, it is important to note that primacy effects are generally
found in performance evaluation tasks, and that recency effects are common to
performance memory tasks (i.e., behavior recall or recognition) (Anderson, 1989). In his
chapter, Anderson (1989) explained this distinction by differentiating between a
functional (i.e., social) and traditional memory concept. Steps in performance evaluation
relying upon summary-based information processing methods use functional memory.
Functional memory dictates that early behaviors are remembered better and are more
heavily relied upon during performance evaluation than is later performance information.
Traditional memory processes refer to those requiring relatively strict data encoding and
retrieval, such as those focused on remembering actual behavior. Traditional memory
processes cause later behaviors to be better remembered and more heavily relied upon
than earlier behaviors. They may come into play when raters deliberately focus on
remembering performer behaviors before assigning a performance rating.
Effect of Rating Goal on the Rating Process

The single most influential situational factor in the rating process may be its purpose (Murphy & Cleveland, 1991). The original goal or focus of observation (consciously chosen or not) strongly influences which behaviors capture raters’ attention, those they encode, and those they store and retrieve for later use (e.g., DeNisi et al., 1984; Foti & Lord, 1987; Woehr & Feldman, 1993). When observing others, individuals focus almost exclusively on information that is directly applicable to their current needs. Humans are “cognitive misers”, sampling behavior in a highly selective way (Bernardin & Beatty, 1984; Fiske & Taylor, 1991). Behaviors that may be useful for observation goals other than those identified by the rater at the beginning of the observation period are unlikely to receive attention, much less more elaborate cognitive processing; they are basically “lost forever” (Balzer, 1986). If a rater is asked to evaluate a performer in a way that was not part of the original observation goal, he or she will have little to no foundation for that judgment. For example, if a rater needs to create an overall performance judgment regarding a performer, he or she will likely form a summary-level judgment of that individual. If later the rater is asked to describe the performer’s specific performance strengths and development needs, the rater is unlikely to be able to perform this task accurately. In other words, unless the observer’s initial processing method is compatible with subsequent evaluation goals, the ratings for the second purpose are likely to be poorly founded—they will be based on inferences drawn from the behaviors observed for the original purpose. To the extent that the original and subsequent
observation purposes disconnect, the rater will have no behavioral information on which to base the later ratings.

Researchers have discovered that rating goal can change the cognitive processes that a rater uses to evaluate a target (DeNisi et al., 1984; Foti & Lord, 1987; Murphy, Philbin, & Adams, 1989; Williams et al., 1985; Woehr & Feldman, 1993; Wyer & Srull, 1980; Wyer & Srull, 1986). For example, Williams and colleagues (1985) propose that performance information is encoded and weighted differently according to the purpose of the rating. Wyer and Srull (1980, 1986) advance that the way in which information about a target is represented cognitively is contingent upon the initial processing objectives of the observer. If an individual observes target performance with the intent of making a judgment about the target, he or she will form an impression in addition to "registering" some behavioral information. That impression will be "stored" in a separate "registry" than the behavioral information. When the observer needs to use that impression, only the impression, and not the behaviors on which it is based, will be recalled (Lichtenstein & Srull, 1987). However, if the observer has no intention to evaluate the target during the observation time, a subset of the behaviors will be relayed to memory.

Rating goal affects the cognitive processes involved in the rating task from its outset. For example, it determines the extent to which raters rely on particular schemas to organize performance information and that to which they depend on the behaviors themselves (Fiske & Taylor, 1991). It can affect the choice of schema that a rater uses to categorize and interpret the information (DeNisi et al., 1984; Foti & Lord, 1987) as well as influence the type of information that is sought and how it is stored (DeNisi et al., 1984). For example, when raters create evaluations for administrative purposes, they tend
to use trait-oriented schemas, and infer personal characteristics from the performer's appearance and actions. When they create evaluations for feedback purposes, they are likely to use goal-oriented schemas, and attend to and store performer actions in relation to the goal that the performer is trying to achieve. When raters anticipate having to interact with ratees (e.g., in a performance feedback session), they tend to process behavioral information from the ratees' perspective. Under these conditions, they also engage in deeper information processing than when no interaction is anticipated (Fiske & Taylor, 1991).

To summarize, rating goal influences the way in which performance ratings are formed. The goal affects the way in which the rater observes the performer and the performance information that is used as a basis for performance ratings. In general, research suggests that when raters adopt goals that lead to summary-based processing, their ratings will be less accurate than when they pursue goals that are conducive to behavior-based processing. More detail about the advantages and disadvantages of behavior-based and summary-based rating methods is presented below.

**Use of Summary-based Versus Behavior-based Performance Ratings**

It has been suggested that basing performance ratings on actual behaviors is superior to making them after organizing and maintaining information with summary impressions. Of course, this assertion assumes that maximizing accuracy is the goal of the rating process (Fiske & Neuberg, 1990; Fiske & Taylor, 1991). In their study, Murphy and Jones (1993) found that the use of categories summarizing performance led
to decreased accuracy of memory for whether an actor performed negative and positive performance behaviors.

However, a robust finding that supports the use of an organizing mechanism (e.g., summary-based schema) in person perception tasks is that people recall information much better with an organizing theme than when they have simple memory goals as they observe a target (Hamilton, 1981). In his review, Hastie (1981) found that both schema-consistent and schema-inconsistent information is remembered better than schema-irrelevant information. This suggests that schema usage in person perception tasks should be encouraged.

More recent research has discovered that the memory advantage gained with schema usage comes at a cost. Ilgen and Feldman (1983) found that, in addition to increased recall of information related to a schema, the number of memory errors based on schematically related biases increased when people used schemas to organize the information they gathered. Recall of information was most accurate when perceivers adopted information memory goals, which kept the ratings free of such biases.

Using summary impressions to form performance ratings has other disadvantages as well. Performance ratings based on previous summary impressions are sensitive to the primacy effect (Fiske & Taylor, 1991; Lichtenstein & Srull, 1987; Srull & Wyer, 1989). Furthermore, these ratings are difficult to change, even if the target emits behavior suggesting an adjustment in the ratings. Lichtenstein and Srull (1987) found that performance judgments created during the observation period (i.e., based on initial overall impressions) were more extreme and were held more confidently than those made
after all relevant behaviors had been observed. Therefore, these judgments seem valid only when performers display a steady level of performance throughout the rating period.

In spite of their weaknesses, ratings based on summary impressions of performer behavior can accurately reflect the individual’s performance level. As mentioned, this is most likely to happen when a performer’s early behavior deviates little from his or her overall average level of performance. A second factor to consider when evaluating the appropriateness of ratings based on summary impressions is the purpose of the ratings. One must restrict their use to decisions about the performer based on his or her overall level of performance. For example, an overall rating of effectiveness may be sufficient to make a hiring decision, but it would not be particularly helpful if an individual wished to pinpoint specific areas for developmental focus (Murphy et al., 1982).

Although the utility of impression-based evaluations over behavior-based evaluations has not been rigorously tested, there is evidence favoring the use of behavior-based ratings. For example, Youtz-Padgett and Ilgen (1989) conclude that a more behavior-based approach to information processing and storage should facilitate rating accuracy. Murphy and colleagues (1982) discovered that raters who are accurate in their reports of frequency of behaviors emitted by performers are more accurate in their ratings of the performers on the corresponding performance dimensions than are raters less accurate in behavioral frequency estimates. That is, raters who most accurately recalled the frequency of target behavior were also most accurate in their performance ratings. Sanchez and De La Torre (1996) also found positive relationships between raters’ ability to recognize performer behavior and their ability to rate the behavior accurately.
Consistent with this notion, researchers have found improved psychometric quality of ratings when raters use techniques to improve their memory for ratee behavior (e.g., notes, performance diaries) (Bernardin & Walter, 1977). To maximize rating accuracy, raters should be encouraged to notice and remember actual behavior rather than rely upon impressions of the ratee as a basis for performance ratings.

On the down side, behavior-based ratings are vulnerable to recency effects (Anderson, 1989; Lichtenstein & Srull, 1987). However, raters using this method to evaluate performance attend to a larger portion of performer behavior than do raters using an impression-based method. Therefore, it is expected that use of this process will lead to more accurate ratings that will the impression-based method.

Raters may consciously choose to use a particular rating method when evaluating performance. However, the rater's goal during the rating process influences the information processing approach that is used (Foti & Lord, 1987; Hastie & Park, 1986; Lichtenstein & Srull, 1987). For example, if raters want to make a promotion decision based on the ratee's overall performance level, they are likely to evaluate the performer as his or her performance is observed. This circumstance is likely to evoke impression-based rating processes. If raters want to make specific recommendations regarding performance improvements after the rating period, they are likely to pay more attention to specific ratee behavior across the rating period (i.e., use behavior-based processing methods).

Overall, behavior-based ratings are prone to recency effects. Impression-based ratings are plagued by primacy effects. While ratings made with the former method
should more accurately reflect performer behavior than those created with the latter process, this relationship may depend on a number of contextual factors. For example, temporal delay between performance observation and performance rating (Murphy & Balzer, 1986) and implicit rater theories regarding performance inferences (Nathan & Alexander, 1988) have been found to affect the way in which information processing method influences the accuracy of performance ratings. In their review, DeNisi and colleagues hypothesize that time pressure in the rating task may affect performance ratings (1984). This idea was confirmed by Ilgen et al. (1993), who found that ratings became more accurate as the time to observe performers increased. Therefore, the effect of time pressure on the rating task is examined in this study.

**Effects of Time Pressure on Performance Ratings**

Time allotted to the rating task has the potential to influence performance ratings. Research suggests that people use fewer pieces of information, less complicated decision rules, and become more susceptible to primacy effects when making decisions under time pressure (Freund et al., 1985; Kruglanski & Freund, 1983; Payne, Bettman, & Johnson, 1988; Wright, 1974). Researchers have yet to examine the effects of these tendencies on the performance rating process.

In their review, Ilgen and colleagues (1993) found that raters who spent less time observing target performance produced ratings less accurate than raters who spent more time observing performance. Time constraints may limit the opportunities that raters have to observe performance. Therefore, raters should spend adequate time collecting and processing ratee information when the accuracy of the ratings is particularly critical.
One strategy shift that raters may use under time demands involves forming and storing overall evaluations of targets instead of remembering performer behavior (Hulland & Kleinmuntz, 1994; Murphy & Balzer, 1986). This may be problematic under circumstances requiring ratings based on specific behavioral instances or memory for the instances on which the rating is based. For example, an employee may approach the supervisor for examples to illustrate a surprisingly high or low performance rating. If the performance ratings had been based upon a summary impression rather than on specific instances of performance, the supervisor may not be able to satisfy the individual’s request.

Time pressure exaggerates the effects of some perceptual biases, such as the primacy effect (Freund et al., 1985; Kruglanski & Freund, 1983). Performers may be “permanently” labeled with the performance level that they had originally displayed during the rating period when raters make evaluations under time pressure. Thus, neither improvements nor decrements in performance may be detected by an individual making performance ratings under time pressure. The result would be that individuals whose performance decreased would be assigned too high a rating; individuals whose performance increased would be given ratings too low. Only in cases where performance remained stable would ratings made under time pressure be unaffected by primacy effects.

It is interesting to note that the processing methods motivated by time pressure during evaluative tasks (i.e., use of less information, creating overall evaluations, susceptibility to primacy effects) are quite similar to those used when using impression-
based processing methods to rate performers. It is possible, then, that time pressure does not meaningfully change the way in which impression-based raters evaluate performance. However, the effects of time pressure on behavior-based ratings may be more substantial. Without time pressure applied to the rating task, behavior-based raters use rating techniques quite different from those utilized by impression-based raters.

One major goal of this study is to find how time pressure affects raters who use behavior-based rating processes versus those using impression-based processes in the performance evaluation task. This study offers the opportunity to study the effects of time pressure without confounding those effects with effects created by differences in the amount of information on which a rater bases the ratings (all raters will be exposed to the same performance information for the same amount of time in this study).

Time pressure is expected to truncate the information search and processing stages of performance evaluation and to make ratings more sensitive to the primacy effect. It is expected to differentially affect raters using behavior-based and impression-based rating processes. Ways in which it may influence each type of accuracy (for raters using each processing method) are described in the subsequent section.

Effects of Processing Method and Time Pressure on Behavior Recognition Accuracy for Early and Late Behavior

As previously discussed, tasks involving performance evaluation are often plagued with primacy effects, and tasks involving more straightforward memory processes are biased by recency effects. Therefore, when raters use impression-based rating processes (i.e., those evaluative in nature), they should demonstrate better behavior
recognition accuracy for early behaviors than for later behaviors. When raters use behavior-based rating processes (i.e., those based on memory), they should display better behavior recognition accuracy for more recent behaviors than for those viewed earlier in the observation period.

When time pressure is applied to the performance rating task, somewhat different results are expected. As discussed, raters forming performance evaluations under time pressure are expected to use less information, make evaluations at an overall performance level, and be more sensitive to primacy effects than raters who perform the rating task at a leisurely pace.

This approach to the rating task is much like that used by raters who use impression-based rating processes, but quite different from those using behavior-based processes (without time pressure). Therefore, behavior-based and impression-based raters should behave more similarly when rating under time pressure than when they create ratings at a leisurely pace. Specifically:

*Hypothesis 1*: Time pressure will interact with processing method in its effects on behavior recognition accuracy for early behaviors. When raters do not feel time pressure to complete the rating task, raters using impression-based rating processes will demonstrate better behavior recognition accuracy for early behaviors than will raters using behavior-based rating processes. Under time pressure, both behavior-based and impression-based raters will demonstrate relatively high levels (and equally high levels) of behavior recognition accuracy for early behaviors (See Figure 1).
Hypothesis 1: Predicted Effect of Time Pressure and Processing Method on Behavior Recognition Accuracy for Early Behaviors
Hypothesis 2: Time pressure will interact with processing method in its effects on behavior recognition accuracy for late behaviors. When raters do not feel time pressure to complete the rating task, raters using behavior-based rating processes will demonstrate better behavior recognition accuracy for late behaviors than will raters using impression-based rating processes. Under time pressure, both behavior-based and impression-based raters will demonstrate relatively low levels (and equally low levels) of behavior recognition accuracy for late behaviors (See Figure 2).

Effects of Processing Method on Rating Accuracy

Raters using impression-based rating processes are expected to rely heavily on behaviors exhibited early in the observation period. Later performance suggesting a change in the initial evaluation of behavior is likely to go unnoticed. This results in ratings too high for individuals whose performance deteriorates over time, and ratings too low for individuals whose performance improves over time. In both cases, the accuracy of ratings created in this manner will suffer due to overreliance on behaviors exhibited early on in the observation period.

Despite their susceptibility to recency effects, ratings made with behavior-based rating processes are expected to be more accurate than are ratings created with impression-based processes. This is because raters using behavior-based processes are expected to observe and consider a larger proportion of the behaviors exhibited during the rating period than are raters using impression-based processes.
Hypothesis 2: Predicted Effect of Time Pressure and Processing Method on Behavior Recognition Accuracy for Late Behaviors
Under time pressure, ratings created by both behavior-based and impression-based raters are expected to be heavily influenced by primacy effects. This will cause both groups of raters to produce ratings of similar accuracy. Unfortunately, it should also cause both groups to create somewhat inaccurate ratings, due to a neglect of behaviors appearing later in the observation period.

_Hypothesis 3_: Time pressure will interact with processing method in its effect on rating accuracy. When raters do not feel time pressure to complete the rating task, raters using behavior-based rating processes will create more accurate ratings than will raters using impression-based rating processes. When ratings are created under time pressure, the accuracy of raters using behavior-based rating processes will not differ from that of raters using impression-based rating processes. Rating accuracy will be relatively low (See Figure 3).

**Level of Performance Rating**

While the accuracy of performance ratings is important, one must not neglect the direct product of the rating process; the rating itself. Important decisions (e.g., hiring, firing, promotion) are often based on the absolute level of the performance rating. Rating level also has the potential to help explain the processes that raters use to arrive at their performance judgments. For example, two raters may observe an individual whose performance improved over time. The first rater might be biased by primacy effects and rate the performance too high. This individual might promote the ratee to a job requiring competencies that he or she does not have. The other rater may be biased by recency effects and assign a rating too low. This rater might recommend that the ratee attends
Hypothesis 3: Predicted Effect of Time Pressure and Processing Method on Rating Accuracy
costly training programs that develop skills that the ratee possesses. Although both raters might be equally inaccurate, the inaccuracy would be due to different information processing methods and biases. Furthermore, the consequences of each type of rating error are not identical. In this study, rating level will be measured and compared with accuracy.

Changes in performance level. Researchers have discovered that differences in performance level can translate into differences in performance ratings, even when average performance is held constant (DeNisi & Stevens, 1981; Jones, Rock, Shaver, Goethals, & Ward, 1968; Rowe, 1967). For example, Jones and colleagues (1968) found that individuals displaying performance that deteriorated over time earned higher performance ratings than those beginning at lower levels, but who improved over time. These authors conclude that evaluations of ability are made quickly, often with inadequate or incomplete information and that raters consider ability a fixed characteristic. This suggests that performers are rated primarily based on their initial behaviors.

Other researchers have found that when individuals display performance improvements over time, they are given higher ratings than when their performance worsens over time (DeNisi & Stevens, 1981). Furthermore, Rowe (1967) found that targets who were portrayed more positively at the end of a description were liked more than targets who were portrayed positively in the beginning, but less well later on. This suggests that people are influenced more by information they encounter late in an interaction session more than by information they absorb toward the beginning.
Potential reasons for inconsistency in results found by these researchers include major differences in study methodology and variable operationalization. For example, the performer task in Jones et al.'s study (1968) involved answering questions from an IQ test. Performance on this task was largely attributable to cognitive ability. The performer task in DeNisi and Steven's (1981) study involved sales performance (measured in dollar sales amounts). Success in this task depends on sales ability, motivation, luck, and other factors. These differences may have affected the study results. A second major difference in these studies is the dependent variables. In Jones et al.'s study (1968), subjects were asked to predict (and recall) the probability of successful performance. In DeNisi and Steven's (1981) work, the target was rated on a number of factors, including recommended raise and promotability. Therefore, it is difficult (and perhaps inappropriate) to draw conclusions across these studies. Overall examination of the research suggests that the evidence regarding the effects of changes in performance on performance ratings is mixed.

It is rarely the case that individuals maintain steady levels of performance on the job (Borman, 1991; Deadrick & Madigan 1990; Henry & Hulin, 1987). Therefore, it seems important to examine the effects of performance pattern on performance ratings. In this study, two types of performance pattern will be examined: steadily increasing performance and steadily decreasing performance. These are performance patterns that have been typically tested in other research on changing performance (e.g., Jones et al., 1968; Rowe, 1967; Scott & Hamner, 1975).
Effects of Processing Method, Performance Pattern, and Time Pressure on Rating Level

Cognitive processes used during the rating task may moderate the effect of performance fluctuations on rating level. For example, if raters employ behavior-based processes when forming performance judgments of others, ratings are likely to reflect the performer's most recent behavior to a greater extent than behavior emitted earlier (due to recency effects). Therefore, performers who improve over time are likely to be overrated; those whose performance declines over time are likely to be underrated.

When raters use impression-based approaches to creating performance ratings, it seems that, due to potential primacy effects, the ratings would simply reflect the level of performance that the individual emitted at the beginning of the rating period. Raters would make an initial categorization decision (e.g., good performer, poor performer), and evaluate the performer based on characteristics of that category. Therefore, performers who improve over time are likely to be underrated; those declining over time are likely to be overrated.

Under time pressure, both impression-based and behavior-based raters are expected to rely heavily on early behaviors when creating their ratings. As discussed, this results in underrating those who improve and overrating those whose performance deteriorates over time.

Hypothesis 4: Time pressure and processing method will interact with performance pattern in their effect on rating level. When raters do not feel time pressure to complete the rating task, behavior-based raters will rate performance that improves over time higher than will impression-based raters. Behavior-based
raters will rate performance that deteriorates over time lower than will
impression-based raters. When rating under time pressure, raters will create
ratings too low when performance improves over time. When performance
deteriorates over time, raters will assign ratings that are too high (See Figure 4).

Table 1 and Figures 1 to 4 summarize the study hypotheses in tabular and
graphical form.

Method

Participants

Two hundred twenty eight undergraduate psychology students in a small private
university in the Southwestern United States acted as subjects (178 for the main study; 50
for a pilot study). Participation in the experiment was rewarded with course credit.

Design

The experimental design was a $2 \times 2 \times 2$, between subjects factorial. The
independent variables were processing method (behavior-based, impression-based)
performance pattern (increasing, decreasing), and time pressure (no pressure, pressure).
The dependent variables were behavior recognition accuracy (for early behaviors and for
late behaviors), rating level, and rating accuracy.

Materials

A videotaped actress in the role of secretary in a university setting was used as
experimental stimuli. Eleven 20-to 150-second behavioral incidents of the secretary
performing on the job were displayed to each subject. Each incident depicted a self-
contained, on-the-job situation that provided behavioral information relevant to two or
<table>
<thead>
<tr>
<th>Rating situation</th>
<th>Recognition accuracy for early behaviors</th>
<th>Recognition accuracy for late behaviors</th>
<th>Rating accuracy</th>
<th>Rating level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No time pressure</strong></td>
<td><strong>Hypothesis 1:</strong> Impression-based raters will demonstrate better behavior recognition accuracy for early behaviors than will behavior-based raters.</td>
<td><strong>Hypothesis 2:</strong> Behavior-based raters will demonstrate better behavior recognition accuracy for late behaviors than will impression-based raters.</td>
<td><strong>Hypothesis 3:</strong> Behavior-based raters will be more accurate than will impression-based raters.</td>
<td><strong>Hypothesis 4:</strong> Behavior-based raters will rate improving performance higher than will impression-based raters. Behavior-based raters will rate declining performance lower than will impression-based raters.</td>
</tr>
<tr>
<td>Increasing performance</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td></td>
</tr>
<tr>
<td>Decreasing performance</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td></td>
</tr>
<tr>
<td><strong>Time pressure</strong></td>
<td>Recognition accuracy for early behaviors will be relatively high; it will not differ between impression-based and behavior-based raters.</td>
<td>Recognition accuracy for late behaviors will be relatively low; it will not differ between impression-based and behavior-based raters.</td>
<td>Behavior-based and impression-based will not differ in their rating accuracy. Rating accuracy will be relatively low.</td>
<td>Improving performance will be rated at a relatively low level; ratings will not differ between impression-based and behavior-based raters. Declining performance will be rated at a relatively high level; ratings will not differ between impression-based and behavior-based raters.</td>
</tr>
<tr>
<td>Increasing performance</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td></td>
</tr>
<tr>
<td>Decreasing performance</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td>Same as above.</td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 4: Predicted Effect of Time Pressure, Processing Method, and Performance Pattern on Rating Level
three of the following performance dimensions: Job Knowledge and Skill, Organizational Ability (ability to organize), Dealing with Faculty and Students, and Working Cooperatively with Other Secretaries. In some incidents, the secretary handled the situation quite effectively; in other situations, her performance was satisfactory; in some incidents, her responses were clearly below standard.

Behavioral incidents were presented in an order that created the performance patterns of interest in this study: steadily improving performance (increasing condition) and steadily declining performance (decreasing condition). The same critical incidents were used for both increasing and decreasing performance conditions; the only difference was the order of presentation. A description of the critical incidents and corresponding true score estimates for the secretary in the study, as well as the order in which they were combined in the experiment, is presented in Appendix A.

The videos were created based on video taped incidents that had been developed by Youtz-Padgett and Ilgen (1989). These authors took the following steps in developing their stimuli. First, five secretaries were interviewed to gather critical incidents relevant to the following performance dimensions: Job Knowledge and Skill, Organizational Ability (ability to organize), Dealing with Faculty and Students, and Working Cooperatively with Other Secretaries. Six graduate students in organizational behavior reviewed the incidents to ensure that they were relevant to the performance dimensions. Women with secretarial experience were hired to act out these critical incidents; incidents were video taped.
Each incident was assigned a true score estimate. Ten full-time secretaries acted as subject matter experts (SMEs) for this step. First, the SMEs were trained regarding typical psychometric rating errors that are of concern in rating tasks (e.g., halo, leniency, central tendency). Next, they practiced rating behavioral incidents and discussed the ratings, thus developing a common frame of reference within the group. For each critical incident that was filmed, the secretaries were given a written description of the event, were shown the incident twice, and were encouraged to take notes as they viewed the incident. After each incident was viewed, the SMEs were asked to categorize it in terms of performance dimension, and to rate the incident on a 7-point Behavior Anchored Rating Scale that was developed for clerical employees by the study authors. Initial criteria for retaining incidents and their true score estimates were that (a) at least 70% of the SMEs agreed on the classification of the critical incident with respect to performance dimension, and (b) the standard deviation of the rating associated with the critical incident was no more than 1.25. Coefficient alphas that were used to calculate the agreement for the classification step were .90 - .96; intra-class correlations for each performance dimension ranged from .81 - .93. This process resulted in a true score estimate for each critical incident, for each performance dimension that was relevant to the event. Complete details about the video tape and true score estimate construction are available in Youtz-Padgett and Ilgen (1989).

To construct the stimuli for this study, performance incidents of interest were copied onto a video tape in the order that would create the desired performance pattern.
The incidents that were used represent only a small portion of the video scenarios that were created for Youtz-Padgett and Ilgen's (1989) study.

**Measures**

Subjects completed the following instruments during the experiment: a behavior observation form (and corresponding practice forms where applicable), a rating form (and corresponding practice forms where applicable), a states list, a behavior check list, and a post-experiment questionnaire (including manipulation check). Each is described below, and is presented in Appendix B.

**Behavior based-processing manipulation.** A behavior observation form was used during training for subjects in the behavior-based condition, and for all subjects in the main part of the experiment. This form was two pages long and was comprised by the names of the performance dimensions, the dimension definitions, and lines for noting behavior observations for each performance dimension. At the top of the form, instructions appeared. The practice behavior observation form was a single-page version of the same form.

For subjects in the behavior-based condition, this form was intended to encourage behavior-based processing as they viewed the secretary. Subjects in the impression-based condition completed the form after they had completed their rating form (described below), primarily to have completed the same activities as the behavior-based group before taking the final two dependent measures.

**Rating accuracy.** This variable was constructed with responses on a form designed to assign a numerical value to the secretary's performance. The rating form was
a single-page instrument that asked subjects how they would rate the secretary on a scale of 1 to 7 for each of the four performance dimensions and for overall effectiveness (1 = below average; 7 = above average). At the bottom of this form, there was a space to make narrative comments. The practice rating form was identical to this form except for its title.

To construct the accuracy measure, performance ratings were compared to the true score estimates for the secretary to get a measure of distance accuracy, the average distance between the rating and the true score estimate across dimensions (McIntyre, Smith, & Hassett, 1984):

\[
\text{Distance Accuracy}_k = \frac{\sum_{i=1}^{d} \left( \sum_{j=1}^{n} \left| t_{ij} - r_{ijk} \right| \right)}{ni}
\]

Where \( k \) refers to the \( k \)th rater; \( n \) is the number of ratees; \( d \) is the number of dimensions; \( r \) refers to subject rating; and \( t \) refers to true scores. Rating accuracy was computed across dimensions, as well as at the individual dimension level. Rating accuracy scores were multiplied by -1; higher scores represent greater accuracy than lower scores.

Behavior recognition accuracy. This variable was constructed with responses to a behavior check list that was comprised by 20 critical incidents. Half described behaviors that the secretary did perform; the remaining half described behaviors that could reasonably have been completed, but ones that the secretary did not actually perform. Of the distracter incidents, half were positive; the other half were negative. Instructions at
the top of the form asked subjects to mark "Y" in front of the behaviors that they thought
the secretary performed and "N" in front of the behaviors that they thought the secretary
did not perform.

Then, hit rates (HR) and false alarm rates (FAR) were calculated. Hit rates
reflect the proportion of times a subject correctly identifies a behavior that was performed
by the secretary. False alarm rates reflect the proportion of times that the subject
incorrectly attributes a behavior as having occurred. Specifically, they are calculated by
finding the proportion of times each subject indicates that a behavior occurred when it did
not actually occur. Memory discrimination and response bias were calculated using the
following formulas (Macmillan & Creelman, 1990):

\[
\text{Memory discrimination} = \text{HR} - \text{FAR}
\]

\[
\text{Response bias} = \text{HR} + \text{FAR}
\]

Then, performance incidents were divided into early (first half of tape) and late
(second half of tape) behaviors. Behavior recognition accuracy measures described above
were computed for early and late behaviors separately.

**Filler task.** This task involved completing a states list, a form asking subjects to
write down as many of the 50 United States as they could. Fifty lines (arranged in two
columns) then appeared: the form was presented on a single page.

**Time pressure.** Subjects were asked on a post-experiment questionnaire the
extent to which they felt pressured by time to complete each experimental task (on a 5-
point, Likert-type scale; 1 = not at all; 5 = to a great extent). The questionnaire was
comprised by a number of short-answer questions (e.g., how well did you understand how
to complete the rating form?), as well as filler questions (to make the primary purpose of the questionnaire less obvious).

**Procedure**

Subjects were run in groups of one to four in either one hour or one and one-half hour sessions (depending on condition). Upon arrival to the experiment, subjects completed a consent form. They were told that the experiment was about secretarial performance observation, that they would be asked to view a secretary performing her job, and that they would be asked to fill out forms to answer questions about what they saw during the experiment.

First, subjects were trained in performance information processing. They were given two copies of a practice form that was intended to encourage the use of either behavior-based or impression-based information processing (i.e., a behavior observation form plus scratch paper for the former approach; a rating form for the latter). Subjects then listened to an instructional audio tape. For those in the behavior-based condition, the tape reviewed the performance dimensions and their definitions, then provided an example critical incident and explained how to make objective behavioral statements. For those in the impression-based condition, the tape reviewed the performance dimensions and their definitions, then provided examples of behavioral anchors (above average, average, and below average levels of performance) for each performance dimension. Then it explained that subjects would make overall evaluations of the secretary's performance.
Subjects were asked to practice completing their forms after watching an example critical incident on the video monitor. Then they were given feedback regarding what their practice forms should look like when complete. A second practice and feedback “session” was conducted; then the practice forms were collected. Training for the behavior-based condition lasted approximately 30 minutes; impression-based subjects were trained for approximately 20 minutes.

Subjects were then told that they would watch the “real” secretary in about a dozen work incidents, that the incidents would last about 17 minutes, and that there will not be an opportunity to view the secretary more than once. The performance incidents were then shown.

After all incidents had been played (approximately 17 minutes), subjects were asked to list as many of the fifty United States as they could for 3 minutes. This task was administered as a filler task (so that subjects could not rehearse behaviors just viewed and were forced to rely on memory to complete the experimental tasks). Subjects were told that the states task was part of a piloting effort for another experiment. They were told that the experimenter needed to know how much of this task could reasonably be completed in a short amount of time. Subjects were given a 3-minute break after the filler task was completed. At this point, the time pressure manipulation was introduced in the appropriate condition (see next section for details).

Next, subjects in the behavior-based condition were asked to complete a behavior observation form. When all subjects had completed these forms, they were collected.
Then they were asked to complete a rating form. When all subjects had completed the rating form, it was collected.

Subjects in the impression-based condition were asked to complete a rating form. When all subjects had completed this form, it was collected. Then, subjects were asked to complete a behavior observation form. When all subjects had completed these forms, they were collected.

At this point, subjects in all conditions had filled out the same dependent measures. The only difference was the order in which the forms were presented. Combined with the training initially given to subjects, this order of form administration is thought to encourage the appropriate information processing mode in each condition (e.g., Foti & Lord, 1987; Hirt, McDonald, & Erickson, 1995; Lichtenstein & Srull, 1987; Woehr & Feldman, 1993).

Subjects were then given a behavioral check list, which asked them to indicate whether the secretary performed a series of behaviors. After all subjects had finished this form, it was collected. Finally, the manipulation check forms were distributed. After they were collected, subjects were asked to leave their names and addresses if they wanted a to receive a form describing the purpose of the experiment after it was completely over.

**Time pressure manipulation.** After their break, subjects in the time pressured condition were told that the experiment was behind schedule and that there were still a number of tasks to complete. To aid in the timely completion of the experiment, the experimenter would have to move them along through the remaining tasks by timing
them. For the impression-based subjects, 5 minutes were allowed for the behavior observation form, 10 seconds were allowed for the rating form, and 1 and 1/2 minutes were allowed for the behavior check list. For the behavior-based subjects, 6 minutes were allowed for the behavior observation form, 25 seconds were allowed for the rating form, and 1 and 1/2 minutes were allowed for the behavior check list. Before distributing each form, the experimenter told subjects its name and the time allotted for it.

The time frames described above were determined through a pilot test of 50 subjects. These subjects rated the extent to which they understood each experimental task and the extent to which they felt pressured by time to complete each task on a 5-point, Likert-type scale (1 = not at all; 5 = to a great extent) on a post-experiment questionnaire. The criteria used to identify an appropriate length of time for each task were that (a) subjects indicated an average time pressure rating of 3.0 (i.e., “to some extent”) or higher, (b) subjects completed an average of 90% of the experimental task or more, and (c) these criteria would be satisfied for two experiment groups or more.

Just before receiving the manipulation check, the subjects were told that they were about to complete the final task in the experiment and that they were back on schedule (i.e., that they were “doing well” with time) so that limiting time for the task was no longer necessary. Subjects in the untimed condition were told that they could have as much time as they wanted to work on each of the forms. The experimenter tried to wait until all subjects had stopped writing before collecting forms to avoid pressuring subjects indirectly.
Results

Manipulation Check

To evaluate the effectiveness of the time pressure manipulation, subjects were asked to complete a questionnaire at the end of the experiment. The questionnaire asked subjects how well they thought they understood how to complete the various forms in the experiment, as well as the extent to which they felt pressured by time to complete each form. Subjects responded on a 5-point Likert-type scale (1 = not at all; 5 = to a great extent). Filler questions were added to the questionnaire to make its purpose less obvious.

T-tests indicate that the time pressure manipulation was successful. Significant differences were found in the perceived pressure involved in the behavior observation task ($M_{untimed} = 1.5; M_{timed} = 3.5; t(176) = -12.54, p < .01$), the rating task ($M_{untimed} = 1.4; M_{timed} = 3.6; t(176) = -15.8, p < .01$), and the behavior check list task ($M_{untimed} = 1.4; M_{timed} = 3.3; t(176) = -11.22, p < .01$) between untimed and timed groups.

General Variable Characteristics and Relationships

Descriptive statistics and correlations for the major variables of interest were calculated; results are presented in Tables 2 and 3. Table 3 reveals that there is some "halo" in the ratings; on average, dimension ratings were moderately correlated (range: $r = .26$ to $r = .44, p < .05$). The overall performance rating was strongly related to each of the individual dimension ratings, but most strongly to ratings of job knowledge ($r = .64, p < .01$).
Table 2

Descriptive Statistics for Major Study Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method&lt;sup&gt;a&lt;/sup&gt;</td>
<td>178</td>
<td>.49</td>
<td>.50</td>
</tr>
<tr>
<td>Performance Pattern&lt;sup&gt;b&lt;/sup&gt;</td>
<td>178</td>
<td>.51</td>
<td>.50</td>
</tr>
<tr>
<td>Time Pressure&lt;sup&gt;c&lt;/sup&gt;</td>
<td>178</td>
<td>.51</td>
<td>.50</td>
</tr>
<tr>
<td>Overall Performance</td>
<td>178</td>
<td>4.03</td>
<td>.94</td>
</tr>
<tr>
<td>Job Knowledge</td>
<td>177</td>
<td>4.56</td>
<td>1.20</td>
</tr>
<tr>
<td>Organizational Ability</td>
<td>176</td>
<td>2.63</td>
<td>.96</td>
</tr>
<tr>
<td>Dealing with Faculty and Students</td>
<td>169</td>
<td>4.49</td>
<td>1.12</td>
</tr>
<tr>
<td>Working Cooperatively with Other Secretaries</td>
<td>156</td>
<td>4.47</td>
<td>1.22</td>
</tr>
<tr>
<td>Distance Accuracy</td>
<td>178</td>
<td>-.88</td>
<td>.43</td>
</tr>
<tr>
<td>Rating Accuracy for Overall Performance</td>
<td>178</td>
<td>-.64</td>
<td>.69</td>
</tr>
<tr>
<td>Rating Accuracy for Job Knowledge</td>
<td>177</td>
<td>-.99</td>
<td>.67</td>
</tr>
<tr>
<td>Rating Accuracy for Organizational Ability</td>
<td>176</td>
<td>-.80</td>
<td>.54</td>
</tr>
<tr>
<td>Rating Accuracy for Dealing with Faculty and Students</td>
<td>169</td>
<td>-.94</td>
<td>.64</td>
</tr>
<tr>
<td>Rating Accuracy for Working Cooperatively with Other Secretaries</td>
<td>156</td>
<td>-1.0</td>
<td>.66</td>
</tr>
<tr>
<td>Hit Rate</td>
<td>176</td>
<td>.93</td>
<td>.10</td>
</tr>
</tbody>
</table>
Table 2, cot’d

Descriptive Statistics for Major Study Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. False Alarm Rate</td>
<td>176</td>
<td>.17</td>
<td>.14</td>
</tr>
<tr>
<td>17. Memory Discrimination for Early Behavior</td>
<td>175</td>
<td>.76</td>
<td>.19</td>
</tr>
<tr>
<td>18. Memory Discrimination for Late Behavior</td>
<td>176</td>
<td>.75</td>
<td>.21</td>
</tr>
<tr>
<td>19. Response Bias for Early Behavior</td>
<td>175</td>
<td>1.11</td>
<td>.19</td>
</tr>
<tr>
<td>20. Response Bias for Late Behavior</td>
<td>176</td>
<td>1.09</td>
<td>.18</td>
</tr>
</tbody>
</table>

Note: a This variable was dummy coded (0 = Behavior-based; 1 = Impression-based).

b This variable was dummy coded (0 = Decreasing performance; 1 = Increasing performance).

c This variable was dummy coded (0 = no time pressure; 1 = time pressure).
<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>.00</td>
<td>-.02</td>
<td>-.15*</td>
<td>-.25**</td>
<td>-.15*</td>
<td>-.30**</td>
<td>-.34**</td>
<td>.09</td>
<td>.10</td>
<td>-.05</td>
</tr>
<tr>
<td>Performance Pattern&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>-.03</td>
<td></td>
<td>-.34**</td>
<td>-.35**</td>
<td>-.19*</td>
<td>-.25**</td>
<td>-.15</td>
<td>.19**</td>
<td>.12</td>
<td>.09</td>
</tr>
<tr>
<td>Time Pressure&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>.02</td>
<td>-.08</td>
<td>.02</td>
<td>-.11</td>
<td>-.04</td>
<td>-.08</td>
<td>.01</td>
<td>-.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.64**</td>
<td>.55**</td>
<td>.48**</td>
<td>.57**</td>
<td>.13</td>
</tr>
<tr>
<td>Job Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.31**</td>
<td>.26*</td>
<td>.40**</td>
<td>.13</td>
<td>.03</td>
<td>.20**</td>
<td></td>
</tr>
<tr>
<td>Organizational Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.39**</td>
<td>.44**</td>
<td>.02</td>
<td>.01</td>
<td>.21**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dealing with Faculty and Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.43**</td>
<td>-.07</td>
<td>-.08</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working Cooperatively with Other Secretaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.19*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.75**</td>
<td>.63**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating Accuracy for Overall Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.36**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rating Accuracy for Job Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3, cot'd

**Intercorrelations for Major Study Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Processing Method&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.07</td>
<td>.19*</td>
<td>-.14</td>
<td>-.17*</td>
<td>.10</td>
<td>-.10</td>
<td>-.16*</td>
<td>.05</td>
<td>-.02</td>
</tr>
<tr>
<td>2. Performance Pattern&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.15*</td>
<td>.12</td>
<td>.10</td>
<td>-.03</td>
<td>-.06</td>
<td>.20**</td>
<td>-.07</td>
<td>.11</td>
<td>-.17*</td>
</tr>
<tr>
<td>3. Time Pressure&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-.11</td>
<td>.02</td>
<td>-.13</td>
<td>-.06</td>
<td>.07</td>
<td>-.04</td>
<td>.00</td>
<td>.06</td>
<td>.11</td>
</tr>
<tr>
<td>4. Overall Performance</td>
<td>-.06</td>
<td>-.13</td>
<td>.13</td>
<td>.19**</td>
<td>.07</td>
<td>.03</td>
<td>.04</td>
<td>.14</td>
<td>.16*</td>
</tr>
<tr>
<td>5. Job Knowledge</td>
<td>-.06</td>
<td>-.02</td>
<td>.09</td>
<td>.11</td>
<td>-.04</td>
<td>.03</td>
<td>.11</td>
<td>-.03</td>
<td>.06</td>
</tr>
<tr>
<td>6. Organizational Ability</td>
<td>-.29**</td>
<td>-.15*</td>
<td>.12</td>
<td>.06</td>
<td>.05</td>
<td>.00</td>
<td>-.03</td>
<td>.09</td>
<td>.05</td>
</tr>
<tr>
<td>7. Dealing with Faculty and Students</td>
<td>-.07</td>
<td>-.12</td>
<td>.03</td>
<td>.14</td>
<td>-.05</td>
<td>.09</td>
<td>.11</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td>8. Working Cooperatively with Other Secretaries</td>
<td>-.10</td>
<td>-.13</td>
<td>.02</td>
<td>.08</td>
<td>.03</td>
<td>.01</td>
<td>.02</td>
<td>.05</td>
<td>.06</td>
</tr>
<tr>
<td>9. Distance Accuracy</td>
<td>.62**</td>
<td>.56**</td>
<td>.62**</td>
<td>.08</td>
<td>-.16*</td>
<td>.17*</td>
<td>.15*</td>
<td>-.08</td>
<td>-.08</td>
</tr>
<tr>
<td>10. Rating Accuracy for Overall Performance</td>
<td>.36**</td>
<td>.27**</td>
<td>.33**</td>
<td>.11</td>
<td>-.16*</td>
<td>.12</td>
<td>.21**</td>
<td>.12</td>
<td>.00</td>
</tr>
<tr>
<td>11. Rating Accuracy for Job Knowledge</td>
<td>.23**</td>
<td>.10</td>
<td>.15</td>
<td>-.01</td>
<td>-.08</td>
<td>.02</td>
<td>.03</td>
<td>-.09</td>
<td>-.08</td>
</tr>
</tbody>
</table>
Table 3, cot'd.

**Intercorrelations for Major Study Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Rating Accuracy for Organizational Ability</td>
<td>--</td>
<td>.19*</td>
<td>.25**</td>
<td>.06</td>
<td>-.12</td>
<td>.09</td>
<td>.14</td>
<td>-.09</td>
<td>-.02</td>
</tr>
<tr>
<td>13. Rating Accuracy for Dealing with Faculty and Students</td>
<td>--</td>
<td>.18**</td>
<td>-.05</td>
<td>-.05</td>
<td>.05</td>
<td>.00</td>
<td>-.03</td>
<td>-.08</td>
<td></td>
</tr>
<tr>
<td>14. Rating Accuracy for Working Cooperatively with Other Secretaries</td>
<td>--</td>
<td>.03</td>
<td>-.12</td>
<td>.17*</td>
<td>.07</td>
<td>-.02</td>
<td>-.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Hit Rate</td>
<td>--</td>
<td>-.15*</td>
<td>.51**</td>
<td>.58**</td>
<td>.28**</td>
<td>.44**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. False Alarm Rate</td>
<td>--</td>
<td>-.76**</td>
<td>-.77**</td>
<td>.75**</td>
<td>.66**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Memory Discrimination for Early Behavior</td>
<td>--</td>
<td>.62**</td>
<td>-.14</td>
<td>-.48**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Memory Discrimination for Late Behavior</td>
<td>--</td>
<td>-.55**</td>
<td>-.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Response Bias for Early Behavior</td>
<td>--</td>
<td>.53**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Response Bias for Late Behavior</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3, cot'd.

**Intercorrelations for Major Study Variables**

*Note.* *This variable was dummy coded (0 = Behavior-based; 1 = Impression-based).*  *b*This variable was dummy coded (0 = Decreasing performance; 1 = Increasing performance).*  *c*This variable was dummy coded (0 = no time pressure; 1 = time pressure).*  

* *p* < .05. ** *p* < .01.
Distance accuracy was strongly related to accuracy for rating each of the
performance dimensions ($r = .56$ to $r = .75$, $p < .01$). Furthermore, rating accuracy for
each individual performance dimension was generally related to rating accuracy for the
other dimensions ($r = .10$, n.s. to $r = .36$, $p < .01$). Raters' ability to accurately evaluate
the target's performance remained fairly stable across performance dimensions.

Significant relationships among measures of behavior recognition accuracy were
discovered. Memory discrimination for behaviors viewed early in the observation period
was closely related to memory discrimination for behaviors viewed later in the
observation period ($r = .62$, $p < .01$). Subjects' ability to distinguish between behavior
that did and did not occur remained consistent across the performance observation period.
A similar relationship emerged for measures of response bias for early and late behaviors
($r = .53$, $p < .01$). Raters' tendency to misattribute behaviors that did not occur to the
performer did not vary as a function of time during the observation period. Relationships
between memory discrimination and response bias examined separately at the early and
late stages of performance were not statistically significant ($r = -.14$, n.s.; $r = -.03$, n.s.,
respectively). This may be due to the high and stable hit rates ($x = 93\%$; $SD = 10\%$) and
low false alarm rates ($x = 17\%$; $SD = 14\%$) that subjects displayed on the behavior
recognition measure.

Measures of distance accuracy and behavior recognition accuracy were minimally
related. The only significant relationships emerging were those between measures of
distance accuracy and memory discrimination for early and late behavior ($r = .17$, $p < .05$;
$r = .15$, $p < .05$, respectively). In other words, ratings became more accurate as the rater's
ability to discriminate among behaviors that did and did not occur improved. Rating
accuracy was not related to one's tendency to misattribute behaviors that were not performed to the ratee (i.e., response bias).

Distance accuracy was significantly related to performance pattern \( (r = .19; p < .01) \). Those rating improving performance were more accurate than those rating deteriorating performance. Furthermore, raters who observed and evaluated the target when her performance improved over time demonstrated better memory discrimination for early behavior than raters observing the performer when her performance deteriorated over time \( (r = .20, p < .05) \). Furthermore, those viewing declining behavior had higher response bias measures for late behavior than did those viewing improving behavior \( (r = -.17, p < .05) \). Raters with higher response bias for late behavior tended to give the target a higher overall rating as well \( (r = .16, p < .05) \). Overall, raters evaluated and recognized target behavior more accurately when her performance improved over time than when it deteriorated over time.

One small relationship between processing method and behavior recognition accuracy emerged. Those in the behavior-based condition demonstrated better memory discrimination for late behaviors than did impression-based raters \( (r = -.16; p < .05) \). That is, behavior-based raters more accurately described what the target did late in the performance period than did impression-based raters. This finding is consistent with study predictions.

**Tests of Hypotheses**

The first two hypotheses predicted an interaction between processing method and time pressure in their effects on behavior recognition accuracy for early and late behaviors. Hypotheses 1 and 2 were written in two parts for ease of discussion.
However, conceptually, they should be tested in a single analysis (testing them separately increases the probability of Type I error). Therefore, two 2 (Time Pressure) x 2 (Processing Method) MANOVAs were conducted. The dependent variables in the first analysis were memory discrimination for early behavior and for late behavior. The second analysis was conducted with response bias for early and late behavior as the dependent variables. These analyses revealed no significant interaction terms or main effects, providing no support for the first two hypotheses (See Tables 4 and 5).

Hypothesis 3 predicted an interaction between time pressure and processing method in their effects on rating accuracy. To test this proposition, a series of 2 (Time Pressure) x 2 (Processing Method) ANOVAs with distance accuracy, as well as rating accuracy for each of the performance dimensions (and overall performance), as the dependent variables was conducted. In all cases, analyses revealed a nonsignificant interaction term and nonsignificant main effects (See Tables 6 to 11). However, at the individual dimension level, analyses revealed a main effect for processing method on rating accuracy for some dimensions. For the Dealing with Faculty and Students dimension, impression-based raters made more accurate ratings than did behavior-based raters, $F(1, 167) = 6.0, p < .01$. For the Working Cooperatively with Other Secretaries dimension, behavior-based raters made more accurate ratings than did impression-based raters, $F(1, 154) = 3.8, p = .05$.

Overall, processing method had limited (and unexpected) effects on rating accuracy. Neither distance accuracy nor individual dimension accuracy was affected by
Table 4

Multiple Analysis of Variance for Memory Discrimination

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early and Late Behavior</td>
<td>.01</td>
<td>1</td>
<td>.84</td>
</tr>
<tr>
<td>Processing Method x Early and Late Behavior</td>
<td>.01</td>
<td>1</td>
<td>.75</td>
</tr>
<tr>
<td>Time Pressure x Early and Late Behavior</td>
<td>.01</td>
<td>1</td>
<td>.42</td>
</tr>
<tr>
<td>Processing Method x Time Pressure x Early and Late Behavior</td>
<td>.03</td>
<td>1</td>
<td>2.02</td>
</tr>
<tr>
<td>Within cells</td>
<td>.02</td>
<td>171</td>
<td></td>
</tr>
</tbody>
</table>
Table 5

**Multiple Analysis of Variance for Response Bias**

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early and Late Behavior</td>
<td>.01</td>
<td>1</td>
<td>.84</td>
</tr>
<tr>
<td>Processing Method x Early and Late Behavior</td>
<td>.01</td>
<td>1</td>
<td>.75</td>
</tr>
<tr>
<td>Time Pressure x Early and Late Behavior</td>
<td>.01</td>
<td>1</td>
<td>.42</td>
</tr>
<tr>
<td>Processing Method x Time Pressure x Early and Late Behavior</td>
<td>.03</td>
<td>1</td>
<td>2.02</td>
</tr>
<tr>
<td>Within cells</td>
<td>.02</td>
<td>171</td>
<td></td>
</tr>
</tbody>
</table>
Table 6

Analysis of Variance for Distance Accuracy

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>.17</td>
<td>1</td>
<td>.92</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.21</td>
<td>1</td>
<td>1.16</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.14</td>
<td>1</td>
<td>.76</td>
</tr>
<tr>
<td>Total</td>
<td>.18</td>
<td>177</td>
<td></td>
</tr>
</tbody>
</table>
Table 7

Analysis of Variance for Rating Accuracy for Overall Performance

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>.91</td>
<td>1</td>
<td>1.9</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.01</td>
<td>1</td>
<td>.02</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>1.7</td>
<td>1</td>
<td>3.5</td>
</tr>
<tr>
<td>Total</td>
<td>.48</td>
<td>177</td>
<td></td>
</tr>
</tbody>
</table>
Table 8

**Analysis of Variance for Rating Accuracy for Job Knowledge**

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>.18</td>
<td>1</td>
<td>.40</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.32</td>
<td>1</td>
<td>.69</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.58</td>
<td>1</td>
<td>1.27</td>
</tr>
<tr>
<td>Total</td>
<td>.46</td>
<td>176</td>
<td></td>
</tr>
</tbody>
</table>
Table 9

Analysis of Variance for Rating Accuracy for Organizational Ability

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>.22</td>
<td>1</td>
<td>.75</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.60</td>
<td>1</td>
<td>2.1</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.33</td>
<td>1</td>
<td>.29</td>
</tr>
<tr>
<td>Total</td>
<td>.29</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>
Table 10

Analysis of Variance for Rating Accuracy for Dealing with Faculty and Students

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>2.4</td>
<td>1</td>
<td>6.0**</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.04</td>
<td>1</td>
<td>.11</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.00</td>
<td>1</td>
<td>.00</td>
</tr>
<tr>
<td>Total</td>
<td>.40</td>
<td>168</td>
<td></td>
</tr>
</tbody>
</table>

Note. ** p < .01.
Table 11

**Analysis of Variance for Rating Accuracy for Working Cooperatively with Other Secretaries**

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>1.6</td>
<td>1</td>
<td>3.8*</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>1.5</td>
<td>1</td>
<td>3.5</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.23</td>
<td>1</td>
<td>.46</td>
</tr>
<tr>
<td>Total</td>
<td>.43</td>
<td>155</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** *p = .05.*
time pressure to complete the rating task.

Hypothesis 4 predicted a 3-way interaction between time pressure, processing method, and performance pattern in their effects on rating level. A series of 2 (Time Pressure) x 2 (Processing Method) x 2 (Performance Pattern) ANOVAS with rating level (for each performance dimension and overall performance) as the dependent variable was conducted (n = 155 to 177). Results are presented in Tables 12 to 16.

The 3-way interaction term was significant only for the dimension of Dealing with Faculty and Students ($F(1, 167) = 4.5, p < .05$). However, the nature of the interaction is different from what was predicted. In conditions without time pressure, all subjects rated decreasing performance higher than increasing performance. In conditions with time pressure, performance pattern had no effect on rating level in the impression-based processing condition. For subjects in the behavior-based condition, when the performer improved over time, she was given a lower rating than when her performance declined over time. Tukey tests reveal that only behavior-based raters who viewed performance that decreased over time created ratings that differed significantly from those created by raters in the other ( pressured) conditions (See Figure 5).

For all performance dimensions, including the overall performance level, main effects for both processing method and performance pattern were significant. F-values are as follows: for Overall Performance, $F(1, 176) = 4.71, p < .05; F(1, 176) = 22.40, p < .01$, respectively; for Job Knowledge, $F(1, 175) = 14.11, p < .01; F(1, 175) = 26.69, p < .01$, respectively; for Organizational Ability, $F(1, 174) = 4.33, p < .05; F(1, 174) = 6.54, p < .01$, respectively; for Working Cooperatively with Other Secretaries, $F(1, 154)$
Table 12

Analysis of Variance for Rating Level for Overall Performance

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>3.74</td>
<td>1</td>
<td>4.71*</td>
</tr>
<tr>
<td>Performance Pattern</td>
<td>17.80</td>
<td>1</td>
<td>22.40**</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.15</td>
<td>1</td>
<td>.19</td>
</tr>
<tr>
<td>Processing Method x performance Pattern</td>
<td>.36</td>
<td>1</td>
<td>.46</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.09</td>
<td>1</td>
<td>.12</td>
</tr>
<tr>
<td>Performance Pattern x Time Pressure</td>
<td>.15</td>
<td>1</td>
<td>.19</td>
</tr>
<tr>
<td>Processing Method x Performance Pattern x Time Pressure</td>
<td>.51</td>
<td>1</td>
<td>.65</td>
</tr>
<tr>
<td>Error</td>
<td>.79</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.89</td>
<td>177</td>
<td></td>
</tr>
</tbody>
</table>

Note.  *p < .05.  **p < .01.
Table 13

Analysis of Variance for Rating Level for Job Knowledge

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>16.6</td>
<td>1</td>
<td>14.11**</td>
</tr>
<tr>
<td>Performance Pattern</td>
<td>31.37</td>
<td>1</td>
<td>26.69***</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>1.46</td>
<td>1</td>
<td>1.25</td>
</tr>
<tr>
<td>Processing Method x performance Pattern</td>
<td>.55</td>
<td>1</td>
<td>.47</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>2.04</td>
<td>1</td>
<td>1.74</td>
</tr>
<tr>
<td>Performance Pattern x Time Pressure</td>
<td>1.27</td>
<td>1</td>
<td>1.08</td>
</tr>
<tr>
<td>Processing Method x Performance Pattern x Time Pressure</td>
<td>.00</td>
<td>1</td>
<td>.00</td>
</tr>
<tr>
<td>Error</td>
<td>1.18</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1.43</td>
<td>176</td>
<td></td>
</tr>
</tbody>
</table>

Note. **p < .01.
Table 14

**Analysis of Variance for Rating Level for Organizational Ability**

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>3.84</td>
<td>1</td>
<td>4.33*</td>
</tr>
<tr>
<td>Performance Pattern</td>
<td>5.80</td>
<td>1</td>
<td>6.54**</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.10</td>
<td>1</td>
<td>.12</td>
</tr>
<tr>
<td>Processing Method x performance Pattern</td>
<td>.24</td>
<td>1</td>
<td>.27</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.06</td>
<td>1</td>
<td>.06</td>
</tr>
<tr>
<td>Performance Pattern x Time Pressure</td>
<td>.28</td>
<td>1</td>
<td>.32</td>
</tr>
<tr>
<td>Processing Method x Performance pattern x Time Pressure</td>
<td>1.78</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Error</td>
<td>.89</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.92</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* *p < .05. **p < .01.
Table 15

Analysis of Variance for Rating Level for Dealing with Faculty and Students

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>20.75</td>
<td>1</td>
<td>19.84**</td>
</tr>
<tr>
<td>Performance Pattern</td>
<td>13.88</td>
<td>1</td>
<td>13.27**</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>3.22</td>
<td>1</td>
<td>3.08</td>
</tr>
<tr>
<td>Processing Method x Performance Pattern</td>
<td>.42</td>
<td>1</td>
<td>.40</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>.17</td>
<td>1</td>
<td>.17</td>
</tr>
<tr>
<td>Performance Pattern x Time Pressure</td>
<td>.05</td>
<td>1</td>
<td>.04</td>
</tr>
<tr>
<td>Processing Method x Performance Pattern x Time Pressure</td>
<td>4.72</td>
<td>1</td>
<td>4.5*</td>
</tr>
<tr>
<td>Error</td>
<td>1.05</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1.25</td>
<td>168</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** *p < .05. **p < .01.
Table 16

Analysis of Variance for Rating Level for Working Cooperatively with Other Secretaries

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>28.55</td>
<td>1</td>
<td>21.8**</td>
</tr>
<tr>
<td>Performance Pattern</td>
<td>5.91</td>
<td>1</td>
<td>4.52*</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>1.77</td>
<td>1</td>
<td>1.35</td>
</tr>
<tr>
<td>Processing Method x Performance Pattern</td>
<td>.02</td>
<td>1</td>
<td>.02</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>1.30</td>
<td>1</td>
<td>.99</td>
</tr>
<tr>
<td>Performance Pattern x Time Pressure</td>
<td>1.30</td>
<td>1</td>
<td>.99</td>
</tr>
<tr>
<td>Processing Method x Performance Pattern x Time Pressure</td>
<td>.65</td>
<td>1</td>
<td>.49</td>
</tr>
<tr>
<td>Error</td>
<td>1.30</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1.50</td>
<td>155</td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05. **p < .01.
Figure 5

Effect of Time Pressure, Processing Method, and Performance Pattern on Rating Level for Dealing With Faculty and Students

With Time Pressure

Without Time Pressure

- Decreasing
- Increasing
= 21.8, \( p < .01 \); \( F (1, 154) = 4.52, \ p < .05 \), respectively. In all cases, when performance declined over time, it was rated higher than when it improved over time. Also, subjects in the behavior-based condition rated the performer higher than subjects in the impression-based condition.

The absence of effect of time pressure on the dependent variables was puzzling. Therefore, the distribution of subjects' perceptions of time pressure to complete the experiment tasks was examined. A positive skew was detected, suggesting that a considerable portion of subjects did not feel as much pressure as the experimenter had intended. Therefore, individuals who indicated a high level of time pressure (i.e., more than "to some extent") were separated from the rest of the subjects. Analyses testing the hypotheses were repeated with this group. Results were identical to those already described except for effects of processing method on rating accuracy for the Job Knowledge dimension. For this dimension, there was a significant interaction between processing method and time pressure, \( F (1, 135) = 7.2, \ p < .01 \). Under time pressure, behavior-based raters were more accurate than impression-based raters for this dimension. Without time pressure, impression-based raters were more accurate than were behavior-based raters (See Table 17 and Figure 6). This interaction does not reflect the effects of time pressure and processing method that were predicted. Tukey tests reveal that behavior-based raters under pressure were more accurate than other raters; rating accuracy did not differ for those in the other conditions.
Table 17

**Analysis of Variance for Rating Accuracy for Job Knowledge (Time Pressured Subjects only)**

<table>
<thead>
<tr>
<th>Source</th>
<th>MS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Method</td>
<td>.88</td>
<td>1</td>
<td>2.3</td>
</tr>
<tr>
<td>Time Pressure</td>
<td>.13</td>
<td>1</td>
<td>.35</td>
</tr>
<tr>
<td>Processing Method x Time Pressure</td>
<td>2.8</td>
<td>1</td>
<td>7.2**</td>
</tr>
<tr>
<td>Total</td>
<td>.40</td>
<td>136</td>
<td></td>
</tr>
</tbody>
</table>

*Note. ** p < .01.*
Figure 6

Effects of Processing Method and Time Pressure on Rating Accuracy for Job Knowledge (Time Pressured Subjects only)
Discussion

This study examined a number of dynamics involved in the performance rating task. It supports the notion that accuracy in rating performance and accuracy in identifying behaviors that ratees perform are two separate constructs. The magnitude of the relationships that were discovered between these two skills is similar to that found in recent research (e.g., Sanchez & De La Torre, 1996). However, the relationships are not as strong in this study as they have been in others (e.g., Murphy et al., 1982, 1984).

The only study manipulation that affected overall (distance) accuracy was performance pattern. Raters watching a performer who improved over time were more accurate than those rating a performer who deteriorated over time. Characteristics of the subject population may help explain this effect. Subjects were drawn from a well reputed and highly selective university. They performed at high levels, and were accustomed to good performance from others around them. The initial poor performance of the secretary may have caught and held their attention better than initial good performance because poor performance is somewhat novel to the raters.

At the individual dimension level, effects for processing method on rating accuracy were discovered. For the Dealing with Faculty and Students dimension, impression-based raters were more accurate than behavior-based raters. It is possible that subjects (university students) had well-developed, internal performance models for this dimension because it is one with a high level of relevance for them. These models would allow subjects to accurately analyze the target’s behavior, even in the absence of external structure for the evaluation task. Subjects who attempted to evaluate behavior in a more systematic way (i.e., behavior-based raters) may have over-processed the information,
leading to ratings of compromised accuracy. Future research could investigate whether experience-based internal models of performance can provide guidelines that allow raters to evaluate performance accurately.

For the Working Cooperatively with Other Secretaries dimension, behavior-based raters were more accurate than were impression-based raters. One possible explanation for this effect is the converse of that offered in the previous paragraph. In their experience interacting with secretaries, subjects are not likely to have paid much attention to the secretaries’ ability to work well with one another. Therefore, more careful information processing (i.e., behavior-based processing) was necessary to arrive at a more accurate evaluation for this dimension.

Although the study did not reveal stable effects of information processing on rating accuracy, the nature of the raters’ inaccuracy differed according to the processing method they used while watching the performer. Behavior-based raters gave performers higher ratings than did impression-based raters. The reason for this effect is not apparent. It may be that the process of recalling behaviors caused raters to evaluate performance more favorably (only subjects in the behavior-based processing condition took this step before rating the performance). In their book, Murphy and Cleveland (1991) point out that encouraging raters to write down performance information shifts information categorization from an automatic to a controlled processing mode. Increased focus on performer behaviors probably also causes the rater to ignore contextual information as the ratee performs tasks. This contextual information may or may not be relevant to the rating task. However, differential use of this information may help explain the differences in level of rating assigned by behavior-based and impression-based raters.
This study shows that consistency in the processing methods used among raters who evaluate candidates is critical. A group of individuals that is considered for a performance-based decision (e.g., hiring or promotion) should be evaluated by raters who are using the same type of performance observation processes. Considerable noise can be introduced into the decision by comparing ratings created by individuals who evaluated performance with different processing methods. One difficulty in establishing common rating methods is that raters are not always consciously aware of their rating goals (Murphy & Cleveland, 1991). These goals can heavily influence information processing methods that are used during the rating process. This study reinforces the importance of not arbitrarily using performance ratings created for one purpose (e.g., promotion decision) for another purpose (e.g., behavioral feedback) (Cleveland, Murphy, & Williams, 1989). The implications of using ratings created with each rating method must be considered in designing the performance rating system that forms the basis for human resources decisions in organizations.

The target was rated higher when her performance deteriorated over time than when it improved over time: Raters were more heavily influenced by the performance that was displayed at the beginning of the observation period than by performance at the end. Results from this study demonstrate the robustness of the primacy effect in the performance rating task. Although raters viewed performance for a relatively short period of time, without interruptions or competing demands, early behavior clearly affected performance ratings more than did action taken later on. The performer was rated higher when her best performance came early than when she took some time to "warm up". This result implies that it is difficult to underestimate the impact of a first impression. In fact,
it may not be possible to “recover” from a bad start. On the flip side, it seems likely that performers who begin to “slide” on the job can do so without jeopardizing others’ impressions of their performance for some time. This speculation assumes that what is written on a performance appraisal form reflects the rater’s impressions of the performer, which may not be the case (e.g., Longenecker et al., 1987). Because the subjects in this study had nothing to lose by rating candidly (i.e., giving the performer the rating they thought she deserved), there is no apparent reason to doubt that the ratings made by subjects in this study reflect their actual impression of the secretary’s performance.

Another explanation for the heavy influence that early behaviors exerted over the performance ratings is that the subjects became bored with the experiment and their attention drifted after the first few incidents. Subjects in the impression-based processing condition did not remember later performance incidents as well as did those in the behavior-based processing condition. However, for all subjects, comparison of mean levels of memory discrimination and response bias for early and late behaviors, as well as the relationship between measures of behavior recognition accuracy for early and late behaviors, suggests that they were roughly as good at identifying behaviors occurring early as they were at identifying those that came later. Therefore, early performance played a much heavier role in determining performance ratings than did later information for reasons other than differences in attention level.

The variables consistently affecting performance rating level in this study (performance pattern and processing method) are controlled by different sources. That is, the ratee has most control over whether initial performance is relatively high versus relatively low. This suggests a way for individuals to manage impressions that others
have of them at work. However, it is the rater who has most control over the type of processing goals (and methods) that are used when creating performance ratings. He or she can influence the performance ratings in this way. Both rater and ratee sources of "noise" in the rating process should be addressed to maximize the "purity" of the ratings.

Performance pattern was related to some measures of behavior recognition accuracy. Raters viewing increasing performance had better memory discrimination scores for early behavior than did raters observing decreasing performance. Furthermore, raters viewing decreasing performance had higher response bias for late behavior. In other words, behavior recognition accuracy was better when raters viewed increasing performance. This finding is consistent with research on the "negativity bias" (e.g., Skowronski & Carlson, 1989). This bias refers to the heavier impact that negative information carries (as opposed to positive information). In this study, the secretary's early poor performance caught and held raters' attention more effectively than did the early positive performance information. The implication for this result is that new employees who perform below expectations are likely to be scrutinized more closely than their high-performing counterparts.

Processing method was minimally related to behavior recognition accuracy. Behavior-based rating processes led to ratings that were somewhat better in memory discrimination for late behaviors than did impression-based processes. However, behavior-based processes require more time and effort on the part of the rater (Fiske & Taylor, 1991; Lichtenstein & Srull, 1987). It is not clear whether the slight advantage in behavior recognition accuracy is worth the additional resources.
The answer to this question may lie in the purpose of the ratings. If they are to be used for hiring or promotion decisions based on an overall performance assessment, perhaps the additional time and energy required to make behavior-based ratings is not well spent. Impression-based ratings should serve this purpose equally well. However, if the rater is gathering feedback to discuss skill improvement in particular areas with the ratee, the effort spent to make ratings (and carefully attend to behavior) with behavior-based processes may lead to a more productive attempt at development.

On average, raters indicated that they felt pressured by time to complete the experiment tasks. However, ratings were basically unaffected by this manipulation. As previously mentioned, the distribution of subjects' perception of time pressure during the experiment was positively skewed. Therefore, the bulk of the subjects did not feel as much pressure as the experimenter intended. However, the experiment hypotheses were not supported even when subjects who indicated genuine pressure to complete their tasks were examined on their own.

It is possible that, even when considering only the most pressured subjects, the level of time pressure did not reach levels high enough to affect the ratings. Another possibility is that time pressure does not affect the rating task unless it is introduced earlier in the evaluation process (e.g., encoding stage). In this study, all subjects viewed the performer in the same incidents without the application of time pressure. In other studies examining the effects of time pressure on decisions, the time pressure was generally introduced before subjects had begun gathering information to help make their decision (Hulland & Kleinmuntz, 1994; Kruglanski & Freund, 1983; Svenson, Edland, & Slovic, 1990; Wright, 1974). Had time pressure been introduced at an earlier stage of the
experiment, ratings may have varied as a function of its intensity.

When considering all subjects, the single significant effect of time pressure was discovered in the rating level of the Dealing with Faculty and Students dimension. For this dimension, the familiar main effects for performance pattern and processing method were evident when raters were not pressured to perform the rating task. However, these effects disappeared completely for impression-based raters when they were under time pressure to rate the performer (the effects remained under time pressure for behavior-based raters). As mentioned previously, during their interactions with secretaries in the past, subjects may have developed internal standards for performance on this dimension. Under time pressure, raters may have been forced to use them to evaluate behavior (there was not time to use any other rating strategy). These standards may have helped structure the performance information so that uniform ratings were created, regardless of changes in performance level. The rating standards may have "protected" raters who were encouraged to use them (i.e., impression-based raters) from biases usually introduced by differences in ratee performance level.

When responses from only the most pressured subjects were examined, accuracy for the Job Knowledge dimension increased for behavior-based raters and decreased for impression-based raters when time was restricted for the rating task. This effect may be a function of the subjects in the study. As bright, high-achieving students in a selective private university, Job Knowledge may be a performance dimension of particular interest to subjects. It is possible that raters tended to over-analyze performance for this dimension, thus detracting from the accuracy of the ratings. This is particularly likely when subjects were encouraged to carefully consider performer behavior (i.e., behavior-
based condition). When time for the rating task was restricted, accuracy for behavior-based raters increased because raters were prevented from over-analyzing the target's behavior. In the non-pressured condition, impression-based subjects were free to engage in some behavior analysis, even though the experiment did not explicitly encourage this. This step may have enhanced rating accuracy. However, in the time pressured condition, impression-based subjects had no time to perform behavior analysis. Thus, accuracy of ratings decreased under these conditions.

In other words, there may be an "ideal" amount of behavior analysis that should take place when performance is evaluated. Significantly more or less processing time may detract from the accuracy of the performance ratings. This idea might be examined in future research.

Before closing this discussion, the lack of comparison with Youtz-Padgett and Ilgen's (1989) study should be addressed. While the stimulus materials in this study represent a subset of those used in these authors' work, considerable differences in independent variables, dependent variables, and study methodology prevents meaningful comparison of study results. For example, changes in performance level was an independent variable of interest in both studies. However, in this study, performance decreased steadily or increased steadily. In the Youtz-Padgett and Ilgen (1989) study, effects based on differences in performance variability were examined. Therefore, results of these studies are not compared.

Alternative Explanations and Study Limitations

This study suffers from a number of limitations. First, the stimuli encompass performance incidents from a single performer; study results may not generalize for this
reason. Second, the nature of the subjects for this research also limits the generalizability. Students had no investment in the ratings that they made for the performer. Neither positive nor negative consequences for doing a good versus poor job at the rating task existed. In general, past research suggests that these conditions would lead to ratings that are lower than those made by those who must interact with the ratee (DeNisi et al., 1984; Ilgen & Feldman, 1983). Undoubtedly, there are other important ways in which the rating task in this study differs from that in the typical organization. These differences may change the dynamics of the rating task considerably. Therefore, the issues raised in this study need to be examined under conditions of stronger external validity.

Another important limitation is that it is not possible to determine whether subjects relied upon impressions of candidates or used actual behavior as a basis for their ratings (Lord & Maher, 1991). Some researchers have compared subjects' behavior recognition errors to those that would be made when using a particular schema to categorize behavior (e.g., effective leader, ineffective leader) (e.g., Phillips, 1984). Because use of a particular schema was not measured in this study, it is not possible to gauge actual schema usage in this way.

The way in which performance information was displayed to subjects represents another study flaw. It was not possible to present performer behavior that was relevant to each of the performance dimensions evenly (while preserving the performance patterns of interest). For example, the bulk of the performance incidents relevant to the Working Cooperatively with Other Secretaries dimension was shown at the beginning of the performance period for subjects in the decreasing performance condition; these incidents were shown at the end of the observation period for those in the increasing condition.
Due to the nature of the behavior recognition task, it was not possible to determine whether behavior recognition accuracy was affected by the way that performance incidents were displayed. However, differences in rating accuracy for one of the individual performance dimensions were found as a function of performance pattern. Subjects in the increasing performance condition were more accurate than those in the decreasing performance condition for the Organizational Ability dimension. When performance improved over time, all of the performance information pertinent to this dimension appeared early. When performance declined over time, all of the information pertinent to this dimension came late in the observation session. This reinforces the conclusion that raters were heavily influenced by performance information appearing early.

Supervisor ratings of employee performance play an important role in human resources research; they are the most frequently used criteria in validation studies (Gatewood & Field, 1990; Krzystof, et al., 1988). Furthermore, data from performance appraisal systems are used to make impactful career decisions (Krzystof, et al., 1988; Longenecker et al., 1987; Schmitt & Klimoski, 1991). Therefore, it behooves us to continue exploring the conditions under which they are best created and the effects of using various rating techniques.

The process of rating job performance is complicated; many issues in this area remain unresolved. A large part of the complexity involves the organizational dynamics that affect the appraisal process (e.g., rater’s goals, organizational rewards, etc.). Because these factors play such an impactful role in determining the rating outcome, we are only at the beginning stages of efforts aimed at improving the appraisal process when studying
it in the laboratory. Field research is required to capture and examine other highly influential factors, many of which are difficult to measure.

Modifying the rating process to ensure accurate ratings requires understanding and control of the rating environment more sophisticated than what we currently have. Rater cognitive, affective, ability, and motivation factors interact with one another in complicated ways during the rating process (Murphy & Cleveland, 1995). It is not clear that it will be possible to control enough of these factors to make the rating process sound. For example, rater goals, organization goals, and organization rewards with respect to the performance appraisal process often do not align (Murphy & Cleveland, 1991). Given this type of rating environment, it should be no surprise that most organizations’ performance rating systems do not produce ratings that form a reliable basis for human resources decisions.

Some research needs in the area include: application for our knowledge of the cognitive components of performance appraisal, a better understanding of the way that rating goals affect rater behavior, and greater attention to the affective components of performance appraisal (Murphy & Cleveland, 1995). This will help us advance our knowledge beyond cognitive components of the rating task. At that point, we can then begin to address the performance evaluation process in addition to the performance judgment process.

Another way to address the problems and complexities in performance ratings is to find other ways to make decisions that have traditionally relied upon their input. Many organizations place little weight on performance ratings due to the problems inherent in their appraisal systems; perhaps researchers should follow suit and spend their energy
investigating other ways to make human resources decisions and change the role and form of supervisor input. For example, in cases where multiple sources of incumbent data are available to make an employment decision (e.g., promotion), there may be no need to use supervisor-generated performance ratings. Instead, the supervisor might be asked to provide behavioral observations to get a sense of the types of things that the employee does on the job. Inferences about the individual’s relative skill strengths and weaknesses might be drawn from the descriptions. To get a more reliable sense of the individual’s absolute level of skill, the organization could require the employee to complete a standardized assessment process. Here performance of all promotion candidates could be evaluated by the same raters, using the same performance standards. This way, skill ratings would be less contaminated by supervisor biases, including those introduced by using different processing methods in the rating task.

This suggestion does not imply that supervisor ratings of performance should be abandoned altogether. Rather, it underscores the need to standardize the rating environment when important decisions are based on the ratings. One way to achieve such standardization is to leave the performance rating task to someone who will rate all performers in a uniform way, without letting cognitive and affective biases differentially contaminate the process.
References


Hastie, R. & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. Psychological Review, 93, 258-268.


S. Wher, D. Hamilton, & D. E. Carlson (Eds.), Person Memory: The cognitive basis of social perception (pp. 227-300). Hillsdale, NJ: Erlbaum.


Appendix A

Critical Incidents
## CRITICAL INCIDENTS

<table>
<thead>
<tr>
<th>Incident</th>
<th>Description</th>
<th>Job Knowledge</th>
<th>Dealing with Faculty and Students</th>
<th>Working with other secretaries</th>
<th>Organization of work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Picks up another secretary's items from the basement while she is there getting references.</td>
<td></td>
<td></td>
<td></td>
<td>6.4</td>
</tr>
<tr>
<td>2</td>
<td>Demonstrates detailed knowledge about the department professors' idiosyncrasies.</td>
<td>6.20</td>
<td></td>
<td></td>
<td>5.9</td>
</tr>
<tr>
<td>3</td>
<td>Repeats a phone message from a caller back to the caller to be sure it is correct.</td>
<td>5.38</td>
<td>5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Does not share confidential information with another secretary despite her request.</td>
<td></td>
<td></td>
<td></td>
<td>4.86</td>
</tr>
<tr>
<td>5</td>
<td>Does not know the information needed for a university form; checks several sources until she finds the correct answer.</td>
<td>4.11</td>
<td></td>
<td></td>
<td>5.38</td>
</tr>
<tr>
<td>6</td>
<td>Says she is too busy to help another secretary; demonstrates ability to perform routine office task (unjamming copier).</td>
<td>4.78</td>
<td></td>
<td></td>
<td>2.0</td>
</tr>
<tr>
<td>7</td>
<td>Denigrates other secretary to a student. Provides details to the student about how to get into a course that is full.</td>
<td>4.56</td>
<td></td>
<td></td>
<td>1.71</td>
</tr>
<tr>
<td>8</td>
<td>Places a package for a professor in her desk, then forgets to give it to the professor.</td>
<td>2.78</td>
<td>4.1</td>
<td></td>
<td>2.83</td>
</tr>
<tr>
<td>Incident</td>
<td>Description</td>
<td>Job Knowledge</td>
<td>Dealing with Faculty and Students</td>
<td>Working with other secretaries</td>
<td>Organization of work</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>9</td>
<td>Agrees to do a typing job for another secretary. Has difficulty finding a quiz she typed for a professor due to clutter on her desk.</td>
<td></td>
<td>3.78</td>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td>10</td>
<td>Misses a deadline for typing a grant proposal even though she was given extra time.</td>
<td></td>
<td>3.3</td>
<td></td>
<td>2.7</td>
</tr>
<tr>
<td>11</td>
<td>Does low priority work (filing) before high priority work (typing exams).</td>
<td></td>
<td>3.4</td>
<td></td>
<td>2.0</td>
</tr>
</tbody>
</table>
## Configurations of Critical Incidents

### Decreasing Performance

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Incident</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Knowledge</td>
<td></td>
<td>6.2</td>
<td>5.4</td>
<td>4.1</td>
<td>4.8</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>4.7</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Dealing with Faculty and Students</td>
<td></td>
<td>5.9</td>
<td>5.0</td>
<td>4.6</td>
<td>4.1</td>
<td>3.8</td>
<td>3.3</td>
<td>3.4</td>
<td>7</td>
<td></td>
<td></td>
<td>4.3</td>
<td>.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working with Other Secretaries</td>
<td></td>
<td>6.4</td>
<td>5.9</td>
<td>4.9</td>
<td>5.4</td>
<td>2.0</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>4.4</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Organizational Ability</td>
<td></td>
<td>2.8</td>
<td>2.3</td>
<td>2.7</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>2.5</td>
<td>.4</td>
<td></td>
</tr>
<tr>
<td>Average performance</td>
<td></td>
<td>6.4</td>
<td>6.1</td>
<td>5.6</td>
<td>4.9</td>
<td>4.7</td>
<td>3.4</td>
<td>3.1</td>
<td>3.2</td>
<td>3</td>
<td>3</td>
<td>2.7</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Increasing Performance

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Incident</th>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Knowledge</td>
<td></td>
<td>2.8</td>
<td></td>
<td>4.8</td>
<td>4.1</td>
<td>5.4</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>4.7</td>
<td>1.3</td>
</tr>
<tr>
<td>Dealing with Faculty and Students</td>
<td></td>
<td>3.4</td>
<td>3.3</td>
<td>3.8</td>
<td>4.1</td>
<td>4.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>4.3</td>
<td>.9</td>
</tr>
<tr>
<td>Working with Other Secretaries</td>
<td></td>
<td>1.7</td>
<td>2.0</td>
<td>5.4</td>
<td>4.9</td>
<td>5.9</td>
<td>6.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>4.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Organizational Ability</td>
<td></td>
<td>2.0</td>
<td>2.7</td>
<td>2.3</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>2.5</td>
<td>.4</td>
</tr>
<tr>
<td>Average performance</td>
<td></td>
<td>2.7</td>
<td>3</td>
<td>3</td>
<td>3.2</td>
<td>3.1</td>
<td>4.7</td>
<td>4.9</td>
<td>5.6</td>
<td>5.1</td>
<td>6.4</td>
<td></td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Experiment Materials
Subject ID # ______________________

Please list examples of what the secretary did to support or fail to support skill in each of the following competencies.

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

_______________________________________________________

_______________________________________________________

_______________________________________________________

**Organizational ability:** The extent to which the secretary is able to organize 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 her time, and can meet deadlines.

_______________________________________________________

_______________________________________________________

_______________________________________________________

**Dealing with faculty and students:** The extent to which the secretary interacts in a courteous and helpful manner with professors, students, callers, or other people with whom she comes into contact on the job (e.g., expressing willingness to help, etc.).

_______________________________________________________

_______________________________________________________

_______________________________________________________

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

_______________________________________________________

_______________________________________________________

_______________________________________________________
Subject ID # ____________________________

Please list examples of what the secretary did to support or fail to support skill in each of the following competencies.

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

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________

**Organizational ability:** The extent to which the secretary is able to organize 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 her time, and can meet deadlines.

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
Behavior Observation Form, page 2

Subject ID # ______________________

Dealing with faculty and students: The extent to which the secretary interacts in a courteous and helpful manner with professors, students, callers, or other people with whom he or she comes into contact on the job (e.g., demonstrating a positive attitude, being willing to help, etc.).

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

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

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________
1. What is your rating of this secretary's **overall level of performance**?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. What is your rating of this secretary's **overall job knowledge and skill**?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. What is your rating of this secretary's **overall organizational ability (ability to organize)**?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. What is your rating of this secretary's **overall skill in dealing with faculty and students**?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. What is your rating of this secretary's **overall skill in working cooperatively with other secretaries**?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please add any comments you may have:

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________
RATING FORM

Subject ID # ____________________________

1. What is your rating of this secretary’s overall level of performance?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td>Average Performance</td>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. What is your rating of this secretary’s overall job knowledge and skill?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td>Average Performance</td>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. What is your rating of this secretary’s overall organizational ability (ability to organize)?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td>Average Performance</td>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. What is your rating of this secretary’s overall skill in dealing with faculty and students?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td>Average Performance</td>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. What is your rating of this secretary’s overall skill in working cooperatively with other secretaries?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Average Performance</td>
<td>Average Performance</td>
<td>Above Average Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please add any comments you may have:

_________________________________________________________________

_________________________________________________________________

_________________________________________________________________
Please name as many of the 50 United States as you can.

1. ____________________________
2. ____________________________
3. ____________________________
4. ____________________________
5. ____________________________
6. ____________________________
7. ____________________________
8. ____________________________
9. ____________________________
10. ____________________________
11. ____________________________
12. ____________________________
13. ____________________________
14. ____________________________
15. ____________________________
16. ____________________________
17. ____________________________
18. ____________________________
19. ____________________________
20. ____________________________
21. ____________________________
22. ____________________________
23. ____________________________
24. ____________________________
25. ____________________________
26. ____________________________
27. ____________________________
28. ____________________________
29. ____________________________
30. ____________________________
31. ____________________________
32. ____________________________
33. ____________________________
34. ____________________________
35. ____________________________
36. ____________________________
37. ____________________________
38. ____________________________
39. ____________________________
40. ____________________________
41. ____________________________
42. ____________________________
43. ____________________________
44. ____________________________
45. ____________________________
46. ____________________________
47. ____________________________
48. ____________________________
49. ____________________________
50. ____________________________
BEHAVIOR CHECK LIST

Subject ID # __________________________

The statements below describe some behaviors that Sue may have exhibited. Please mark "Y" in front of the behaviors that you think she performed on the video tape. Please mark "N" in front of the behaviors that you think she did not perform on the video tape.

1. ___ Gets the pages out of order when stapling tests for a professor.
2. ___ Compliments another secretary on how organized she is.
3. ___ Picks up another secretary's items from the basement while she was there checking references.
4. ___ Forgets to take a professor's assigned readings to the library to be put on reserve.
5. ___ Repeats a phone message from a caller back to the caller to be sure that it is correct.
6. ___ Accidentally hangs up on a caller.
7. ___ Does low priority work, such as filing, before typing exams, which should be given top priority.
8. ___ Gives incorrect information to a student about a course being taught in Sweden.
9. ___ Demonstrates an ability to perform routine maintenance procedures (e.g., adding toner or removing stuck papers) on the copy machine.
10. ___ 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 home phone number.
11. ___ Checks several different sources in order to find out what to place in several of the boxes on a form rather than leaving them blank.
12. ___ Types several charts and graphs for a professor without any errors.
13. ___ Puts a package that came for a professor in her desk and then forgets to give it to the professor.
14. ___ Agrees to teach another secretary how to do an office procedure that the other secretary is not familiar with.
15. ___ Agrees to work with another secretary in order to get a convention manuscript typed for a professor on short notice.
16. ___ Misses a deadline for typing a grant proposal for a professor even though she was given extra time to complete the assignment.
17. ___ Has difficulty finding a paper she typed due to clutter on her desk.
18. ___ Is unable to answer a simple question asked by a student about dropping a course and does not know the appropriate place to find the answer to the question.
19. ___ Volunteers to do some work for a professor that is not part of her regular job description.
20. ___ Makes a derogatory remark about a co-worker when answering a student's question.
POST-EXPERIMENT QUESTIONNAIRE

1. How well do you think you understood the following tasks in the experiment?
   a) Completing the behavior observation form:
   b) Completing the rating form:
   c) Completing the behavior check list:

2. To what extent do you have confidence that Sue performs well on the job?
   
   Not at all 2 3 4 5
   To some extent To a great extent

3. To what extent did you feel pressured (by time) to complete the behavior observation form?
   
   Not at all 2 3 4 5
   To some extent To a great extent

4. To what extent did you feel pressured (by time) to complete the rating form?
   
   Not at all 2 3 4 5
   To some extent To a great extent

5. To what extent did you feel pressured (by time) to complete the behavior check list?
   
   Not at all 2 3 4 5
   To some extent To a great extent

6. To what extent would you like to work in a job where Sue was your primary secretarial support?
   
   Not at all 2 3 4 5
   To some extent To a great extent

Please explain briefly: