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
MOOD AND PERFORMANCE APPRAISAL QUALITY

DISSERTATION

Presented in Partial Fulfillment of the Requirement for the Degree
Doctor of Philosophy in the Graduate School of
The Ohio State University

By
Robert Franciscus Calderón, M. A.

The Ohio State University

1998

Dissertation Committee:
Professor James T. Austin, Advisor
Professor Robert S. Billings
Professor Patrick Raymark
Professor Mary Roznowski

Approved by
James T. Austin
Advisor
Psychology Department
ABSTRACT

The effect of mood at both time of observing performance behaviors and time of recalling performance behaviors on performance appraisal quality was examined. It was hypothesized that raters in a negative mood would read both negative behaviors which are congruent to their mood as well as positive behaviors in an attempt to regain a desired positive mood. For raters in a positive mood, however, it was hypothesized that they would read only positive behaviors in an attempt to maintain their positive mood. This difference in total number of behaviors read, then, was hypothesized to affect how many behaviors could be recalled and therefore directly affect overall performance appraisal quality. Two hundred and seventy-nine (279) undergraduate students served as participants in the study. The participants were subjected to either a positive or negative mood manipulation at both time of reading and time of recalling performance behaviors. The purpose of the performance appraisal was also manipulated to include both administrative and developmental appraisal conditions for a total of eight experimental conditions. Results indicated that raters in a negative mood did not observe significantly more behaviors than raters in a positive mood. However, results suggest that the interaction of mood at both time of observing and time of recalling behaviors had some effect on the total number of behaviors recalled. Results suggest that neither halo error
nor the accuracy of ratings was affected by mood. Lastly, results suggest that the purpose of the appraisal had no effect on performance appraisal quality. Overall, the direction of several of the results in the study offers some suggestions that mood at both time of observing and time of recalling behaviors may have some effect on performance appraisal quality. Further research is now needed which will take the findings and suggestions of this study into account so that the quality of performance appraisals within organizations can be improved.
Dedicated to My Mother
ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Dr. James T. Austin for his guidance throughout this research. I also wish to thank Dr. Robert S. Billings, Dr. Patrick Raymark, Dr. Mary Roznowski, and Dr. Doug Wolfe for their helpful suggestions and comments. I am grateful to Teresa Bramble, Tammy Mott, Regina Niesen, Dave Plavac, Liat Raby, and Colleen Safford for their valuable input into the study design and data collection. I thank Jennifer Deal for discussing various aspects of the methodology of this dissertation with me. I also wish to thank my colleagues Daniel Russell and Mary Shane Connelly for their assistance with the data analysis and general discussions they had with me about the dissertation. Finally, I would like to thank Allison Roby for motivating and supporting me tremendously throughout the past year.
VITA

August 19, 1969 . . . . . . . . . . . . . Born - Wichita, Kansas

1991 . . . . . . . . . . . . . . . . . . . B.A. Psychology & Statistics. Northwestern University

1994 . . . . . . . . . . . . . . . . . . . M.A. Psychology. The Ohio State University

1993 - 1997 . . . . . . . . . . Graduate Teaching Assistant. The Ohio State University

1997 - present . . . . . . . . . . Research Scientist. American Institutes for Research

PUBLICATIONS


FIELDS OF STUDY

Major Field: Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Performance Appraisal</td>
<td>2</td>
</tr>
<tr>
<td>Affect</td>
<td>4</td>
</tr>
<tr>
<td>Differentiated Affect</td>
<td>5</td>
</tr>
<tr>
<td>Undifferentiated Affect (Mood)</td>
<td>7</td>
</tr>
<tr>
<td>Influence of mood on consumer evaluations</td>
<td>7</td>
</tr>
<tr>
<td>Mood at time of observation</td>
<td>7</td>
</tr>
<tr>
<td>Mood at time of recall</td>
<td>9</td>
</tr>
<tr>
<td>Mood at both time of observation and time of recall</td>
<td>10</td>
</tr>
<tr>
<td>Automatic vs. Controlled Processes</td>
<td>11</td>
</tr>
<tr>
<td>Relationship of Mood &amp; Cognitive Organization</td>
<td>12</td>
</tr>
<tr>
<td>Relationship of Mood &amp; Performance Appraisals</td>
<td>15</td>
</tr>
<tr>
<td>Personality</td>
<td>18</td>
</tr>
<tr>
<td>Current Research Project</td>
<td>20</td>
</tr>
<tr>
<td>2. Method</td>
<td>28</td>
</tr>
<tr>
<td>Study #1</td>
<td>28</td>
</tr>
<tr>
<td>Subjects</td>
<td>28</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Measures</td>
<td>29</td>
</tr>
<tr>
<td>Experimental Task</td>
<td>30</td>
</tr>
<tr>
<td>Design &amp; Procedures</td>
<td>31</td>
</tr>
<tr>
<td>Data Analyses and Discussion</td>
<td>31</td>
</tr>
<tr>
<td>Study #2</td>
<td>34</td>
</tr>
<tr>
<td>Subjects</td>
<td>34</td>
</tr>
<tr>
<td>Measures</td>
<td>35</td>
</tr>
<tr>
<td>Experimental Task</td>
<td>37</td>
</tr>
<tr>
<td>Design &amp; Procedures</td>
<td>38</td>
</tr>
<tr>
<td>Data Analyses and Discussion</td>
<td>41</td>
</tr>
<tr>
<td>Study #3</td>
<td>43</td>
</tr>
<tr>
<td>Subjects</td>
<td>43</td>
</tr>
<tr>
<td>Measures</td>
<td>44</td>
</tr>
<tr>
<td>Experimental Task</td>
<td>48</td>
</tr>
<tr>
<td>Design &amp; Procedures</td>
<td>49</td>
</tr>
<tr>
<td>3. Results</td>
<td>54</td>
</tr>
<tr>
<td>Manipulation Check</td>
<td>54</td>
</tr>
<tr>
<td>Hypothesis 1</td>
<td>60</td>
</tr>
<tr>
<td>Hypotheses 2a, 3a, and 4a</td>
<td>62</td>
</tr>
<tr>
<td>Hypotheses 2b, 3b, and 4b</td>
<td>65</td>
</tr>
<tr>
<td>Hypotheses 2c, 3c, and 4c</td>
<td>69</td>
</tr>
<tr>
<td>Hypotheses 5a, 5b, 5c, and 5d</td>
<td>71</td>
</tr>
<tr>
<td>Hypotheses 6a, 6b, 6c, and 6d</td>
<td>80</td>
</tr>
<tr>
<td>Hypotheses 7a, 7b, 7c, and 7d</td>
<td>86</td>
</tr>
<tr>
<td>Secondary Analyses</td>
<td>92</td>
</tr>
<tr>
<td>4. Discussion</td>
<td>97</td>
</tr>
<tr>
<td>Explanations</td>
<td>97</td>
</tr>
<tr>
<td>Internal Validity</td>
<td>107</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>109</td>
</tr>
<tr>
<td>Statistical Conclusion Validity</td>
<td>110</td>
</tr>
<tr>
<td>External Validity</td>
<td>111</td>
</tr>
<tr>
<td>Future Work</td>
<td>113</td>
</tr>
<tr>
<td>Bibliography</td>
<td>117</td>
</tr>
<tr>
<td>Appendices</td>
<td>122</td>
</tr>
<tr>
<td>Appendix A</td>
<td>123</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ANOVA, means and standard deviations for the positive and negative mood ratings using the PANAS for raters in the positive, negative, and neutral mood conditions in session 1</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>ANOVA, means and standard deviations for the positive and negative mood ratings using the PANAS for raters in the positive, negative, and neutral mood conditions in session 2</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>ANOVA, means and standard deviations for the behaviors read by raters in the positive and negative mood conditions</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>ANOVA, means and standard errors for the behaviors recalled for the raters in the positive and negative mood conditions at both time 1 and time 2</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2</td>
<td>68</td>
</tr>
<tr>
<td>6</td>
<td>ANOVA, means and standard errors for the differential accuracy scores for the raters in the positive and negative mood conditions at both time 1 and time 2</td>
<td>72</td>
</tr>
<tr>
<td>7</td>
<td>ANOVA, means and standard errors for the behaviors read and the behaviors recalled for the raters in the positive and negative mood conditions at both time 1 and time 2 conducting the appraisal for developmental purposes</td>
<td>73</td>
</tr>
<tr>
<td>8</td>
<td>Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2 conducting the appraisal for developmental purposes</td>
<td>77</td>
</tr>
</tbody>
</table>
9 ANOVA. means and standard errors for the differential accuracy scores for the raters in the positive and negative mood conditions at both time 1 and time 2 conducting the appraisal for administrative and developmental purposes ......................................................... 81

10 ANCOVA for the behaviors recalled by the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their emotional stability score .......................................................... 83

11 Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their emotional stability score .......................................................... 87

12 ANCOVA for the behaviors recalled by the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their extraversion score .......................................................... 89

13 Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their extraversion score .......................................................... 93
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Experimental design of Study #1</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>Cell sizes for Study #2</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>Experimental design for Study #2</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>Cell sizes for Study #3</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>Experimental design of Study #3</td>
<td>53</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Rating quality is frequently a concern for anyone involved in employee performance appraisals (Tsui & Barry, 1986). Rating quality can affect administrative, training, and developmental processes in organizations. Employees, managers, and researchers, therefore, are all interested with rating quality, especially if objective performance criteria are not available. However, the presence of objective performance criteria cannot ensure quality ratings. Therefore, great care is needed to either control for or at least be aware of the variables that can affect rating quality. For example, the performance appraisal process cannot be thought of as a "cold" process (i.e., one not involving emotion: Murphy & Cleveland, 1995). It is unreasonable to believe that raters will gather, and eventually remember, information on ratees in a vacuum. Rather, mood will likely play some role in the gathering and recall of information whether the performance criteria are objective or subjective. The purpose of this research is to examine the effects of rater mood on the performance appraisal process. Specifically, how does mood affect the rater's observation and recall of performance criteria? A brief discussion of performance appraisals is provided first. Next, a brief review of affect is
given which is followed by a review of mood with respect to evaluations. This section is followed by an explanation of what mood provides to performance appraisal quality.

Next, a short discussion of personality and how it potentially relates to mood is provided. Lastly, the hypotheses are summarized in order to provide a more global picture of what specifically about mood is of interest and therefore being examined, with respect to performance appraisals. Any relationship between mood and the performance appraisal process that is discovered here can then be addressed in an attempt to ensure higher rating quality.

**Performance Appraisal**

The concept of formally evaluating employees has been around for centuries (Murphy & Cleveland, 1995). However, only recently (i.e., the past 50 years) has the concept of rating employees been an accepted practice in a large number of organizations. According to Murphy and Cleveland (1995), performance appraisals have followed two basic trends over the past 30 years. The first trend deals with how the appraisal methods have moved from more trait-based approaches to more behavioral and results-oriented approaches. This approach has allowed the appraisal of an employee to be more objective as well as provide more specific feedback to employees based upon these more objective criteria. The second trend deals with how the uses of performance appraisal information have expanded. Specifically, for quite some time the performance appraisal system was used primarily for administrative purposes such as salary or promotion decisions. However, performance appraisals have more recently begun to be used for
employee development and feedback, legal documentation, and research (Murphy & Cleveland, 1995). This expansion of performance appraisal usage has allowed for both supervisors and employees to use all of the behavioral information about an employee and work toward raising performance as well as increasing satisfaction levels of both the supervisor and the employee.

As it stands, though, performance appraisals are one of the least popular aspects of the human resources systems in many organizations (Murphy & Cleveland, 1995). Neither supervisors nor subordinates look forward to participating in the performance appraisal process and they are rarely satisfied with the overall results of the process. This dissatisfaction is often due to the fact that the system used by their organization does not help raise performance levels. Therefore, the quality of performance appraisals is an issue that needs to be addressed in order for the satisfaction of both supervisors and employees to be increased as well as for valuable information that can be used by the organization to be obtained (e.g., for employee development, training, and legal documentation). Furthermore, since the appraisal system is usually connected with a negative mood, level of affectivity may need to be examined to determine what effect it has on the entire appraisal process (e.g., while observing behaviors throughout the appraisal cycle as well as while completing the actual appraisal forms). Affect, and more specifically mood, will be discussed next in order to explain how this construct may play a key role in determining the quality of a performance appraisal system within an organization.
Affect

Affect is a term that has been used in numerous ways throughout the psychological literature. In the performance appraisal literature, affect has mainly been defined as a response of liking (Murphy & Cleveland, 1995). In these instances, affect is primarily represented as an emotional reaction to a specific individual. However, affect has also been defined as mood, which represents transient undirected affect, as well as temperament, which represents chronic, more trait-like, undirected affect (Murphy & Cleveland, 1995). Although neither of these latter definitions of affect has received as much exploration as liking in the performance appraisal literature, both appear to have something to offer with regard to the performance appraisal process. For example, because an individual is always experiencing a mood of some sort, this mood may have some effect on the performance appraisal process. It is this conceptualization of affect (i.e., mood which may fluctuate frequently) that will be examined in this research. Although temperament may also offer valuable information with regard to performance appraisal quality (e.g., raters who have an upbeat disposition may rate individuals differently than raters who have a surely disposition), the author is more interested in examining mood. Furthermore, in order to isolate the potential effects of mood on performance appraisal quality, only this conceptualization (i.e., mood) will be examined.

One further relevant distinction concerning affect is the notion of differentiated vs. undifferentiated affect (Longenecker, Jaccoud, Sims, & Gioia, 1992). Longenecker et al. (1992) state that differentiated affect is affect that is directed at something (e.g., a subordinate or the performance appraisal process) whereas undifferentiated affect is a
general affective state (e.g., mood). A brief discussion of these two types of affect follows.

Differentiated Affect

Much of the literature dealing with performance evaluations describes affect as being directed at an object (i.e., differentiated affect; Murphy & Cleveland, 1995). The “likability”, or positive perception, of an individual being rated appears to be the most common operational definition of differentiated affect. When using this distinction, positive and negative differentiated affect refer to how much a rater likes or dislikes the ratee. Several examples of how differentiated affect affects performance evaluations are provided below.

Tsui and Barry (1986) were interested in examining how a rater’s differentiated affect toward a ratee influenced overall ratings. They examined how positive, neutral, and negative differentiated affect influenced rating accuracy. Their results suggest that raters with positive differentiated affect are the most lenient while raters with negative differentiated affect are the least lenient. Furthermore, raters with extreme levels of differentiated affect (i.e., highly positive or highly negative) exhibit the most halo error. Lastly, raters experiencing extreme levels of differentiated affect tend to avoid ratings near the center of the rating scale.

Cardy and Dobbins (1986) were also interested in how differentiated affect influences the accuracy of ratings. Their results suggest that when differentiated affect varies from ratee to ratee (which most likely will happen in a work setting), overall
accuracy decreases. In other words, how much a rater likes a ratee will influence the accuracy of that rater’s final performance evaluation when the rater is responsible for several ratees (presumably differing in how much they are liked).

Ferris, Judge, Rowland, and Fitzgibbons (1994) examined the effects of positive differentiated affect on both subordinate ratings and resource provisions by a supervisor (e.g., inside information, flexible work schedules). Their results suggest that positive differentiated affect leads to increases in supervisor ratings. Furthermore, positive differentiated affect also leads to greater resource allocation indirectly (i.e., through higher ratings) as well as directly. This finding suggests that “liked” subordinates are given higher ratings as well as limited resources. Unfortunately, accuracy was not addressed so the deservedness of the higher ratings and resource provisions can only be assumed.

Lastly, a study by Lefkowitz and Battista (1995) presented results that suggest that a supervisor’s liking of a subordinate had a significant effect on performance ratings (i.e., liking leads to higher ratings, disliking leads to lower ratings). Furthermore, their results suggest that effects of differentiated affect may strengthen over time. This finding suggests that the longer a rater knows a liked ratee, the stronger the differentiated affect will be in relation to their performance evaluations (i.e., ratings will continue to increase over time for liked ratees or decrease over time for disliked ratees).

To summarize, most of the studies that have examined the influence of differentiated affect on performance evaluations have shown that extreme levels of differentiated affect (i.e., positive or negative) influence a rater’s final performance
evaluation. These ratings have been shown to be inaccurate and to exhibit considerable halo error due to the fact that they are based upon level of differentiated affect as opposed to actual levels of performance. Furthermore, most of this research is concerned only with the recall of behaviors in order to determine an overall performance evaluation. A more interesting question than what behaviors are recalled (which is what much of the literature has focused on) may be what type of information is available to the rater, regardless of their affect, at time of recall. The observing and storage of performance information along with undifferentiated affect (i.e., mood) would appear to offer an alternative explanation to the final recall process (as opposed to simply studying likability). Therefore, the effect of mood on what is focused on during the observation of behaviors and subsequent recall of these behaviors will be discussed next.

Undifferentiated Affect (Mood)

It was mentioned earlier that most of the literature dealing with performance evaluations has focused on affect that is directed at something (i.e., likability or positive perception). Therefore, studies that have examined the relationship between mood and other evaluations (e.g., consumer & cognitive organization (Isen, 1984)) will be used to build arguments for any relationship between mood and performance evaluations.

Influence of mood on consumer evaluations.

Mood at time of observation. There is a fair amount of literature that discusses the effects of mood on consumer evaluations of products. Srull (1983a), in his first study, was
interested in how a potential consumer’s mood would affect their ratings of products in advertisements. His results suggest that products featured in positive advertisements are rated more favorably by individuals in a positive mood than by individuals in a neutral or negative mood. Furthermore, products featured in negative advertisements are rated less favorably by individuals in a negative mood than by individuals in either neutral or positive moods. These results suggest that individuals who observe information while in a specific mood (i.e., positive, neutral, or negative) will later recall information in a "mood congruent" (Srull, 1983a) fashion (i.e., an individual in a positive mood rates the product positively, an individual in a negative mood rates the product negatively).

Srull (1983b) examined how a consumer's mood would affect their rating of both familiar and unfamiliar products in advertisements. His results suggested that individuals who are in a positive mood will rate unfamiliar products more favorably than individuals in a negative or neutral mood. However, there was no difference in ratings by individuals in the different moods for products that were familiar to all raters. These results suggest that individuals who observe information in a specific mood will later recall information in a "mood congruent" fashion (i.e., individuals in a positive mood will rate more positively, individuals in a negative mood will rate more negatively) only for information that is of low familiarity to them.

Srull (1984) extended his earlier research by examining how an individual's mood would affect product ratings if the individuals knew they would be making future evaluations. Srull (1984) believed that if it were made explicit to the individuals that future evaluations would occur, an evaluation based upon their mood would be stored
independently from specific information presented to them. His results suggest that an advertisement will be rated more favorably by individuals in a positive mood than individuals in a neutral or negative mood if the individuals are aware that they will be rating the product in the future. Mood had no effect on ratings for those individuals that were not made aware that future ratings would occur. This finding is important because a rater in a performance appraisal setting is always aware that future evaluations of their subordinates will occur and may, therefore, be affected by their moods while making performance appraisal ratings.

Mood at time of recall. Curren and Harrich (1994) were interested in examining whether the importance (i.e., important or unimportant decisions) of the rating along with rater mood would influence the ratings of consumer products. Their results suggest that individuals who are in a positive mood will rate a product more favorably than individuals in a negative mood only if the evaluation is deemed unimportant. If the evaluation was important, then there were no rating differences between individuals in positive or negative moods. This finding is interesting because the evaluation process should be a very important process (both to the employee as well as the organization). Furthermore, the purpose of the performance appraisal (e.g., personnel decisions vs. developmental) could be considered to vary in level of importance. These are issues that will be addressed to ensure that any specific relationships between mood and ratings are interpreted correctly.

Lawson (1985) was interested in examining the effects of mood on product recall and desirability. His results suggest that a "cognitive loop" (i.e., mood congruency effect:
Isen, Shalker, Clark, & Karp. (1978) may occur which means that a person’s mood acts as a recall cue for information that is consistent with the individual’s mood. Specifically, an individual in a negative mood at time of recall will remember more negative than positive information. Furthermore, the individual will rate the product as less desirable than individuals in a positive mood. Also, individuals in a positive mood at time of recall will remember slightly more positive information than negative information.

A study by Knowles, Grove, and Burroughs (1993) presented results that suggest that individuals in a positive mood at time of recall will remember more information about an advertisement than individuals in a negative mood. Furthermore, their results suggest that mood at time of recall has no effect on advertisement or brand evaluations. These results contradict the findings of Srull (1983a) which suggest that further research is needed in order to determine what relationships exist between mood and evaluation.

Mood at both time of observation and time of recall. Lastly, Srull in his third study (1983a) examined the effects of mood at time of observation as well as mood at time of recall. His results suggest that individuals in a positive mood recall more negative than positive attributes of products. Furthermore, individuals in a negative mood recall more positive attributes than individuals in a positive mood. These results are interesting because the findings are somewhat counter-intuitive. One way the literature has described these types of contradictory findings is by suggesting that different processes (i.e., controlled vs. automatic: Sinclair, 1988) occur depending on the mood of the individual. This process distinction will be described next and reasons for its importance will be presented.
**Automatic vs. Controlled Processes**

When a rater is experiencing an automatic process, the rater encounters a stimulus with which they are familiar or have had considerable experience (Sinclair, 1988). The processing of the stimulus then proceeds without conscious awareness on the part of the rater. In other words, upon observing a ratee perform a given task, the rater automatically processes behaviors as relevant or not to the appraisal decision without consciously attending to the encoding process. This automatic process may also occur whenever a rater encounters a piece of information about the ratee that was consistent with the rater’s expectations about the ratee. In relation to mood, then, an automatic process may occur when the rater rates another individual quickly, focusing primarily on positive behaviors (Sinclair, 1988). This process is most likely to occur for individuals who are already in a positive mood. Specifically, an individual in a positive mood possesses a strong desire to maintain their positive mood. Therefore, these raters will only gather enough behavioral information to maintain their positive mood. Furthermore, since congruently positive information will help them best maintain their positive mood the majority of the behavioral information they gather will be positive (Robbins & DeNisi, 1994).

Conversely, a controlled process is under the conscious control of the rater and is seen as intentional (Sinclair, 1988). A controlled process may occur when a rater encounters information that is inconsistent with their expectations of the ratee. In this case, active attention (i.e., controlled process) by the rater is necessary. In relation to mood, then, a controlled process is a process where an individual may take their time rating another individual and focus on all available behaviors (Baron, 1993; Ikegami.
1989; Isen & Daubman. 1984; Robbins & DeNisi. 1994; Sinclair. 1988). This process is most likely to occur for individuals who are in a negative mood. Specifically, an individual in a negative mood will focus on the negative material, which is congruent with their mood, as well as positive material in an attempt to "recover their mood" (Isen & Daubman. 1984). That is, this individual prefers to be in a positive mood and will therefore attempt to move from their current negative mood to a more positive mood (i.e., mood recovery). This notion suggests that a rater who is in a negative mood will attend to more information (both positive and negative) than an individual in a positive mood. If this suggestion is the case, then these differences may significantly affect both what is focused on during the observation and subsequent recall processes of raters. These differences will be discussed later with respect to the hypotheses of this study.

**Relationship of Mood to Cognitive Organization**

There is also some literature that discusses the relationship of mood and cognitive organization (Isen. 1984). Isen and Daubman (1984) examined how mood influences the categorization of stimuli. Their results suggest that positive mood causes a rater to categorize stimuli into broader categories than when a rater is in a neutral mood (i.e., raters who were in the control condition). This finding suggests that raters tend to be more lenient and exhibit more halo error if they are in a positive mood. This idea is further strengthened by a follow-up study by Isen and Daubman (1984) where the results suggest that positive mood causes a rater to include ambiguous stimuli in categories into which they would not normally fit. Considering these results in a performance evaluation
setting, it is suggested that positive mood may lead a rater to include an average employee in a more positive rating category. Lastly, as was mentioned earlier, these researchers suggest that negative mood may not be symmetric with positive mood. That is, although negative mood does lead to more negative ratings, these ratings are not as extreme as those shown for positive mood. Isen and Daubman (1984) suggest that a type of "mood recovery" may take place. That is, individuals may try to focus on positive material in order to alleviate a negative mood. Murphy and Cleveland (1995) support this idea by stating that raters try to avoid negative ratings because they can bring about negative moods. This concept of "recovery" that Isen and Daubman (1984) suggest appears to be one manner in which raters try to deal with a negative mood.

Isen, Johnson, Mertz, and Robinson (1985) discovered similar results to those of Isen and Daubman (1984). Specifically, individuals in a positive mood associated neutral words with words they would normally not be matched with (i.e., unusual associates). These results support what was suggested earlier regarding the possibility that a rater in a performance setting may include an average employee in a more positive setting if the rater is in a positive mood.

Kraiger, Billings, and Isen (1989) were interested in how mood affects job perceptions as well as job satisfaction. Their results suggest that individuals in a positive mood will characterize their job more positively and offer higher ratings of job satisfaction than individuals in a neutral mood. These results are useful in that they are consistent with what would be expected if an automatic process were occurring. That is,
individuals in a positive mood will focus on positive details and therefore produce more positive ratings.

A study by Teasdale and Fogarty (1979) examined how mood affects the recall of positive and negative memories. Their results suggest that individuals in a negative mood recall negative memories quicker than they recall positive memories. These results support the idea that raters in a negative mood have congruent behaviors (i.e., negative mood, negative behaviors) readily accessible. However, as was suggested earlier (Isen & Daubman, 1984), raters should desire a positive state and should therefore attempt to "recover" their mood by focusing on incongruent behaviors (i.e., negative mood, positive behaviors), which was not found in this study.

Lastly, Isen et al. (1978) presented findings that complement those of Teasdale and Fogarty (1979). Specifically, their results suggest that individuals in a positive mood recall positive memories more quickly than they recall negative memories. These results offer further support for the hypothesized automatic process of those individuals in a positive mood. That is, individuals in a positive mood will attempt to maintain their positive mood. They will focus on positive events (or behaviors in an evaluation), and these positive memories will therefore be more accessible due to the fact that they are the only ones available. Although none of the studies mentioned in this section specifically examines the effects of mood on evaluations, they are useful in providing guidance as to how mood may trigger either controlled or automatic processing. As mentioned earlier, if individuals process information differently depending upon their mood, then an
individual's mood may significantly affect what behaviors that individual observes and recalls in a performance appraisal process.

**Relationship of Mood to the Performance Appraisal Process**

Unfortunately, there is little literature examining the effect of mood on the performance appraisal process. Longenecker et al. (1992) surveyed executives to examine the role of mood in performance appraisal settings. They examined how mood is related to the appraisal process in general, the subordinates being rated, and typical events in the performance appraisal interview itself. Longenecker et al. (1992) discovered that mood was related to nearly all aspects of appraisal judgment. Specifically, the executives were generally apprehensive about the performance appraisal process with regards to how subjective it was. The executives also stated that a rater's mood influences a ratee's evaluation. Furthermore, raters agree that being in a good mood can inflate ratings whereas being in a bad mood can deflate ratings. Executives also state that the subordinate's reaction to either praise or reprimand will lead to positive or negative affect. Lastly, raters indicated that feelings about a subordinate can influence the subordinate's ratings. These results suggest that mood influences the recall process in performance evaluations.

Heide and Grønhaug (1993) also examined the effects of mood on evaluations. They had participants rate the country of Norway as a potential travel destination. Participants were provided with a list of aspects that might influence their decision to visit Norway and they were asked to provide ratings for each of these. Their results
suggest that final evaluations are consistently in a mood-congruent direction. That is, positive mood leads to more positive evaluations whereas negative mood leads to more negative evaluations. These findings offer support for the survey results of Longenecker et al. (1992).

Baron (1993) extended the format of the studies already mentioned by including the qualifications of the ratee being evaluated. His results suggest that highly qualified individuals are rated higher regardless of the mood of the rater. However, ambiguously qualified individuals are rated higher by raters in a positive mood than those in a negative mood, similar to the findings of Isen and Daubman (1984). That is, raters may tend to be more lenient and exhibit more halo (in a positive direction for congruence) if they are in a good mood (i.e., an automatic process where the rater only focus on positive aspects of a ratee). Lastly, unqualified individuals were rated the lowest overall. An interesting discovery was made, however, for the unqualified individuals. Results suggest that raters in a negative mood rate the unqualified ratees higher (yet still negative overall) than raters in a positive mood. One potential explanation for this finding deals with mood recovery (Ikegami, 1989; Isen & Daubman, 1984; Robbins & DeNisi, 1994). This concept suggests that a rater in a negative mood will attempt to recover their mood and become more positive by searching for more positive information (i.e., a controlled process where the rater focuses on both the positive and negative aspects of the ratee). This rater will in turn evaluate a ratee more positively than a rater already in a positive mood.

Lastly, Sinclair (1988) examined the effect of order of information acquisition and mood on performance evaluations. His results suggest that a rater in a negative mood
will display less halo error than raters in either neutral or positive moods. Halo was measured by examining correlations among specific dimensions (e.g., delivering lectures, answering questions, availability). The raters in the negative mood produced the lowest correlation between these dimensions suggesting the least halo error. Furthermore, raters in a negative mood displayed greater accuracy than raters in either a positive or neutral mood. Accuracy was measured using the correlation between number of behaviors viewed on a dimension (e.g., availability) and the rating given to the ratee on that dimension. For example, if the individual viewed a large number of positive behaviors, then a high correlation (suggesting greater accuracy) would be present provided the individual rated the ratee positively. Furthermore, if the individual viewed an equal number of positive and negative behaviors, then a neutral rating would be expected (resulting in a high correlation implying greater accuracy). Sinclair (1988) discovered that individuals in a negative mood viewed the most behaviors resulting in the highest correlation which suggests their ratings were most accurate. Sinclair (1988) suggests that these mood differences may be explained by examining automatic vs. controlled processes (supporting what was mentioned earlier). Unfortunately, the study did not specifically examine the influence of mood on the encoding process.

The literature mentioned thus far suggests that rater mood affects evaluations. Unfortunately, little has been done to examine how mood at time of observation as well as mood at time of recall can affect performance appraisal ratings. For instance, positive mood may not simply lead a rater to focus only on positive information at time of recall simply to maintain their positive mood. Instead, a rater in a positive mood may only have
certain information available to them due to their mood at the time of observation. That is, these rater may have only focused on the positive behaviors while observing performance and therefore can only remember these positive behaviors at time of recall. A final factor that will be discussed with respect to evaluations is personality. Rusting and Larsen (1995) suggest that there may be correlations between several personality variables and mood states.

**Personality**

Personality measures specifically designed to predict job performance have become popular over the past several years (Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991). Personality measures, specifically the "Big Five" personality dimensions, have been shown to predict job performance quite well. Part of the improvement appears to be because these personality dimensions can better predict job performance than personality measures in the past have (e.g., MMPI used for clinical purposes: Barrick & Mount, 1991). Both Barrick and Mount (1991) and Tett et al. (1991) performed meta-analyses to examine the ability of personality measures to predict job performance. In both studies, the results suggest that the "Big Five" personality dimensions may be a useful tool in selection procedures as well as other personnel decisions such as performance appraisals.

Although the "Big Five" personality dimensions of the subordinate have been the primary focus in the literature with respect to personnel decisions the personality of the supervisor may also affect the performance appraisal process. For example, Rusting and
Larsen (1995) suggest that several personality dimensions may be correlated with mood states. If this relationship does exist, then personality may also have an affect on the performance appraisal process. Therefore, personality will be examined in this study to examine any relationships (either direct or indirect) with performance appraisal quality.

Much of the personality research has suggested that there are five higher-order personality dimensions (the so-called "Big Five": John, 1990; McCrae & Costa, 1985; Norman, 1963). The first dimension is extraversion which describes the ambition or surgency of an individual as well as the sociability of an individual. The second dimension is emotional stability or neuroticism which consists of being anxious, hostile, depressed, emotional, worried, embarrassed or insecure (Barrick & Mount, 1991; McCrae & Costa, 1985). The third dimension is agreeableness or likability which consists of being courteous, cooperative, good-natured, and tolerant (Barrick & Mount, 1991). The fourth dimension is conscientiousness which consists of being hardworking, achievement-oriented, and persevering (Barrick & Mount, 1991; Digman, 1990). The final dimension is openness to experience which consists of being cultured, imaginative, and intelligent (Barrick & Mount, 1991).

The specific dimensions of interest with respect to mood are extraversion and emotional stability. Rusting and Larsen (1995) as well as John (1990) suggest that extraverts experience more positive moods than introverts as well as desire more positive moods than introverts (suggesting a correlation between extraversion and positive mood). Rusting and Larsen (1995) and John (1990) also suggest that individuals low in emotional stability desire moods that are less activated (e.g., negative, suggesting a
negative correlation between emotional stability and negative mood). It is because of these potential correlations, as well as the suggestion that personality variables may be useful in personnel decisions (such as performance appraisals: Barrick & Mount, 1991) that the "Big Five" personality dimensions will be included in this study for further examination.

Current Research Project

Much of the research involving mood mentioned earlier suggests that the categorization of behaviors can take place through either an automatic or controlled process depending upon the mood of the rater at the time of the evaluation. The same should hold true for the observation period since it is part of the overall evaluation process. That is, a rater in a positive mood should exhibit a greater automatic observation process. A rater already in a positive mood will work to maintain their current mood. This rater will primarily focus on positive performance behaviors gathered about the ratee in order to maintain their positive mood. On the other hand, a rater in a negative mood will work to recover a positive mood state (i.e., individuals desire to be in a positive mood and will work to achieve this). This rater will focus on more performance behaviors (i.e., negative as well as positive) since they will be more attuned to the negative behaviors (congruence), as well as more interested in mood recovery (incongruence: controlled overall process).

Hypothesis 1: Raters who are in a negative mood will attempt to read more performance behaviors of a ratee on average than raters who are in a positive mood.
Furthermore, the mood of the rater, which should influence the number of behaviors a rater focuses on, should also affect the number of behaviors a rater will recall. Specifically, a rater who experiences negative mood congruency (i.e., negative mood at time of observation as well as at time of recall) will recall more behaviors than a rater experiencing positive mood congruency in an attempt to experience mood recovery at time of recall. This rater will have focused on both positive as well as negative performance behaviors and will therefore be able to remember more total behaviors than the rater who experiences positive mood congruency and only focuses on positive behaviors.

**Hypothesis 2a:** Raters who are in a negative mood at time of observation and time of recall will remember more behaviors on average during the rating process than raters who are in a positive mood at both time of observation and time of recall.

Also, because more behaviors will be recalled by a rater who experiences negative mood congruency, there should be more behaviors available to justify the performance ratings they give. Therefore, there should be less error (specifically halo error) as well as greater accuracy by raters that experience negative mood congruency.

**Hypothesis 2b:** Raters who are in a negative mood at time of observation and time of recall will exhibit less halo error on average during the rating process than raters who are in a positive mood at both time of observation and time of recall.

**Hypothesis 2c:** Raters who are in a negative mood at time of observation and time of recall will exhibit greater accuracy on average during the rating process than raters who are in a positive mood at both time of observation and time of recall.
Unfortunately, though, most raters do not experience consistent moods throughout the entire rating process. A rater may experience positive moods during the observation period and then be in a negative mood at time of recall. These raters will attempt to recover their positive moods. However, unlike raters who experience negative mood congruency, the raters who are only in a negative mood at time of recall will not have focused on as many behaviors that can be used to justify their ratings. Because these raters focused on mainly positive behaviors during their observational stages (i.e., automatic process to maintain their positive mood), these raters will make their ratings based upon a limited number of behaviors that are read.

**Hypothesis 3a:** Raters who are in a positive mood at time of observation but a negative mood at time of recall will remember, and therefore base their ratings, on fewer behaviors on average than raters who experience negative mood congruency.

Furthermore, because fewer behaviors will be remembered by those raters that only experience negative mood at time of recall, there should be fewer behaviors available to justify the performance ratings they give. Therefore, there should be more error (specifically halo error) as well as less accuracy by raters that only experience negative mood at time of recall compared to raters that experience negative mood congruency.

**Hypothesis 3b:** Raters who are in a positive mood at time of observation but a negative mood at time of recall will exhibit more halo error on average during the rating process than raters who experience negative mood congruency.
Hypothesis 3c: Raters who are in a positive mood at time of observation but a negative mood at time of recall will exhibit less accuracy on average during the rating process than raters who experience negative mood congruency.

On the other hand, a rater who is in a negative mood during the observation period (i.e., a more controlled process during observation) and then a positive mood during the recall period will be more aware of all potential behaviors than raters who experience positive mood congruency (i.e., an automatic process during observation). These raters who shift from negative mood at time of observation to positive mood at time of recall, therefore, should base their ratings on all of the behaviors they read (which should be greater than the behaviors read by individuals in a positive mood congruency).

Hypothesis 4a: Raters who are in a negative mood at time of observation but a positive mood at time of recall will remember, and therefore base their ratings, on more behaviors on average than raters who experience positive mood congruency.

Furthermore, since more behaviors will be read by those raters who experience negative mood at time of observation, there should be more behaviors available at time of recall to justify the performance ratings given. Therefore, there should be less error (specifically halo error) as well as greater accuracy by raters who experience negative mood at time of observation than those who experience positive mood congruency.

Hypothesis 4b: Raters who are in a negative mood at time of observation but a positive mood at time of recall will exhibit less halo error on average during the rating process than raters who experience positive mood congruency.
Hypothesis 4c: Raters who are in a negative mood at time of observation but a positive mood at time of recall will exhibit greater accuracy on average during the rating process than raters who experience positive mood congruency.

Performance appraisals are used in numerous personnel processes (e.g., administrative, developmental, training processes; see Murphy & Cleveland, 1995). It was mentioned earlier that mood may only affect ratings that are deemed unimportant (Curren & Harrich, 1994). If this were the case, then mood may affect ratings differently depending on the purpose of the rating. For instance, since ratings used in administrative processes may be directly tied to financial gains/losses, these ratings may be considered more important than ratings used in developmental processes. If this were the case, then ratings used in administrative processes may follow a more controlled process and be unaffected by mood. Specifically, raters who use ratings for administrative purposes should gather adequate information during the observation stage and recall all of this information (both accurately and without error) during the recall stage (regardless of mood). Raters who use the appraisal for developmental purposes, however, should follow the patterns of the raters mentioned in the previous hypotheses (and be affected by mood because of the lower importance of their ratings).

Hypothesis 5a: Raters who are conducting an appraisal for developmental purposes will follow the same pattern of observing behaviors as the raters mentioned in hypothesis 1 whereas raters who are conducting an appraisal for administrative purposes will not be affected by mood at all with respect to observing behaviors.
Hypothesis 5b: Raters who are conducting an appraisal for developmental purposes will follow the same patterns of recalling behaviors as the raters mentioned in hypotheses 2a, 3a, and 4a whereas raters who are conducting an appraisal for administrative purposes will not be affected by mood at all with respect to recalling behaviors.

Hypothesis 5c: Raters who are conducting an appraisal for developmental purposes will follow the same patterns of halo error as the raters mentioned in hypotheses 2b, 3b, and 4b whereas raters who are conducting an appraisal for administrative purposes will not be affected by mood at all with respect to halo error.

Hypothesis 5d: Raters who are conducting an appraisal for developmental purposes will follow the same patterns of accuracy as the raters mentioned in hypotheses 2c, 3c, and 4c whereas raters who are conducting an appraisal for administrative purposes will not be affected by mood at all with respect to accuracy.

Lastly, it was mentioned that several personality dimensions may be correlated with mood states (Rusting & Larsen, 1995). Specifically, extraverts may desire more positive moods than introverts and individuals with low emotional stability may desire less activated (i.e., negative) moods than individuals with high emotional stability (Rusting & Larsen, 1995). If this is the case, then an individual’s personality may couple with their mood to affect their performance appraisal quality. For instance, an individual low on emotional stability who is in a negative mood may use an even more controlled process than an individual who is in a negative mood but high on emotional stability. Similarly, an extravert who is in a positive mood may use an even more automatic process than an introvert in a positive mood.
Hypothesis 6a: Raters who are low on emotional stability and in a negative mood will attempt to read more performance behaviors of a ratee on average than any other raters.

Hypothesis 6b: Raters who are low on emotional stability and are in a negative mood at both time of observation and time of recall will remember more behaviors on average during the rating process than any other raters.

Hypothesis 6c: Raters who are low on emotional stability and are in a negative mood at both time of observation and time of recall will exhibit less halo error on average during the rating process than any other raters.

Hypothesis 6d: Raters who are low on emotional stability and are in a negative mood at both time of observation and time of recall will exhibit greater accuracy on average during the rating process than any other raters.

Hypothesis 7a: Raters who are extraverted and in a positive mood will read fewer performance behaviors of a ratee on average than any other raters.

Hypothesis 7b: Raters who are extraverted and are in a positive mood at both time of observation and time of recall will remember fewer behaviors on average during the rating process than any other raters.

Hypothesis 7c: Raters who are extraverted and are in a positive mood at both time of observation and time of recall will exhibit more halo error on average during the rating process than any other raters.

Hypothesis 7d: Raters who are extraverted and are in a positive mood at both time of observation and time of recall will exhibit less accuracy on average during the rating process than any other raters.
This study, then, is an attempt at examining what effect mood has on both the number of behaviors a rater will read as well as the number of behaviors a rater will recall during a performance appraisal session. This study is also aimed at determining what effect, if any, mood has on both halo error as well as the accuracy of the ratings given during the performance appraisal session. Furthermore, the purpose of the appraisal (i.e., administrative vs. developmental) is examined to determine whether the effects of mood are consistent across the different types of performance appraisals. Finally, individual differences (i.e., personality) are included to determine if a rater’s traits can strengthen any of the differences that are discovered within the performance appraisal process based upon the rater’s mood.
CHAPTER 2

METHOD

Two initial studies were conducted in preparation for the main tests of the hypotheses. The first of these initial studies was used to gather ratings of the performance behaviors and examine rating categories to be used for the performance appraisal. The second of these initial studies examined the procedures to be used in the primary study. The results from the second study suggested that several minor changes be made to the primary study. Appendix A contains the materials and measures used in the primary study.

Study #1

Subjects

Thirty-eight (38) undergraduates (14 M & 24 F) taking introductory psychology courses at a large Midwestern university participated in this study. Participants received one credit toward their course requirement for serving as participants in the study. The participants were required to participate at two separate times (48 hours apart). All participants were exposed to the same condition in small groups (n ranging from 2-5).
All participants were treated according to the ethical guidelines of the American Psychological Association (APA, 1982). These participants were used to evaluate potential performance behaviors for the subsequent study. Their task was to rate the positivity/negativity of performance behaviors as well as create rating dimensions to adequately describe the different performance behaviors. No mood manipulations were used in this first study.

Measures

At the beginning of the first experimental session participants completed an adjective checklist (Goldberg, 1992; Saucier, 1994). Forty adjectives were used to obtain scores for the "Big Five" personality dimensions (8 adjectives per dimension). Estimated coefficient alphas were .77 for the emotional stability score. .84 for the extraversion score, .86 for the openness to experience score, .79 for the conscientiousness score, and .81 for the agreeableness score (all comparable to the estimates Saucier (1994) reported). The participants also completed the Positive and Negative Affect Scale (PANAS: Watson, Clark, & Tellegen, 1988) which contains 20 items designed to measure the individual’s mood state. There are ten items designed to measure positive mood (PA: e.g., excited, enthusiastic) and ten items designed to measure negative mood (NA: e.g., upset, irritable). Each item is anchored by how the individual feels “right now”. Estimated coefficient alphas were .91 for positive mood state and .74 for negative mood state (both comparable to the estimates Watson et al. (1988) reported).
At the beginning of the second experimental session, participants completed Form S of the NEO-PI (Costa & McCrae, 1985). Sixty statements were used to obtain scores for the "Big Five" personality dimensions. Participants then completed the PANAS (Watson et al., 1988) again. Estimated coefficient alphas were .89 for positive mood state and .75 for negative mood state (again comparable to those reported by Watson et al. (1988)).

Experimental Task

The experimental task in the first session consisted of participants providing ratings of 49 potential performance behaviors. These potential performance behaviors were critical incidents that research assistants had either experienced themselves or had communicated to them at some point during their college tenure. A group of five research assistants generated positive and negative tasks until a satisfactory group of tasks was available to be used in the experiment. Participants were informed that they would be reading descriptions of potential behaviors of a college professor (i.e., behaviors that the college professor may have performed). Participants then rated the behaviors on a scale of 1 (very negative) to 10 (very positive). This rating was completed in order to ensure that an equivalent number of positive as well as negative behaviors for subsequent studies would be present. It was also meant to ensure that equivalent extremes for both types of behaviors (i.e., positive & negative) would be present.

The experimental task in the second session consisted of the participants sorting the behaviors from the first session into specific rating categories. Participants were also
asked to create names for any behaviors or groups of behaviors that did not fit into one of the rating areas already developed. This task was completed to ensure that all of the behaviors provided during the observation stage (i.e., initial session) would be adequately represented on the appraisal form that was to be used in subsequent studies.

Design & Procedures

All participants began the first session by completing an adjective checklist to measure the "Big Five" personality dimensions (Saucier, 1984). Participants then completed the PANAS (Watson et al., 1988). Lastly, participants rated the 49 potential performance behaviors. Participants were reminded of their second session requirements and excused.

All participants returned 48 hours after their initial session to complete the second session. Following this, the PANAS was completed again. Lastly, the participants placed the performance behaviors from the previous session into specific rating categories. All participants were then debriefed according to the ethical guidelines of the American Psychological Association (APA, 1982) and participants were informed that they could receive more information regarding the experiment by inquiring in a subsequent academic term. Figure 1 offers a visual representation of the experimental design.

Data Analyses and Discussion

The ratings of the potential positive and negative behaviors were analyzed in order to ensure that an equivalent number of positive and negative behaviors would be
Figure 1: Experimental design of Study #1
available for subsequent studies. The average rating of the potential positive behaviors was 8.43 with a standard deviation of 0.73. The average rating of the potential negative behaviors was 3.07 with a standard deviation of 0.87. The potential positive and negative behaviors were significantly different from each other ($t_{46.05}=30.29$ $p<0.00$). Furthermore, the average ratings were on the border between very negative and negative for the potential negative behaviors and very positive and positive for the potential positive behaviors. Lastly, individual behaviors were examined in order to ensure that there were no behaviors between 4 and 7 (the neutral range). Five behaviors fell into this range (all potentially negative behaviors) and these items were reworked in an attempt to make them more negative. These results suggest that an adequate number of positive and negative behaviors were available for the primary study. Furthermore, these results suggest that primarily extreme behaviors (i.e., very positive & very negative) were available which should help strengthen findings in the primary study.

The sorting of behaviors was also examined to ensure that all of the behaviors provided during the observation stage would be adequately represented on the appraisal form to be used in the primary study. A group of 3 subject matter experts (SMEs) (i.e., myself and 2 research assistants) examined the different sorting schemes and determined there were three primary categories represented by the majority of the behaviors. These included: communication skills, defined as a description of how the individual communicates to others; research skills, defined as a behavior that describes the individual’s research; and organizational skills, defined as a behavior that describes how organized the individual is. For those behaviors that did not completely fit into one of
these categories, that behavior was reworded so as to represent one of the three skill categories better. The results suggest that the three skill categories were an adequate representation of all the behaviors that could be read during the initial session and subsequently rated (on these 3 dimensions) for the primary study.

Study # 2

Subjects

Forty-eight (48) undergraduates (17 M & 31 F) taking introductory psychology courses at a large Midwestern university participated in this study. Participants received one credit toward their course requirement for serving in the study. The participants were required to participate at two separate times (48 hours apart). Nineteen (19) participants were randomly assigned to the consistent mood conditions. Eleven (11) participants were in the positive-positive condition and eight were run in the negative-negative condition. The remaining twenty-nine (29) participants were then run in the two shifted mood conditions (15 in the positive-negative condition and 14 in the negative-positive condition). This provided a 2x2 factorial design with 2 between-subject factors. The between-subject factors were mood at time one (positive or negative) and mood at time two (positive or negative). This study was designed to examine the procedures to be used in the primary study paying particular attention to the effectiveness of the mood manipulation. Figure 2 provides a visual representation of the final cell sizes.
Measures

Two types of measures were used in this study: self-report and behavioral measures. At the beginning of the first experimental session, participants completed an adjective checklist (Goldberg, 1992; Saucier, 1994). Forty adjectives were used to obtain scores for the "Big Five" personality dimensions (8 adjectives per personality dimension). Estimated coefficient alphas were .78 for the emotional stability score, .76 for the extraversion score, .82 for the openness to experience score, .77 for the conscientiousness score, and .88 for the agreeableness score (all comparable to estimates Saucier (1994) reported). The participants also completed the Positive and Negative Affect Scale (PANAS; Watson et al., 1988) following the mood manipulation. It contains 20 items designed to measure the individual's mood state. There are 10 items designed to measure positive mood (PA: e.g., excited, enthusiastic) and 10 items designed to measure negative mood (NA: e.g., upset, irritable). Each item is anchored by how the individual feels "right now". Estimated alphas were .87 for positive mood state and .84 for negative mood state (all similar to what Watson et al., (1988) reported).

Behavior measures were the other category of response variable in the first session. Participants had the opportunity to view an hypothetical ratee's performance behaviors. There were two piles of behaviors that contained 24 positive and 24 negative (48 total) potential behaviors that could be viewed. The number of behaviors each participant viewed as well as the order of the viewing made up the encoding section for each participant (i.e., rater).
<table>
<thead>
<tr>
<th>Consistent Condition</th>
<th>Positive at Time 1</th>
<th>Negative at Time 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive -&gt; Positive</td>
<td>N=11</td>
<td>Negative -&gt; Negative</td>
</tr>
<tr>
<td>Inconsistent Condition</td>
<td>Positive -&gt; Negative</td>
<td>N=15</td>
</tr>
</tbody>
</table>

Figure 2: Cell sizes for Study #2
Self-report and behavioral measures were also used in the second session. At the beginning of the second experimental session, participants completed the PANAS (Watson et al., 1988) once again following the mood manipulation. Estimated alpha coefficients were .92 for positive mood state and .88 for negative mood state (all similar to what Watson et al., (1988) reported). The final scale was a demographic questionnaire that assessed gender and any prior experience the participants had with performance appraisals.

Behavioral measures were the other category of response variable in the second session. Participants completed a performance appraisal form for the ratee from the first session. Participants rated each ratee on three specific behavioral categories (i.e., communication skills, research skills, and organizational skills) as well as provided an overall rating. A 5-item Likert scale was used for this purpose ranging from 1: poor to 5: good. The number of behaviors each participant wrote down to justify a rating as well as the rating itself made up the recall section of the session for each participant (i.e., rater).

**Experimental Task**

The experimental task in the first session consisted of participants reading descriptions of as few or as many potential behaviors as they felt necessary in order to provide an accurate performance appraisal during the following session. Each participant was provided with two piles of behaviors. One pile contained 24 positive behaviors while the other contained 24 negative behaviors of the imaginary ratee. Participants were
instructed to read the behaviors in any order they liked and to place the behaviors they had read into one pile (i.e., a third pile). The positive and negative behaviors were chosen based upon the ratings received in the initial study. Minor changes were made to several of the negative behaviors in order to ensure the "negativity" of each statement.

The experimental task in the second session consisted of the participants providing ratings of the imaginary ratee from the previous session. Each participant rated the ratee on three specific behavioral categories (i.e., communication skills, research skills, and organizational skills) as well as overall performance. Participants were instructed to write down comments to justify the ratings given before providing an actual rating. The specific behavioral categories were chosen based upon the sorting task in the initial study.

Design & Procedures

The experimental design was a 2x2 factorial with 2 between-subject factors. The between-subject factors were mood at time one (positive or negative) and mood at time two (positive or negative).

All participants began by completing the "Big Five" adjective checklist (Saucier, 1994). To induce mood, they were asked to think of a recent memory. The induction of the desired mood consisted of the participant being asked to recall and write out a brief description of a recent experience that was either positive or negative. Approximately one-half of the participants were randomly assigned to the positive mood state condition while the other half were randomly assigned to the negative mood state condition. They
were also asked to write out how they felt during the experience (i.e., what feelings occurred as a result of the event). Lastly, they were asked to continue thinking about the event throughout the duration of the experiment. Participants then completed the PANAS (Watson et al., 1988) as a mood manipulation check.

The information gathering portion of the experiment then began as each participant was presented with a fairly vague description of the individual they would be rating. The occupation of the ratee was a college professor. The description consisted of several statements that described the professor in neutral terms (i.e., The professor has been at OSU for several years. The professor is currently teaching one undergraduate upper level course as well as a graduate level seminar. The professor has a broad range of academic interests and is currently involved in several research projects). Each participant then read as many performance behaviors of the professor as they wanted so as to be able to provide an accurate performance appraisal in the following session. Each participant understood which pile contained negative and which pile contained positive behaviors. However, the participants were unaware as to how many total behaviors were in each pile. The participants also had the opportunity to write down any notes about the specific behaviors and were told that these notes would be provided in the following session. Once each participant decided they had read enough behaviors to adequately rate the ratee during the subsequent session, the participant signaled to the experimenter that they were finished and stopped reading the behaviors. Once they finished reading all of the behaviors they chose to review, the first session was completed. In order to assure that no participant would attempt to read all of the possible behaviors, a pile of bogus
sheets was included at the bottom of each pile of positive and negative behaviors.

Furthermore, once a participant finished reading all of the behaviors they chose to review they were given a bogus questionnaire to complete. This technique was used in order to deter a participant from continuing to read more behaviors than they felt necessary simply because other participants in the session were still reading.

The second experimental session constituted the recall portion of the experiment. It was held 2 days after the initial session. Participants began by completing the mood induction where approximately one half of the participants remained in the consistent mood condition, while the other half were shifted to the inconsistent mood condition. Participants in the consistent mood conditions were asked to remember and write down their mood inducing experiences from the previous session. They were also asked to update their experiences if anything new had occurred (along with additional feelings). Participants in the inconsistent mood conditions were asked to recall and write down either a negative or positive mood experience (which ever was the opposite of their session one manipulation). The PANAS (Watson et al., 1988) was then administered as a mood manipulation check. Each participant then rated the professor from the previous session. They rated the professor on the 3 specific behavioral categories (i.e., communication skills, research skills, and organizational skills) as well as provided an overall performance rating. Each participant was instructed to write down as many behaviors as possible for each rating area as a means of justification before making a rating and were given information regarding rating errors they should avoid (i.e., severity, leniency, central tendency, & halo). They were also given their notes on the performance
behaviors from the previous session. Upon completion of the ratings, participants were
given a final demographic questionnaire. All participants were then debriefed and
informed they could receive a report of the experimental results by inquiring in a
subsequent academic term. Figure 3 offers a visual representation of the experimental
design for Study #2.

Data Analyses and Discussion

The mood manipulations were analyzed via the PANAS (Watson et al., 1988) to
determine whether or not the current manipulations would elicit the desired moods (i.e.,
positive & negative). One-way analyses of variance suggested that neither the positive
\((F_{1,46}=1.99, p=.17)\) nor the negative \((F_{1,46}=0.37, p=.55)\) mood manipulations were
successful during the initial session. Specifically, these results suggest that those
individuals who were placed within the negative mood manipulation did not experience a
more negative mood than those placed within the positive mood. Similarly, those
individuals placed within the positive mood manipulation did not experience a more
positive mood than those placed within the negative mood. However, one-way analyses
during the second session were more promising. In the second session, both the positive
\((F_{1,46}=5.32, p=.03)\) and the negative \((F_{1,46}=6.95, p=.01)\) mood manipulations were
significant. This finding suggests that at least for the second session, those individuals
who were in the positive mood condition experienced a more positive mood than
individuals in the negative mood condition whereas those in the negative mood condition
experienced a more negative mood than those individuals in the positive mood condition.

41
Figure 3: Experimental design of Study #2
Still, due to the weak findings with regard to the initial mood manipulation, a stronger mood manipulation was necessary in order to ensure that the desired mood states were achieved in the primary study. In order to strengthen the mood manipulation, several Velten statements (Velten. 1968) were added to both the positive and negative mood manipulations. Specifically, the statements "this is great! I really do feel good! I am elated about things!" and "sometimes I can’t wait to get out of bed in the morning!" were used in conjunction with the positive mood manipulation and the statements "every now and then I feel so tired and gloomy that I’d rather sit than do anything" and "I have too many bad things in my life" were used in conjunction with the negative mood manipulation. The rationale for including these statements was that if the participant was first lead to imagine feeling the Velten (1968) statements, they would have some reference as to which events were positive or negative in their lives. This technique would then lead to more positive and/or negative moods being experienced by that individual.

Study # 3

Subjects

Three hundred and forty-nine (349) undergraduates (127 M & 222 F) taking an introductory psychology course at a large Midwestern university participated in this study. Participants received one credit toward their course requirement for serving as participants in the study. The participants were required to participate at two separate times (48 hours apart). One hundred and forty-one (141) participants were randomly
assigned to the consistent mood conditions. Seventy-one (71) participants were in the positive-positive condition and seventy (70) were run in the negative-negative condition. Another one hundred and forty (140) participants were run in the two shifted mood conditions (70 in the positive-negative condition and 70 in the negative-positive condition). The remaining sixty-eight (68) participants were run in a neutral-neutral control condition. Also, one hundred and seventy (170) participants were randomly assigned to an administrative appraisal condition (distributed evenly for the most part through the mood state conditions) while the remaining one hundred and seventy-nine (179) participants were run in the developmental appraisal condition (distributed evenly for the most part through the mood state conditions). This provided a 2x2x2 factorial design with 3 between-subject factors. The between-subject factors were mood at time one (positive or negative), mood at time two (positive or negative), and rating purpose (administrative or developmental). There was also a control condition where participants were assigned to a neutral mood at both times one and two. This study was designed to examine the hypotheses stated earlier. Figure 4 provides a visual representation of the final cell sizes.

Measures

Two types of measures were used in this study: self-report and behavioral measures. At the beginning of the first experimental session, participants completed the Mood Survey (Underwood & Froming, 1980). It contains 15 items designed to measure the individual's temperament. There are 8 items designed to measure the level of an
### Appraisal to be used for Administrative Purposes

<table>
<thead>
<tr>
<th></th>
<th>Positive -&gt; Positive</th>
<th>Negative -&gt; Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Positive -&gt; Negative</td>
<td>Negative -&gt; Positive</td>
</tr>
<tr>
<td>N</td>
<td>37</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Neutral -&gt; Neutral</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

### Appraisal to be used for Developmental Purposes

<table>
<thead>
<tr>
<th></th>
<th>Positive -&gt; Positive</th>
<th>Negative -&gt; Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Positive -&gt; Negative</td>
<td>Negative -&gt; Positive</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Neutral -&gt; Neutral</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Cell sizes for Study #3
individual's temperament (e.g., "I usually feel quite cheerful") and 7 items designed to measure an individual's reactivity to temperament relevant events (e.g., "I may change from happy to sad and back again several times in a single week"). Each item is anchored by how well a statement "best represents your usual or general behavior." Estimated coefficient alphas were .72 for temperament level and .80 for reactivity to temperament relevant events. Individuals then completed an adjective checklist (Goldberg, 1992; Saucier, 1994). Forty adjectives were used to obtain scores for the "Big Five" personality dimensions (8 adjectives per personality dimension). Each item is anchored by "how accurately the trait describes them." Estimated coefficient alphas were .79 for the emotional stability score, .81 for the extraversion score, .83 for the openness to experience score, .78 for the conscientiousness score, and .84 for the agreeableness score (all comparable to the estimates Saucier (1994) reported). The participants also completed a revised version of the Positive and Negative Affect Scale (PANAS: Watson et al., 1988) following the mood manipulation. It contains 20 items designed to measure the individual's mood state. There are 10 items designed to measure positive mood (PA: e.g., excited, enthusiastic) and 10 items designed to measure negative mood (NA: e.g., upset, irritable). The negative items were the same as those used by Watson et al. (1988). The positive items, however, were slightly different. There has been some concern by Staats, Partlo, Holzapfel, & Morris (1996) with three items on the positive scale (i.e., interested, alert, & attentive). The argument is that these words are measuring a cognitive dimension of positive mood instead of an emotional dimension (Mayer, Salovey, Gomberg-Kaufman, & Blainey, 1991; Staats et al., 1996). Therefore, three positive
words that are markers of emotional positive mood (Mayer et al., 1991) were used in place of the three cognitive markers (i.e., pleasant, cheerful, and joyful). Each item is anchored by how the individual feels "right now". Estimated alpha coefficients were .92 for positive mood and .88 for negative mood (all comparable to the estimates Watson et al. (1988) reported).

Behavior measures were the other category of response variable in the first session. Participants had the opportunity to view an imaginary ratee's performance behaviors. There were two piles of behaviors that contained 24 positive and 24 negative (48 total) potential behaviors that could be viewed. The number of behaviors each participant viewed as well as the order of the viewing made up the encoding section for each participant (i.e., rater).

Self-report and behavioral measures were also used in the second session. At the beginning of the second experimental session, participants completed the revised PANAS (Watson et al., 1988) once again following the mood manipulation. Estimated alpha coefficients were .94 for positive mood state and .89 for negative mood state (all comparable to the estimates Watson et al., (1988) reported). The participants also completed a retrieval task designed to determine whether an individual could accurately identify those behaviors they read during the initial session. There were 10 behaviors listed (5 positive & 5 negative) consisting of 5 that were actual behaviors that could have been read during the initial session whereas the remaining 5 were behaviors that were never presented to the participants. This scale was used to determine whether a more controlled (i.e., high score) or automatic (i.e., low score) process was used by the
participants in reading the behaviors. The final scale was a demographic questionnaire
that assessed gender, and any experiences the participants ever had with performance
appraisals.

Behavioral measures were the other category of response variable in the second
session. Participants completed a performance appraisal form for the ratee from the first
session. Participants rated each ratee on three specific behavioral categories (i.e.,
communication skills, research skills, and organizational skills) as well as provided an
overall rating. A 5-item Likert scale was used for this purpose ranging from 1:poor to
5:good. The number of behaviors each participant wrote down to justify a rating as well
as the rating itself made up the recall section for each participant (i.e., rater).

Experimental Task

The experimental task in the first session consisted of the participants viewing as
few or as many potential behaviors as they felt necessary in order to provide an accurate
performance appraisal in the following session. Each participant was provided with two
piles of behaviors. One pile contained 24 positive behaviors while the other contained 24
negative behaviors of the imaginary ratee. The participants were instructed to read the
behaviors in any order and to place the behaviors they read into one pile (i.e., a third pile).
The positive and negative behaviors were the same behaviors used in Study #2.

The experimental task in the second session consisted of the participants
providing ratings of the imaginary ratee from the previous session. Each participant rated
the ratee on three specific behavioral categories (i.e., communication skills, research

48
skills, and organizational skills) as well as overall performance. The participants were instructed to write down comments to justify the ratings given before providing an actual rating. The specific behavioral categories were the same as those used in Study #2.

**Design & Procedures**

The experimental design was a 2x2x2 factorial with 3 between-subject factors. The between-subject factors were mood at time one (positive or negative), mood at time two (positive or negative), and rating purpose (administrative or developmental). There was also a control condition where participants were assigned to a neutral mood at both times one and two.

All participants began by completing the Mood Survey designed to assess their temperaments (Underwood & Froming, 1980). Next, they completed the “Big Five” adjective checklist (Saucier, 1994). They were then asked to think of a recent memory in order to induce mood. The induction of the desired mood consisted of the participant being asked to “imagine feeling” statements based upon the Velten (1968) task mentioned earlier (i.e., either positive or negative statements). They were then asked to recall and write out a brief description of a recent experience that was either positive or negative. Approximately two-fifths of the participants were assigned to the positive mood state condition while another two-fifths were assigned to the negative mood state condition. They were also asked to write out how they felt during the experience (i.e., what feelings occurred as a result of the event). Lastly, they were asked to continue thinking about the event throughout the rest of the experiment. The remaining one-fifth of the participants
served as the control group and were asked to write down what they remembered from their most recent psychology class. The participants then completed the revised PANAS (Watson et al., 1988) as a mood manipulation check.

The information gathering portion of the experiment then began, as before, by presenting each participant with a fairly vague description of the individual they would be rating. The occupation of the ratee was again a college professor and the description consisted of several statements that described the professor in neutral terms (i.e., The professor has been at OSU for several years. The professor is currently teaching one undergraduate upper level course as well as a graduate level seminar. The professor has a broad range of academic interests and is currently involved in several research projects). Each participant then read as many performance behaviors of the professor as they wanted so as to be able to provide an accurate performance appraisal during the following session. Approximately one-half of the participants were informed that they were gathering behaviors in order to make a rating that was to be used for administrative purposes. The other one-half was informed that they were gathering behaviors in order to make a rating that was to be used for developmental purposes. Each participant understood which pile contained negative and which pile contained positive behaviors. However, the participants were unaware as to how many total behaviors were in each pile. The participants also had an opportunity to write down any notes about the specific behaviors and were told that these notes would be provided to them in the following session. Once each participant decided they had read enough behaviors to adequately rate the ratee during the subsequent session, the participant signaled to the experimenter that

50
they were finished and stopped reading the behaviors. Once they finished reading all of the behaviors they chose to review, the first session was complete. In order to assure that no participant attempted to read all of the possible behaviors, a pile of bogus sheets was included at the bottom of each pile of positive and negative behaviors. Furthermore, once a participant finished reading all of the behaviors they chose to review they were given a bogus questionnaire to complete. This technique was once again used in order to deter participants from continuing to read more behaviors than they felt necessary simply because other participants in the session were still reading.

The second experimental session constituted the recall portion of the experiment. It was held 2 days after the initial session. Participants began by completing the mood induction. Approximately three-fifths of the participants remained in the consistent mood condition (positive, negative, and neutral), while the other two-fifths shifted to the inconsistent mood condition (positive or negative mood only). Participants in the consistent positive or negative mood conditions were asked to remember and write down their mood inducing experiences from the previous session. They were also asked to update their experiences if anything new had occurred (along with additional feelings). The participants in the consistent neutral conditions were asked to once again write down what they remembered from their most recent psychology class. Participants in the inconsistent mood conditions were asked to recall and write down either a positive or negative mood experience (which ever was the opposite of their session one manipulation). The revised PANAS (Watson et al., 1988) was then administered. Each participant then rated the professor from the previous session according to their rating.
purposes (i.e., administrative or developmental). Unfortunately, though, a manipulation check for purpose was not included in this experiment. They then rated the professor on the 3 specific behavioral categories (i.e., communication skills, research skills, and organizational skills) as well as provided an overall performance rating. Each participant was instructed to write down as many behaviors as possible for each rating area as a means of justification before making a rating and were given information regarding rating errors they should avoid (i.e., severity, leniency, central tendency, & halo). They were also given their notes on the performance behaviors from the previous session. Upon completion of the ratings, the participants were given the retrieval questionnaire. Lastly, the participants received a final demographic questionnaire. All participants were then debriefed and informed that they could receive a report of the experimental results when the project was completed by inquiring in a subsequent academic term. Figure 5 offers a visual representation of the experimental design for Study #2.
Figure 5: Experimental design of Study #3
CHAPTER 3

RESULTS

This section begins by examining the effectiveness of the mood manipulation at both time of behavior reading and time of recall. The hypotheses are then examined and the statistical analyses performed to test them are provided. Lastly, several secondary analyses are provided which were not a primary focus of the experiment.

Manipulation Check

The manipulation check was examined first to ensure that the participant’s mood during the session was what it was intended to be. A one-way analysis of variance indicated a significant main effect ($F_{2,348} = 25.00, p=.00$) of mood on the positive scale of the PANAS (Watson et al., 1988). Table 1 provides the ANOVA as well as the means and standard deviations for the three mood conditions examined in session 1. This significant main effect accounted for 13% of the variance in the positive scale score. Follow-up analyses indicated that individuals in the positive mood condition experienced a greater degree of positive mood than those in either the negative mood condition ($t_{275} = 6.48$) or the neutral mood condition ($t_{211} = 5.23$). Individuals in the neutral
### Dependent Variable: Positive Mood at Time 1

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>3751.61</td>
<td>1875.81</td>
<td>25.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>346</td>
<td>25960.74</td>
<td>75.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>348</td>
<td>29712.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.13

<table>
<thead>
<tr>
<th>Mood Type</th>
<th>Positive Mood at Time 1</th>
<th>Negative Mood at Time 1</th>
<th>Neutral Mood at Time 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=141</td>
<td>n=136</td>
<td>n=72</td>
</tr>
<tr>
<td>Positive Mood Rating</td>
<td>33.05</td>
<td>26.45</td>
<td>26.22</td>
</tr>
<tr>
<td>SD</td>
<td>8.36</td>
<td>8.60</td>
<td>9.33</td>
</tr>
</tbody>
</table>

### Dependent Variable: Negative Mood at Time 1

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>1225.43</td>
<td>612.71</td>
<td>12.44</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>345</td>
<td>16987.05</td>
<td>49.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>347</td>
<td>18212.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.07

<table>
<thead>
<tr>
<th>Mood Type</th>
<th>Positive Mood at Time 1</th>
<th>Negative Mood at Time 1</th>
<th>Neutral Mood at Time 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=140</td>
<td>n=136</td>
<td>n=72</td>
</tr>
<tr>
<td>Negative Mood Rating</td>
<td>15.99</td>
<td>20.01</td>
<td>16.61</td>
</tr>
<tr>
<td>SD</td>
<td>6.23</td>
<td>8.14</td>
<td>6.10</td>
</tr>
</tbody>
</table>

Table 1: ANOVA, means and standard deviations for the positive and negative mood ratings using the PANAS for raters in the positive, negative, and neutral mood conditions in session 1.
and negative mood conditions, however, experienced nearly the same level of positive mood ($t_{206, 25}=.17$). A one-way analysis of variance indicated a significant main effect ($F_{2,345}=12.44, p=.00$) of mood as well on the negative scale of the PANAS (Watson et al., 1988). The significant main effect accounted for 7% of the variance in the negative scale score. Follow-up analyses indicated that individuals in the negative mood condition experienced a greater amount of negative mood than those in the positive mood condition ($t_{275.05}=4.61$). However, there was not a significant difference between those in the neutral and positive condition ($t_{210, 25}=.70$) or the neutral and negative condition ($t_{206, 10}=3.40$). These results suggest that the mood manipulation was successful for at least the negative and positive mood conditions (i.e., individuals in the negative mood condition experienced a significantly greater amount of negative mood than those in the positive mood, and individuals in the positive mood condition experienced a significantly greater amount of positive mood than those in the negative mood condition) for the initial session. This is further supported when the neutral mood condition is examined. Specifically, the neutral mood can serve as a baseline condition. In relation to the positive mood manipulation, the results suggest that individuals in the positive mood condition were significantly more positive than individuals in the baseline (i.e., neutral) condition which was expected. Furthermore, the individuals who were in the negative mood condition were not significantly different with respect to positive mood to individuals in the baseline (i.e., neutral) condition which was also expected. Similar results were also provided in the negative mood manipulation. Specifically, individuals in the negative mood condition were significantly more negative than individuals in the
baseline (i.e., neutral) condition which was expected. Furthermore, the individuals who were in the positive mood condition were not significantly different with respect to negative mood to individuals in the baseline (i.e., neutral) condition which also expected.

The mood manipulation during the second session was examined next. A one-way analysis of variance indicated a significant main effect ($F_{2,320} = 31.41, p = .00$) of mood on the positive scale of the PANAS (Watson et al., 1988). Table 2 provides the ANOVA as well as the means and standard deviations for the three mood conditions examined in session 2. This significant main effect accounted for 15% of the variance in the positive scale score. Follow-up analyses indicated that individuals in the positive mood condition experienced a greater degree of positive mood than those in either the negative mood condition ($t_{158, 6} = 8.03$) or the neutral mood condition ($t_{158, 6} = 5.77$). The individuals in the neutral and negative mood conditions, however, experienced nearly the same level of positive mood ($t_{158, 6} = 0.04$). A one-way analysis of variance indicated a significant main effect ($F_{2,320} = 9.47, p < .00$) of mood as well on the negative scale of the PANAS (Watson et al., 1988). The significant main effect accounted for 5% of the variance in the negative scale score. Follow-up analyses indicated that individuals in the negative mood condition experienced a greater level of negative mood than those in the positive mood condition ($t_{158, 6} = 3.88$). However, there was not a significant difference between those in the neutral and positive condition ($t_{158, 6} = 0.09$) or the neutral and negative condition ($t_{158, 6} = 3.35$). These results suggest that the mood manipulation was successful for at least the negative and positive mood conditions (i.e., individuals in the negative mood condition experienced a significantly greater amount of negative mood
### Dependent Variable: Positive Mood at Time 2

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>5480.32</td>
<td>2740.16</td>
<td>31.41</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>346</td>
<td>30187.62</td>
<td>87.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>348</td>
<td>35667.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.15

<table>
<thead>
<tr>
<th>Mood Type</th>
<th>Positive Mood Rating</th>
<th>Negative Mood Rating</th>
<th>Neutral Mood Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=137</td>
<td>n=140</td>
<td>n=72</td>
</tr>
<tr>
<td>M</td>
<td>34.28</td>
<td>26.15</td>
<td>26.21</td>
</tr>
<tr>
<td>SD</td>
<td>9.05</td>
<td>9.32</td>
<td>9.90</td>
</tr>
</tbody>
</table>

### Dependent Variable: Negative Mood at Time 2

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>906.46</td>
<td>453.23</td>
<td>9.47</td>
<td>0.00</td>
</tr>
<tr>
<td>Error</td>
<td>346</td>
<td>16551.88</td>
<td>47.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>348</td>
<td>17458.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.05

<table>
<thead>
<tr>
<th>Mood Type</th>
<th>Positive Mood Rating</th>
<th>Negative Mood Rating</th>
<th>Neutral Mood Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=137</td>
<td>n=140</td>
<td>n=72</td>
</tr>
<tr>
<td>M</td>
<td>15.58</td>
<td>26.15</td>
<td>26.21</td>
</tr>
<tr>
<td>SD</td>
<td>6.45</td>
<td>7.74</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Table 2: ANOVA. means and standard deviations for the positive and negative mood ratings using the PANAS for raters in the positive, negative, and neutral mood conditions in session 2.
than those in the positive mood, and those in the positive mood condition experienced a
significantly greater amount of positive mood than those in the negative mood condition
for the second session as well. Once again, this is further supported when the neutral
mood condition is examined and serves as a baseline condition. In relation to the positive
mood manipulation, the results suggest that individuals in the positive mood condition
were significantly more positive than individuals in the baseline (i.e., neutral) condition
which was expected. Furthermore, the individuals who were in the negative mood
condition were not significantly different with respect to positive mood to individuals in
the baseline (i.e., neutral) condition which was also expected. Similar results were also
provided in the negative mood manipulation. Specifically, individuals in the negative
mood condition were significantly more negative than individuals in the baseline (i.e.,
neutral) condition which was expected. Furthermore, the individuals who were in the
positive mood condition were not significantly different with respect to negative mood to
individuals in the baseline (i.e., neutral) condition which also expected. However, it
cannot be stated that a "neutral" mood manipulation necessarily occurred at either time 1
or time 2. Rather, it can only be suggested that a condition similar to a baseline condition
was produced by the neutral condition. Therefore, since there were no hypotheses that
included the neutral condition, the individuals in this condition were dropped under the
assumption that they did not actually experience a mood manipulation, which was the
focus of the study. The remaining two hundred and seventy-nine (279) participants (100
males and 179 females) were used in the subsequent analyses. All alpha estimates were
recalculated but no changes were detected.
Hypothesis 1

Hypothesis 1 stated that raters in a negative mood would read (i.e., observe) more performance behaviors of a ratee on average than raters who were in a positive mood. A one-way analysis of variance was performed to examine what effect mood had on the total number of behaviors read. Results were expected to yield a significant effect of mood on behaviors read with follow-up analyses showing that raters in a negative mood read significantly more behaviors than individuals in a positive mood.

The one-way analysis of variance for mood on total number of behaviors read was not significant ($F_{1,27} = .08, p = .78$). Table 3 provides the ANOVA as well as the means and standard deviations for the behaviors read by the raters. Although the difference between the average number of behaviors read was in the suggested direction, the difference was not statistically significant and the effect size was negligible ($R^2 = .02$). There was no support, therefore, for Hypothesis 1. The number of both positive and negative behaviors read was also examined to see if raters in a negative mood would read more specific behaviors (i.e., positive or negative) than raters in a positive mood. The rationale here was that raters in a negative mood would read more negative behaviors than raters in a positive mood because these behaviors are seen as mood congruent. Furthermore, raters in a negative mood would also read more positive behaviors than raters in a positive mood because of their attempt to recover their mood state (i.e., move from a negative to a positive mood state). Neither the one-way analysis of variance for mood on positive behaviors ($F_{1,27} = .25, p = .61$) nor the one-way analysis of variance for
Table 3: ANOVA. means and standard deviations for the behaviors read by raters in the positive and negative mood conditions.
mood on negative behaviors ($F_{1,277}=1.30, p=.25$) was significant. Although the differences between the average number of behaviors read were in the suggested direction, these differences were not statistically significant and the effect sizes were small ($R^2=.03$ and $R^2=.07$ respectively). As was mentioned earlier, the ANOVA means and standard deviations for the behaviors read by the raters in these mood conditions are located in Table 3.

**Hypotheses 2a, 3a, and 4a**

Hypothesis 2a stated that raters who were in a negative mood at both time of reading the behaviors and time of recalling the behaviors would recall, and therefore base their ratings, on more behaviors during the rating process than raters who were in a positive mood at both times. Secondly, hypothesis 3a stated that raters who were in a positive mood at time of reading the behaviors but a negative mood at time of recalling the behaviors would recall fewer behaviors during the rating process than raters who were in a negative mood at both times. Lastly, hypothesis 4a stated that raters who were in a negative mood at time of reading the behaviors but a positive mood at time of recalling the behaviors would recall more behaviors on average during the rating process than raters who were in a positive mood at both times. A two-way analysis of variance of mood at time 1 and mood at time 2 on number of behaviors recalled indicated a significant effect ($F_{1,274}=4.18, p=.01$) suggesting that mood did have some effect on the number of behaviors recalled. Specifically, mood accounted for 5% of the variance in total behaviors recalled. Table 4 provides the ANOVA means, and standard errors for
### ANOVA Summary

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>191.73</td>
<td>63.91</td>
<td>4.18</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>274</td>
<td>4191.23</td>
<td>15.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>277</td>
<td>4382.97</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.05

### Behavioral Recall Means and Standard Errors

<table>
<thead>
<tr>
<th>Behavior Type</th>
<th>Positive / Positive</th>
<th>Negative / Negative</th>
<th>Positive / Negative</th>
<th>Negative / Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Behaviors</td>
<td>n=71</td>
<td>n=70</td>
<td>n=70</td>
<td>n=66</td>
</tr>
<tr>
<td></td>
<td>M 8.28</td>
<td>M 8.39</td>
<td>M 9.00</td>
<td>M 10.39</td>
</tr>
<tr>
<td></td>
<td>SE 0.46</td>
<td>SE 0.46</td>
<td>SE 0.47</td>
<td>SE 0.48</td>
</tr>
<tr>
<td>Positive Behaviors</td>
<td>M 4.85</td>
<td>M 4.64</td>
<td>M 5.54</td>
<td>M 6.02</td>
</tr>
<tr>
<td></td>
<td>SE 0.36</td>
<td>SE 0.36</td>
<td>SE 0.36</td>
<td>SE 0.37</td>
</tr>
<tr>
<td>Negative Behaviors</td>
<td>M 3.44</td>
<td>M 3.75</td>
<td>M 3.43</td>
<td>M 4.38</td>
</tr>
<tr>
<td></td>
<td>SE 0.35</td>
<td>SE 0.35</td>
<td>SE 0.36</td>
<td>SE 0.36</td>
</tr>
</tbody>
</table>

Table 4: ANOVA. Means and standard errors for the behaviors recalled for the raters in the positive and negative mood conditions at both time 1 and time 2.
the behaviors recalled. Furthermore, examining the partial sums of squares suggested that it was not just mood at time 1 ($F_{1,274}=2.56, p=.11$) or mood at time 2 ($F_{1,274}=1.88, p=.17$) that accounted for most of this variance in total behaviors recalled, but rather the combination of mood at time 1 and time 2 ($F_{1,274}=8.42, p=.00$).

A number of comparisons were performed to follow up these analyses. A significant difference between raters in a positively mood congruent condition (i.e. positive mood at both time 1 & 2) and raters in a negatively mood congruent condition (i.e., negative mood at both time 1 & 2) was expected. The results of the follow-up Bonferroni test ($t_{139.01}=1.37, p=.08$) indicated that the difference between raters in the negatively mood congruent condition and raters in a positively mood congruent condition was not statistically significant (although it was in the suggested direction). Therefore, there was not complete support for hypothesis 2a.

A significant difference between raters in a positive mood at time of reading behaviors but a negative mood at time of recall (i.e., positive at time 1 but negative at time 2) and raters in a negatively mood congruent condition (i.e., negative mood at both times 1 & 2) was also expected. The results of the Bonferroni test ($t_{138.01}=-7.77, p=.00$) indicated that although there was a significant difference between the number of behaviors recalled by the raters who switched from a positive mood during the reading of behaviors to a negative mood during the recalling of behaviors and those raters who remained in a negatively congruent mood, the direction of the difference was not expected. Specifically, raters who were in a positive mood at time 1 but a negative
mood at time 2 recalled significantly more total behaviors than raters in a negatively mood congruent condition. Therefore, there was not support for hypothesis 3a.

Lastly, a significant difference between raters in a negative mood at time of reading behaviors but a positive mood at time of recalling behaviors (i.e., negative at time 1 but positive at time 2) and raters in a positively congruent mood condition (i.e., positive mood at both times 1 & 2) was expected. The results of the Bonferroni test ($t_{135.01}=26.11, p=.00$) indicated that raters in a negative mood at time of reading behaviors but a positive mood at time of recalling behaviors did recall significantly more behaviors on average than raters in a positively congruent mood condition. Therefore there was support for hypothesis 4a.

**Hypotheses 2b, 3b, and 4b**

Hypothesis 2b stated that raters in a negative mood at both time of reading the behaviors and time of recalling the behaviors would exhibit less halo error on average during the rating process than raters who were in a positive mood at both times. Secondly, hypothesis 3b stated that raters in a positive mood at time of reading the behaviors but a negative mood at time of recalling behaviors would exhibit more halo error on average during the rating process than raters who were in a negative mood at both times. Lastly, hypothesis 4b stated that raters in a negative mood at time of reading behaviors but a positive mood at time of recalling behaviors would exhibit less halo error on average during the rating process than raters who were in a positive mood at both times.
Balzer and Sulsky (1992) offer two conceptual definitions of halo error. These are: “general impression halo” which occurs when the rater’s overall impression of a ratee leads the rater to evaluate all aspects of performance in a consistent manner with this overall impression, and “dimensional similarity halo” which occurs when the rater tends to rate similar dimensions as conceptually related. Balzer and Sulsky (1992) also examined existing literature to determine the most frequently used operational definition of halo error. They discovered that the use of intercategory correlations (which is basically the same as the dimensional similarity halo) was the most used operational definition of halo error. Therefore, this intercategory correlation measure of halo was used to examine the effect of mood on halo error. It was believed, then, that the correlations between the specific rating dimensions would be lower for raters who were in a negative mood at both times compared to raters who were in a positive mood at both times. Furthermore, it was believed that the correlations between the specific rating dimensions would be greater for raters who were in a positive mood at time 1 but a negative mood at time 2 compared to raters who were in a negative mood at both times. Lastly, it was believed that the correlations between the specific rating dimensions would be lower for raters who were in a negative mood at time 1 but a positive mood at time 2 compared to raters who were in a positive mood at both times.

Since the correlations between the specific rating dimensions were independent with respect to the different mood conditions, a Fisher’s $z$ transformation and comparison was computed to determine if the intercategory correlations were different from one another. There was not a significant difference between any of the intercategory
correlations for raters who were in a negative mood at both times compared to raters who were in a positive mood at both times. Specifically, neither the intercategory correlations between the research and communication dimensions ($z=1.14, p=.13$), the intercategory correlations between the communication and organization dimensions ($z=-.42, p=.34$), nor the intercategory correlations between the research and organization dimensions ($z=-.56, p=.29$) were significant. Therefore there was not support for hypothesis 2b. Table 4 provides the intercategory correlations for the raters assigned to these mood conditions.

There was not a significant difference between any of the intercategory correlations for raters who were in a positive mood at time 1 but a negative mood at time 2 compared to raters who were in a negative mood at both times either. Specifically, neither the intercategory correlations between the research and communication dimensions ($z=.22, p=.41$), the intercategory correlations between the communication and organization dimensions ($z=.16, p=.44$), nor the intercategory correlations between the research and organization dimensions ($z=-1.24, p=.11$) were significant. Therefore there was not support for hypothesis 3b. As was mentioned above, table 4 provides the intercategory correlations for the raters assigned to these mood conditions.

Lastly, there was not a significant difference between two of the intercategory correlations for raters who were in a negative mood at time 1 but a positive mood at time 2 compared to raters who were in a positive mood at both times either. Specifically, neither the intercategory correlations between the research and communication
Table 5: Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2. *p < .05.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive / Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.41*</td>
<td>.20</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td></td>
<td>.37*</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative / Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.23</td>
<td>.26*</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td></td>
<td>.45*</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive / Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.27*</td>
<td>.29*</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td></td>
<td>.27*</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative / Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.20</td>
<td>.26*</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td></td>
<td>.15</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
dimensions (z=-1.30, p=.10), nor the intercategory correlations between the research and organization dimensions (z=-1.38, p=.08) was significant. However, there was a significant difference between the intercategory correlations between the communication and organization dimensions (z=1.78, p=.04). Therefore there was partial support for hypothesis 4b. As was mentioned above, table 5 provides intercategory correlations for the raters assigned to these mood conditions.

Hypothesis 2c, 3c, and 4c

Hypothesis 2c stated that raters in a negative mood at both time of reading behaviors and time of recalling behaviors would exhibit greater accuracy during the rating process than raters who are in a positive mood at both times. Secondly, hypothesis 3c stated that raters in a positive mood at time 1 but a negative mood at time 2 would exhibit less accuracy during the rating process than raters who are in a negative mood at both times. Lastly, hypothesis 4c stated that raters in a negative mood at time 1 but a positive mood at time 1 would exhibit greater accuracy during the rating process than raters who are in a positive mood at both times.

There are several forms of accuracy mentioned in the literature. However the one measure that is mentioned both in general performance appraisal literature (Murphy & Balzer, 1989; Sulsky & Balzer, 1988) as well as literature dealing with affect and appraisal accuracy (Cardy & Dobbins, 1986) is differential accuracy. Differential accuracy is calculated by means of the following equation: \( DA^2 = \frac{1}{kn \Sigma (x_{i+}-x_{i-}+x_{-i+}-x_{-i-})} \) where a smaller number means greater accuracy. In this equation \( x_{ij} \) and \( t_{ij} \)
are the rating and true score for ratee $i$ on item $j$; $x_i$ and $t_i$ are the mean rating and mean true score for ratee $i$; $x_j$ and $t_j$ are the mean rating and mean true score for item $j$; and $x$ and $t$ are the mean rating and mean true score over all ratees and items. Also, $k$ represents the number of items being rated and $n$ represents the number of ratees (Murphy & Balzer, 1989; Sulsky & Balzer, 1988). Therefore, differential accuracy was the accuracy measure used in this study.

It should be noted that one of the main concerns with using accuracy scores was addressed in this study; particularly a true score. Specifically, since an equal number of positive and negative behaviors were available for the rater to view, a neutral rating on the different rating dimensions would have been the true score. Therefore, the true score for item $j$, the mean true score for ratee $i$, the mean true score for item $j$, and the mean true score over all ratees were all the same (i.e., $3$) and therefore canceled each other out. It should be noted, though, that the number of behaviors the raters read could have also served as a true score. However, it is believed that the overall results would not have been significantly effected. Nonetheless, the accuracy scores will be recalculated in the near future using these new true scores to examine how the overall accuracy scores change.

It was believed that the differential accuracy score would be lower on average for raters in a negatively congruent mood condition than for raters in a positively congruent mood condition, therefore suggesting greater accuracy. Furthermore, it was believed that the differential accuracy score would be greater on average for raters in a positive mood at time 1 but a negative mood at time 2 than for raters in a negatively congruent mood.
condition, therefore suggesting lower accuracy. Lastly, it was believed that the
differential accuracy score would be lower on average for raters in a negative mood at
time 1 but a positive mood at time 2 than for raters in a positively congruent mood
condition, therefore suggesting greater accuracy. A two-way analysis of variance of
mood at time 1 and mood at time 2 on accuracy scores was not significant ($F_{2,55} = .33,
p = .80$). Therefore, none of the differences were statistically significant and the effect size
was negligible ($R^2 = .00$). There was not support, therefore, for hypotheses 2c, 3c, or 4c.
Table 6 provides the ANOVAs as well as the means and standard deviations for the
accuracy scores for the raters assigned to these mood conditions.

Hypotheses 5a, 5b, 5c, and 5d

Hypothesis 5a stated that raters who were conducting an appraisal for
developmental purposes would follow the same pattern of reading behaviors as the raters
mentioned in hypothesis 1 whereas raters who were conducting an appraisal for
administrative purposes would not be affected by mood at all. A two-way analysis of
variance for mood at time 1 and purpose on total number of behaviors read was not
significant ($F_{2,55} = .27, p = .84$). Furthermore, the combination of mood at time 1 and
purpose accounted for only 0.3% of the variance in number of behaviors read. This
suggests that purpose did not have any effect on the total number of behaviors read.
Therefore, there was no support for hypothesis 5a. Table 7 provides the ANOVA as well
as the means and standard errors for the behaviors read by the raters.
## Table 6: ANOVA. means and standard errors for the differential accuracy scores for the raters in the positive and negative mood conditions at both time 1 and time 2.
### Dependent Variable: Behaviors Read

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>84.09</td>
<td>28.03</td>
<td>0.27</td>
<td>0.84</td>
</tr>
<tr>
<td>Error</td>
<td>275</td>
<td>28057.58</td>
<td>102.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>278</td>
<td>28141.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.00

<table>
<thead>
<tr>
<th>Behavior Type</th>
<th>Positive / Admin</th>
<th>Negative / Admin</th>
<th>Positive / Develop</th>
<th>Negative / Develop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=71</td>
<td>n=67</td>
<td>n=70</td>
<td>n=73</td>
</tr>
</tbody>
</table>

Total Behaviors Read

<table>
<thead>
<tr>
<th>M</th>
<th>24.69</th>
<th>25.35</th>
<th>24.00</th>
<th>24.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>1.20</td>
<td>1.22</td>
<td>1.21</td>
<td>1.21</td>
</tr>
</tbody>
</table>

### Dependent Variable: Behaviors Recalled

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>37.63</td>
<td>12.54</td>
<td>0.79</td>
<td>0.50</td>
</tr>
<tr>
<td>Error</td>
<td>274</td>
<td>4345.34</td>
<td>15.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>277</td>
<td>4382.97</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.01

<table>
<thead>
<tr>
<th>Behavior Type</th>
<th>Positive / Admin</th>
<th>Negative / Admin</th>
<th>Positive / Develop</th>
<th>Negative / Develop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=71</td>
<td>n=67</td>
<td>n=70</td>
<td>n=73</td>
</tr>
</tbody>
</table>

Total Behaviors Recalled

<table>
<thead>
<tr>
<th>M</th>
<th>9.60</th>
<th>8.70</th>
<th>9.01</th>
<th>8.67</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>0.49</td>
<td>0.47</td>
<td>0.48</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 7: ANOVA. means and standard errors for the behaviors read and the behaviors recalled for the raters in the positive and negative mood conditions at both time 1 and time 2 conducting the appraisal for administrative and developmental purposes.
Hypothesis 5b stated that raters who were conducting an appraisal for developmental purposes would follow the same patterns of recalling behaviors as the raters mentioned in hypotheses 2a, 3a, and 4a whereas raters who were conducting an appraisal for administrative purposes would not be affected by mood at all. A two-way analysis of variance for mood at time 2 and purpose on total number of behaviors recalled was not significant ($F_{1,274}=0.79, p=.50$). Furthermore, the combination of mood at time 2 and purpose accounted for only 0.9% of the variance in number of behaviors recalled. This suggests that purpose did not have any effect on the total number of behaviors recalled. Therefore, there was no support for hypothesis 5b. Table 7 also provides the ANOVA as well as the means and standard deviations for the behaviors recalled by the raters.

Hypothesis 5c stated that raters who were conducting an appraisal for developmental purposes would follow the same patterns of halo error as the raters mentioned in hypotheses 2b, 3b, and 4b whereas raters who were conducting an appraisal for administrative purposes would not be affected by mood at all. Specifically, it was hypothesized that raters in a negative mood at both time of reading the behaviors and time of recalling the behaviors and conducting the appraisal for developmental purposes would exhibit less halo error during the rating process than raters who were in a positive mood at both times and conducting the appraisal for developmental purposes. Furthermore, it was hypothesized that raters in a positive mood at time of reading the behaviors but a negative mood at time of recalling behaviors and conducting the appraisal for developmental purposes would exhibit more halo error during the rating process than
raters who were in a negative mood at both times and conducting the appraisal for developmental purposes. Lastly, it was hypothesized that raters in a negative mood at time of reading the behaviors but a positive mood at time of recalling behaviors and conducting the appraisal for developmental purposes would exhibit less halo error during the rating process than raters who were in a positive mood at both times and conducting the appraisal for developmental purposes. The intercategory correlation measure of halo error (Balzer & Sulsky, 1992) was used once again to examine the effect of mood and purpose on halo error.

It was believed, then, that the correlations between the specific rating dimensions would be lower for raters who were in a negative mood at both times and conducting the appraisal for developmental purposes compared to raters who were in a positive mood at both times and conducting the appraisal for developmental purposes. Furthermore, it was believed that the correlations between the specific rating dimensions would be greater for raters who were in a positive mood at time 1 but a negative mood at time 2 and conducting the appraisal for developmental purposes compared to raters who were in a negative mood at both times and conducting the appraisal for developmental purposes. Lastly, it was believed that the correlations between the specific rating dimensions would be lower for raters who were in a negative mood at time 1 but a positive mood at time 2 and conducting the appraisal for developmental purposes compared to raters who were in a positive mood at both times and conducting the appraisal for developmental purposes.

Once again, since the correlations between the specific rating dimensions were independent with respect to the different mood and purpose conditions, a Fisher’s $z'$
transformation and comparison was computed to determine if the intercategory correlations were different from one another. There was not a significant difference between any of the intercategory correlations for raters who were in a negative mood at both times and conducting the appraisal for developmental purposes compared to raters who were in a positive mood at both times and conducting the appraisal for developmental purposes. Specifically, neither the intercategory correlations between the research and communication dimensions (z=1.43, p=.08), the intercategory correlations between the communication and organization dimensions (z=-.37, p=.37), nor the intercategory correlations between the research and organization dimensions (z=.50, p=.31) were significant. Therefore there was not support for this section of hypothesis 5c.

Table 8 provides the intercategory correlations for the raters assigned to these mood and purpose conditions.

There was not a significant difference between any of the intercategory correlations for raters who were in a positive mood at time 1 but a negative mood at time 2 and conducting the appraisal for developmental purposes compared to raters who were in a negative mood at both times and conducting the appraisal for developmental purposes either. Specifically, neither the intercategory correlations between the research and communication dimensions (z=.39, p=.35), the intercategory correlations between the communication and organization dimensions (z=.14, p=.44), nor the intercategory correlations between the research and organization dimensions (z=.18, p=.43) were significant. Therefore there was not support for this section of hypothesis 5c either. As
<table>
<thead>
<tr>
<th>Positive / Positive / Developmental</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.48*</td>
<td>.13</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td>--</td>
<td>.46*</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative / Negative / Developmental</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.17</td>
<td>.22</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td>--</td>
<td>.35*</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive / Negative / Developmental</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td>--</td>
<td>.39*</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative / Positive / Developmental</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Communication Skills</td>
<td>--</td>
<td>.09</td>
<td>.37*</td>
</tr>
<tr>
<td>2. Research Skills</td>
<td>--</td>
<td>--</td>
<td>.13</td>
</tr>
<tr>
<td>3. Organization Skills</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 8: Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2 conducting the appraisal for developmental purposes. *p < .05.
was mentioned above, table 8 provides the intercategory correlations for the raters assigned to these mood and purpose conditions.

Lastly, there was not a significant difference between two of the intercategory correlations for raters who were in a negative mood at time 1 but a positive mood at time 2 and conducting the appraisal for developmental purposes compared to raters who were in a positive mood at both times and conducting the appraisal for developmental purposes either. Specifically, neither the intercategory correlations between the communication and organization dimensions ($z = -1.01, p = .16$), nor the intercategory correlations between the research and organization dimensions ($z = 1.41, p = .08$). However, there was a significant difference between the intercategory correlations between the communication and research dimensions ($z = 1.73, p = .04$). Therefore there was partial support for this section of hypothesis 5c. As was mentioned above, table 8 provides intercategory correlations for the raters assigned to these mood and purpose conditions.

Hypothesis 5d stated that raters who were conducting an appraisal for developmental purposes would follow the same patterns of accuracy as the raters mentioned in hypotheses 2c, 3c, and 4c whereas raters who were conducting an appraisal for administrative purposes would not be affected by mood at all. Specifically, it was hypothesized that raters in a negative mood at both time of reading behaviors and time of recalling behaviors and conducting the appraisal for developmental purposes would exhibit greater accuracy during the rating process than raters who were in a positive mood at both times and conducting the appraisal for developmental purposes. Secondly, it was hypothesized that raters in a positive mood at time 1 but a negative mood at time 2 and
conducting the appraisal for developmental purposes would exhibit less accuracy during the rating process than raters who were in a negative mood at both times and conducting the appraisal for developmental purposes. Lastly, it was hypothesized that raters in a negative mood at time 1 but a positive mood at time 1 and conducting the appraisal for developmental purposes would exhibit greater accuracy during the rating process than raters who were in a positive mood at both times and conducting the appraisal for developmental purposes. Once again differential accuracy (Cardy & Dobbins. 1986; Murphy & Balzer. 1989; Sulsky & Balzer. 1988) was the accuracy measure used in this study.

It was believed that the differential accuracy score would be lower on average for raters in a negatively congruent mood condition and conducting the appraisal for developmental purposes than for raters in a positively congruent mood condition and conducting the appraisal for developmental purposes, therefore suggesting greater accuracy. Furthermore, it was believed that the differential accuracy score would be greater on average for raters in a positive mood at time 1 but a negative mood at time 2 and conducting the appraisal for developmental purposes than for raters in a negatively congruent mood condition and conducting the appraisal for developmental purposes, therefore suggesting lower accuracy. Lastly, it was believed that the differential accuracy score would be lower on average for raters in a negative mood at time 1 but a positive mood at time 2 and conducting the appraisal for developmental purposes than for raters in a positively congruent mood condition and conducting the appraisal for developmental purposes, therefore suggesting greater accuracy. A three-way analysis of variance of
mood at time 1, mood at time 2, and purpose on accuracy scores was not significant
\( (F_{(7,26)}=1.10, p=.36) \). Therefore, none of the differences were statistically significant and
the effect size was moderate \( (R^2=.06) \). There was no support, therefore, for hypothesis
5d. Table 9 provides the ANOVA as well as the means and standard deviations for the
differential accuracy scores for the raters assigned to these mood and purpose conditions.

Hypotheses 6a, 6b, 6c, and 6d

Hypothesis 6a stated that raters low on emotional stability and in a negative mood
would read more performance behaviors of a ratee on average than any other raters (i.e.,
low on emotional stability and in a positive mood, high on emotional stability and in a
negative mood, and high on emotional stability and in a positive mood). A one-factor
analysis of covariance using a combined design with personality as the covariate was
performed to examine what effect personality (i.e., emotional stability) and mood had on
the total number of behaviors read. Results were expected to yield a significant effect of
personality and mood on behaviors read with follow-up analyses showing that raters low
on emotional stability and in a negative mood read significantly more behaviors than any
of the remaining three groups.

The one-factor analysis of covariance of mood and personality on total number of
behaviors read was not significant \( (F_{1,275}=0.28, p=.84) \). Furthermore, the combination of
personality and mood accounted for only .003% of the variance in number of behaviors
read. Therefore, there was no support for hypothesis 6a. Table 10 provides the
Dependent Variable: Differential Accuracy Scores

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>0.86</td>
<td>0.12</td>
<td>1.10</td>
<td>0.36</td>
</tr>
<tr>
<td>Error</td>
<td>261</td>
<td>29.30</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>268</td>
<td>30.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.03

<table>
<thead>
<tr>
<th>Accuracy Score</th>
<th>Positive / Develop</th>
<th>Negative / Develop</th>
<th>Positive / Develop</th>
<th>Negative / Develop</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0.60</td>
<td>0.73</td>
<td>0.65</td>
<td>0.71</td>
</tr>
<tr>
<td>SE</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy Score</th>
<th>Positive / Admin</th>
<th>Negative / Admin</th>
<th>Positive / Admin</th>
<th>Negative / Admin</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0.66</td>
<td>0.55</td>
<td>0.71</td>
<td>0.63</td>
</tr>
<tr>
<td>SE</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 9: ANOVA. means and standard errors for the differential accuracy scores for the raters in the positive and negative mood conditions at both time 1 and time 2 conducting the appraisal for administrative and developmental purposes.
ANCOVA for the individuals assigned to these mood conditions with respect to their personality scores.

Hypothesis 6b stated that raters low on emotional stability and in a negative mood at both time of reading and time of recalling the behaviors would recall, and therefore base their ratings, on more behaviors on average during the rating process than any other raters (i.e., the remaining groups of mood at time 1, mood at time 2, and level of emotional stability). A two-factor analysis of covariance using a combined design with personality as the covariate was performed to examine what effect personality (i.e., emotional stability) and mood had on the total number of behaviors recalled. Results were expected to yield a significant effect of personality and mood at both time of reading and time of recalling the behaviors with follow-up analyses showing that raters low on emotional stability and in a negative mood at both time 1 and time 2 recalled significantly more behaviors than any of the remaining seven groups.

The two-factor analysis of covariance for personality and mood on total number of behaviors recalled was significant ($F = 2.86, p = .01$) suggesting that personality and mood did have some effect on the number of behaviors recalled. Specifically, the combination of personality and mood accounted for 7% of the variance in total number of behaviors recalled. However, the partial sum of squares suggests that it may have not been the interaction of personality, mood at time 1, and mood at time 2 that accounted for most of this variance in total behaviors recalled ($F_{1,7} = 0.87, p = .35$), but rather only mood at time 2 ($F_{1,7} = 6.99, p = .01$), as well as the interaction of personality and mood at time 2 ($F_{1,7} = 6.69, p = .02$). It should be noted that both the correlation between
### Table 10: ANCOVA for the behaviors read and the behaviors recalled by the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their emotional stability score.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>84.98</td>
<td>28.33</td>
<td>0.28</td>
<td>0.84</td>
</tr>
<tr>
<td>Error</td>
<td>275</td>
<td>28056.70</td>
<td>102.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>278</td>
<td>28141.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.00

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>302.60</td>
<td>43.23</td>
<td>2.86</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>270</td>
<td>4080.37</td>
<td>15.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>277</td>
<td>4382.97</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.07
emotional stability and mood at time 1 ($r = -.31, p = .00$ for negative mood; $r = .17, p = .00$ for positive mood) as well as the correlation between emotional stability and mood at time 2 ($r = -.28, p = .00$ for negative mood; $r = .23, p = .00$ for positive mood) were significant and in the intended direction. Nonetheless, since the specific personality, mood at time 1, and mood at time 2 interaction was not significant with respect to the total number of behaviors recalled there was not support for hypotheses 6b. Table 10 provides the ANCOVA for the behaviors recalled by individuals assigned to these mood conditions with respect to their personality scores as well.

Hypothesis 6c stated that raters low on emotional stability and in a negative mood at both time of reading and time of recalling the behaviors would exhibit less halo error on average during the rating process than any other raters (i.e., the remaining groups of mood at time 1, mood at time 2, and level of emotional stability). Once again the intercategory correlation measure of halo (Balzer & Sulsky, 1992) was used to examine the effect of mood on halo error. However, in order to obtain the intercategory correlation measure of halo, emotional stability could not be used as a continuous variable. A median split was one possibility of how to dichotomize emotional stability and allow for the intercategory correlation measure to be calculated and provide practical meaning. The primary problem with this procedure was that it would not differentiate raters adequately who scored on the middle of the scale. Instead, more extreme values were desired so that a label of “low” and “high” emotional stability would provide an accurate distinction between the two and yet not lose too much information. Two methods were examined to achieve this. The first method was a simple trichotomy of the
emotional stability scores. In this instance, raters who scored in the upper one-third would be labeled high on emotional stability, raters who scored in the lower one-third would be labeled low on emotional stability, and raters who scored in the middle one-third would not be used in this analysis as they did not provide a meaningful distinction on the emotional stability scale. The second method was to label anyone who scored more than \( \frac{1}{2} \) of a standard deviation above the mean as high on emotional stability, anyone who scored less than \( \frac{1}{2} \) of a standard deviation below the mean as low on emotional stability, and then not use anyone who scored within \( \frac{1}{2} \) of a standard deviation of the mean as they did not provide a meaningful distinction on the emotional stability scale (Holloway, Russell, & Gustafson, 1998). These two methods were compared and the cutoffs for raters low and high on emotional stability were the same. Therefore, these groupings of raters were used in calculating the intercategory correlation measure of halo error. It was believed, then, that the correlations between rating dimensions would be lower for raters who were low on emotional stability and in a negatively congruent mood condition compared to all other raters (i.e., the remaining 7 groups).

There was not a significant difference between any of the intercategory correlations for raters who were low on emotional stability and in a negatively congruent mood condition compared to all other raters. Specifically, neither the intercategory correlations between the research and communication dimensions (ranging from \( z = .89, p = .19 \) to \( z = -.15, p = .56 \)), the intercategory correlations between the communication and organization dimensions (ranging from \( z = .95, p = .17 \) to \( z = .12, p = .55 \)), nor the intercategory correlations between the research and organization dimensions (ranging
from \( z=1.47, \ p=.07 \) to \( z=-.13, \ p=.55 \) were significant. Therefore there was not support for hypothesis 6c. Table 11 provides the intercategory correlations for the raters assigned to these mood conditions with respect to their personality scores.

Hypothesis 6d stated that raters low on emotional stability and in a negative mood at both time of reading and time of recalling the behaviors would exhibit greater accuracy on average during the rating process than any other raters. Differential accuracy (Cardy & Dobbins, 1986; Murphy & Balzer, 1989; Sulsky & Balzer, 1988) was once again the measure of accuracy used. A two-factor analysis of covariance using a combined design with personality as the covariate was performed to examine what effect personality (i.e., emotional stability) and mood had on the differential accuracy scores. Results were expected to yield a significant effect of personality and mood at both time of reading and time of recalling the behaviors with follow-up analyses showing that differential accuracy scores would be lower for raters low on emotional stability and in a negatively mood congruent condition than for other raters, therefore suggesting greater accuracy. The two-factor analysis of covariance for personality and mood on accuracy scores was not significant \((E_{7,26})=1.70, \ p=.11\). Furthermore, the combination of personality and mood accounted for 4% of the variance in the accuracy scores. Therefore, there was not support for hypothesis 6d.

**Hypothesis 7a, 7b, 7c, and 7d**

Hypothesis 7a stated that raters who are extraverted and in a positive mood would read fewer performance behaviors of a ratee on average than any other raters (i.e.,
Table 11: Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their emotional stability score. *p < .05.
extraverted and in a negative mood, introverted and in a negative mood, and introverted and in a positive mood). A one-factor analysis of covariance using a combined design with personality as the covariate was performed to examine what effect personality (i.e., extraversion/introversion) and mood had on the total number of behaviors read. Results were expected to yield a significant effect of personality and mood on behaviors read with follow-up analyses showing that raters who were extraverted and in a positive mood read significantly fewer behaviors than any of the remaining three groups.

The one-factor analysis of covariance for personality and mood on total number of behaviors read was not significant ($F_{(3,272)}=0.23, p=.87$). Furthermore, the combination of personality and mood accounted for only .003% of the variance in number of behaviors read. Therefore, there was no support for hypothesis 7a. Table 12 provides the ANCOVA for the behaviors read by the individuals assigned to these mood conditions with respect to their personality scores.

Hypothesis 7b stated that raters who are extraverted and in a positive mood at both time of reading and time of recalling the behaviors would recall, and therefore base their ratings, on fewer behaviors on average during the rating process than any other raters (i.e., the remaining groups of mood at time 1, mood at time 2, and extraversion/introversion). A two-factor analysis of covariance using a combined design with personality as the covariate was performed to examine what effect personality and mood had on the total number of behaviors recalled. Results were expected to yield a significant effect of personality and mood at both time of reading and time of recalling behaviors with follow-up analyses showing that extraverted individuals and in a positive
<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>70.92</td>
<td>23.64</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>Error</td>
<td>272</td>
<td>27825.15</td>
<td>102.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>275</td>
<td>27896.07</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**R-Square: 0.00**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>291.04</td>
<td>41.58</td>
<td>2.72</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>267</td>
<td>4086.95</td>
<td>15.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>274</td>
<td>4377.99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**R-Square: 0.07**

Table 12: ANCOVA for the behaviors read and the behaviors recalled by the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their extraversion score.
mood at both time 1 and time 2 recalled significantly fewer behaviors than any of the remaining seven groups.

The two-factor analysis of covariance for personality and mood on total number of behaviors recalled was significant ($F_{7,267}=2.72, p=.01$) suggesting that personality and mood did have some effect on the number of behaviors recalled. Specifically, the combination of personality and mood accounted for 7% of the variance in total number of behaviors recalled. However, the partial sums of squares suggests that it may have not been the interaction of personality, mood at time 1, and mood at time 2 that accounted for most of this variance in total behaviors recalled ($F_{1,267}=1.65, p=.20$), but rather only personality ($F_{1,267}=3.80, p=.05$). It should be noted that both the correlation between extraversion and mood at time 1 ($r=.21, p=.00$ for positive mood; $r=-.18, p=.00$ for negative mood) as well as the correlation between extraversion and mood at time 2 ($r=.23, p=.00$ for positive mood; $r=-.28, p=.00$ for negative mood) were significant and in the intended direction. Nonetheless, since the specific personality, mood at time 1, and mood at time 2 interaction was not significant with respect to the total number of behaviors recalled there was not support for hypotheses 7b. Table 12 provides the ANCOVA for the behaviors recalled by individuals in these mood conditions with respect to their personality scores as well.

Hypothesis 7c stated that raters who are extraverted and in a positive mood at both time of reading and time of recalling the behaviors would exhibit more halo error on average during the rating process than any other raters (i.e., the remaining groups of mood at time 1, mood at time 2, and extraversion/introversion). Once again the intercategory
correlation measure of halo (Balzer & Sulsky, 1992) was used to examine the effect of mood on halo error. Furthermore, both the trichotimization of scores and the $\frac{1}{2}$ standard deviation above and below the mean was used to label raters as either introverted or extraverted. Once again, the two procedures produced the same groupings. These groupings were then used in calculating the intercategory correlation measure of halo. It was believed, then, that the correlations between rating dimensions would be greater for raters who were extraverted and in a positively congruent mood condition compared to all other raters (i.e., the remaining 7 groups).

There was not a significant difference between the majority of the intercategory correlations for raters who were extraverted and in a positively congruent mood condition compared to all other raters. Specifically, although the intercategory correlations between the research and communication dimensions were significantly different when compared to raters who were extraverted and in a negatively congruent mood condition ($z=2.05$, $p=.02$), the remaining correlations between the research and communication dimensions (ranging from $z=1.54$, $p=.06$ to $z=3.38$, $p=.33$) were not significant. Furthermore, although the intercategory correlations between the communication and organization dimensions were significantly different when compared to raters who were extraverted and in negative mood at time 1 but a positive mood at time 2 ($z=2.03$, $p=.02$), the remaining correlations between the communication and organization dimensions (ranging from $z=1.21$, $p=.11$ to $z=.54$, $p=.29$) were not. Lastly, the intercategory correlations between the research and organization dimensions (ranging from $z=1.21$, $p=.11$ to $z=.41$, $p=.34$) were not significant. Therefore there was not support for hypothesis 7c. Table 13
provides the intercategory correlations for the raters assigned to these mood conditions with respect to their personality scores.

Hypothesis 7d stated that raters who are extraverted and in a positive mood at both time of reading and time of recalling the behaviors would exhibit less accuracy on average during the rating process than any other raters (i.e., the remaining groups of mood at time 1, mood at time 2, and extraversion/introversion). Differential accuracy (Cardy & Dobbins, 1986; Murphy & Balzer, 1989; Sulsky & Balzer, 1988) was once again the measure of accuracy used. A two-factor analysis of covariance using a combined design with personality as the covariate was performed to examine what effect personality (i.e., extraversion) and mood had on the differential accuracy scores. Results were expected to yield a significant effect of personality and mood at both time of reading and time of recalling the behaviors with follow-up analyses showing that differential accuracy scores would be greater for raters who are extraverted and in a positive mood congruent condition than for other raters, therefore suggesting lower accuracy. The two-factor analysis of covariance for personality and mood on accuracy scores was not significant ($F_{(1,258)} = 0.30, p = .95$). Furthermore, the combination of personality and mood accounted for only 0.8% of the variance in the accuracy scores. Therefore, there was not support for hypothesis 7d.

Secondary Analyses

A two-way analysis of variance was calculated to detect any mean differences between the ratings given (both on the 3 specific rating scales as well as the overall
Table 13: Intercategory correlations for the raters in the positive and negative mood conditions at both time 1 and time 2 taking into account their extraversion score. *p < .05.
rating) based upon the rater's mood at both time of reading and time of recalling the behaviors. It was believed that a mood at time 1 by a mood at time 2 interaction may occur with rater's in a positively congruent mood condition providing the highest ratings (suggesting lower accuracy and subsequently, lower performance appraisal quality).

The two-way analysis of variance for mood at time 1 and mood at time 2 on the ratings was not significant for either the communication skills ($F_{3,274} = 0.69, \ p = .56$), the research skills ($F_{3,268} = .38, \ p = .77$), or the overall rating ($F_{3,273} = 1.43, \ p = .23$). However, the two-way analysis of variance was significant for the organization skills ($F_{3,274} = 2.61, \ p = .05$) suggesting that mood did have some effect on how the organization skills were rated. Specifically, mood accounted for $3\%$ of the variance in the ratings given to organization skills. However, the partial sums of squares suggests that it may have not been the interaction of mood at time 1 and mood at time 2 that accounted for most of this variance in the rating ($F_{1,274} = 2.90, \ p = .09$), but rather only mood at time 1 ($F_{1,274} = 4.40, \ p = .04$).

A one-way analysis of variance was also calculated to detect any mean differences between the ratings given based upon the rater's mood only at time of recalling the behaviors. It was believed that a main effect of mood may occur with rater's in a positive mood providing the highest ratings (suggesting lower accuracy and subsequently, lower performance appraisal quality).

The one-way analysis of variance for mood at time 2 on the ratings was not significant for any of the specific ratings ($F_{1,276} = 1.70, \ p = .19$ for communication skills;
$F_{1,270} = 0.05, p = .82$ for research skills; ($F_{1,276} = 0.40, p = .53$ for organization skills) or the overall rating ($F_{1,275} = 0.06, p = .81$). Furthermore, mood did not account for any of the variance in the ratings given. This suggests that mood at time of recalling the behaviors did not have any effect on the ratings given.

Rater's temperament was also examined to determine whether it had an effect on either behaviors read or behaviors recalled. A one-factor analysis of covariance using a combined design with temperament as the covariate was performed to examine what effect temperament and mood had on the total number of behaviors read (paying particular attention to the value added by temperament). Results were expected to yield a nonsignificant effect of temperament on the behaviors read. The one-factor analysis of covariance for temperament and mood on total number of behaviors read was nonsignificant ($F_{1,275} = 0.57, p = .63$) suggesting that temperament did not add any value above what was discussed in the original hypotheses that dealt with mood.

A two-factor analysis of covariance using a combined design with temperament as the covariate was also performed to examine what effect temperament and mood at both time of reading as well as time of recalling behaviors had on the total number of behaviors recalled. Results were again expected to yield a nonsignificant effect of temperament on the behaviors recalled. The two-factor analysis of covariance for temperament, mood at time 1 and mood at time 2 on total number of behaviors recalled was significant ($F_{1,270} = 2.15, p = .04$) suggesting that temperament and mood did have some effect on the number of behaviors recalled. However, the partial sums of squares suggests that it may have not been temperament, but rather the interaction of mood at
time 1 and mood at time 2 ($F_{1,270} = 8.16, p = .00$) that accounted for most of the variance in number of behaviors recalled (which followed the same patterns as the original hypotheses mentioned earlier). Furthermore, the amount of variance accounted for (5%) was similar to the amount of variance accounted for by mood at time 1 and mood at time 2 in the original hypotheses. Therefore, temperament did not appear to add any value to the number of behaviors recalled beyond that of mood.
CHAPTER 4

DISCUSSION

This section begins by offering explanations regarding the findings from the study. The four main types of validity are then examined and an explanation of how each was addressed is provided. Lastly, suggestions for future work regarding mood and its relation to performance appraisal quality are offered.

Explanation of Findings

There was weak support for most of the hypotheses suggesting that the differences regarding performance appraisal quality based upon what mood a rater was in during the performance appraisal process may not have been as strong as originally believed. The first area of interest dealt with how many behaviors were read during the session preceding the actual performance appraisal with respect to the mood of the rater. Mood did not have the intended effect on the number of behaviors read (although the direction of the results were in the suggested direction). One reason these effects may have not been as strong was that there was only one session preceding the actual performance appraisal where raters could read behaviors. Typically raters will observe a ratee’s
behavior over an extended period of time (e.g., 6 months or 1 year depending on the frequency of appraisals). Therefore, although the raters in a negative mood a larger number of behaviors on average than the raters in a positive mood, more than one session may have been necessary in order for significant differences to appear. Specifically, it may not be that in every instance where a rater has the opportunity to observe a ratee's behaviors a rater will necessarily observe more behaviors if they are in a negative mood. Instead, over an extended period of time raters who are typically in a negative mood while reading behaviors will indeed observe more behaviors on average than raters who are typically in a positive mood while reading behaviors (i.e., more of a cumulative process vs. one instance). Of course, over time, mood may become more like temperament if the same mood continues to occur. If this is the case, then it will no longer be mood but rather temperament that causes any changes in the number of behaviors observed throughout the performance appraisal process.

The second area of interest dealt with how many behaviors were recalled during the actual performance appraisal with respect to the mood of the rater. Similar to the observation of behaviors, mood did not have the intended effect on the number of behaviors recalled with respect to those raters who remained in a mood congruent condition (although the direction of the results were in the suggested direction). As was the case for the observation of behaviors, this finding may have been due to the fact that there was only one session preceding the actual performance appraisal where raters could read the behaviors to be recalled. Therefore, although the number of behaviors recalled by raters in a negatively mood congruent condition was slightly larger than the number of
behaviors recalled by the raters in a positively mood congruent condition. More than one observation session may have been necessary in order for significant differences in the average number of behaviors recalled to appear. Specifically, it may not be that while conducting a performance appraisal a rater is able to recall more of a ratee’s behaviors if they are in a negatively congruent mood if the rater considers only one behavioral instance (i.e., one observation session). Instead, taking into consideration numerous behavioral instances (i.e., several observation sessions) throughout the performance appraisal cycle, raters who are typically in a negative mood while reading behaviors and are in a negative mood while recalling behaviors may indeed recall more behaviors on average than raters who are typically in a positive mood while reading behaviors and are in a positive mood while recalling behaviors (i.e., an appraisal based upon more of a cumulative process vs. one behavioral instance). Of course, as was mentioned above, if the rater is typically in a negative mood while observing behaviors then this rater may actually be experiencing negative temperament instead of negative mood. If this is the case, then it may be temperament and not mood that is causing a difference in the total number of behaviors recalled. If more sessions are to be included, then temperament will need to be examined closely in relation to mood.

Mood did not have the intended significant effect on the number of behaviors recalled with respect to those raters who switched from a positive mood at time of observing to a negative mood at time of recalling the behaviors compared to those raters who remained in a negatively mood congruent condition either. However, in this case it was not that the effect was simply weak yet in the intended direction. Rather, raters who
were in a positive mood at time 1 but a negative mood at time 2 actually recalled
significantly more behaviors on average than raters in a negatively mood congruent
condition (which was opposite of the suggested direction). One possible explanation is
that raters who switched from a positive mood at time 1 to a negative mood at time 2 may
have felt some concern when it was time to recall (i.e., base their ratings) behaviors.
Specifically, these raters may have only been able originally to recall positive behaviors.
due to their positive mood at time of observing behaviors, which were inconsistent with
their mood state at time of recalling behaviors. Therefore, these raters may have
attempted to recall as well as create additional behaviors that would be consistent with
their mood state. If this occurred, then raters who switched from a positive mood at time
1 to a negative mood at time 2 could have recalled more behaviors (i.e., those that were
actually read combined with those that were created for consistency purposes) than raters
who were in a negatively mood congruent condition who only recalled those behaviors
they actually read. It follows, then, that this again may have been due to the fact that
there was only one session preceding the actual performance appraisal. Specifically, if a
rater was typically in a positive mood while observing behaviors but in a negative mood
while recalling behaviors, they might still feel concern regarding the inconsistency of
their mood state and the types of behaviors recalled (i.e., negative mood but recalling
more positive behaviors). However, the more behavioral instances where these positive
behaviors occurred they could recall would help to increase their sense of reliable ratings
(i.e., an appraisal based upon more of a cumulative process vs. one behavioral instance).
Mood did have the intended effect, however, on the number of behaviors recalled with respect to those raters who switched from a negative mood at time of reading to a positive mood at time of recalling the behaviors compared to those raters who remained in a positively mood congruent condition. Specifically, raters who switched from a negative mood at time 1 to a positive mood at time 2 did recall significantly more behaviors on average compared to raters in a positively mood congruent condition. This finding suggests that raters who were in a negative mood at time 1 but a positive mood at time 2 actually recalled several of the negative behaviors they read to compliment the positive behaviors they read whereas those who remained in a positively congruent mood primarily recalled only those behaviors that they read (i.e., positive behaviors). One potential explanation as to why this difference in behaviors recalled did occur while the differences mentioned above did not deals with the desire of an individual to remain in a positive mood (Robbins & DeNisi, 1994; Sinclair, 1988). Specifically, if a rater was in a consistently positive mood at both time of observing and recalling behaviors they would not be concerned with the lack of negative behavioral instances while making their rating but rather make the rating based upon those behavioral instances that are readily available to them (i.e., an automatic process). On the other hand, the rater who was in a negative mood while observing behaviors but a positive mood while recalling the behaviors would recall primarily the positive behaviors to remain in a positive mood, but also some of the negative behaviors since there was an inconsistency in their mood states (which could lead to some concern). One interesting thought, though, is that if there was more than one session preceding the actual performance appraisal the strength of the effect would
probably remain unchanged. Specifically, if a rater was typically in a negative mood while observing behaviors but in a positive mood while recalling behaviors, they might feel some concern regarding the inconsistency of their mood state and the types of behaviors recalled (i.e., positive mood but recalling that there were negative behaviors). The more behavioral instances they read, however, would help to increase their sense of reliable ratings (i.e., an appraisal based upon more of a cumulative process vs. one behavioral instance). Furthermore, if a rater was typically in a positive mood while observing behaviors and a positive mood while recalling behaviors, they would recall primarily positive behaviors (which was consistent with what they read) and not experience this concern (i.e., follow an automatic process that would maintain their positive mood). Once again, though, the concept of mood versus temperament needs to be considered when numerous observational sessions are provided. It may be that temperament, and not mood, will cause any differences in the number of behaviors recalled when numerous observational sessions are provided.

The third area of interest dealt with how much halo error would be present during the rating process. The findings with respect to halo error suggest that mood did not have the intended effect on halo error. These results suggest that the ratings on the different dimensions may not only contribute to the overall performance rating, but also to each other (suggesting halo error). One explanation for this may be the assumption by the rater that a higher level of communication is necessary in order to produce a higher quality of research. Specifically, they may believe that the professor needs to communicate their findings adequately to get recognized and therefore create a bond between these two
dimensions. Furthermore, the rater in this study (i.e., students) may not be totally familiar with all aspects of a professor’s research. Along these same lines, the assumption by a rater that a higher level of organization is necessary in order to produce a higher quality of research may have occurred. If this were the case, then a strong correlation between the two factors would make sense. Furthermore, since the participants (i.e., undergraduate students) were probably more familiar with a professor’s organization skills compared to their research skills (e.g., through their interactions with their own professors), they may have assumed this correlation to be stronger than it necessarily was. Lastly, one explanation for this may be the assumption by a rater that a higher level of organization will always lead to a higher level of communication. If this were the case, then a strong correlation between the two factors would make sense. Furthermore, since the participants (i.e., undergraduate students) were probably more familiar with a professor’s organization skills with respect to their communication skills (e.g., via a lecture), they may have assumed this correlation to be stronger than it necessarily was. The general knowledge of the rater in relation to these dimensions may have been valuable in determining if the potential biases mentioned above were indeed true. Specifically, if the raters did not totally understand the breadth of the rating dimensions, then halo error may have occurred due to the rater’s lack of knowledge with respect to the performance dimensions.

The fourth area of interest dealt with how much accuracy would be present during the rating process. The findings with respect to accuracy suggest that neither mood nor appraisal purpose had the intended effect on the accuracy scores. Although in several
instances the average rating scores were in the intended direction (e.g., the differential accuracy scores for raters in the negatively congruent mood condition were lower than the differential accuracy scores for the raters in the positively congruent mood condition which suggests greater accuracy). None of these differences were significant. There are several explanations as to why accuracy was not affected in the hypothesized manner. One is that the rating dimensions may not have been specific enough. Specifically, the dimensions that the participant rated the professor on may have been too general to make ratings that actually varied. For example, the rating dimension labeled "organization skills" may have been too broad to allow for differences in accuracy scores. Providing dimensions that were more concrete (e.g., management of projects or supervision and staff development) and contained specific tasks that were mapped to a BARS scale may have allowed for greater variability in the ratings. This in turn may have allowed mood to have a greater impact on accuracy scores. However, as was mentioned above, had the raters possessed adequate knowledge in relation to the different rating dimensions, the breadth of the rating dimensions may have been fine.

Another explanation deals with the potential relationship between accuracy and halo error. The instructions on the performance appraisal form asked the raters to avoid central tendency, along with other rater errors, when in fact ratings right in the middle of the rating scale (i.e., central tendency) may have been the most accurate. Therefore an overall decrease in accuracy may have been built into the study. Specifically, to avoid the rater errors described in the performance appraisal directions, and therefore halo error, the participants may have avoided providing accurate ratings. To support this explanation it
can be noted that a number of the hypotheses regarding halo error mentioned above were supported. This suggests that the accurate ratings may have been very difficult to achieve.

A final explanation may pertain to when in the rating process the rater actually provided a rating on each dimension. Specifically, although raters were told to write down behaviors they could use to justify a rating before making their final rating on each dimension, many stated they did the reverse. That is, these raters provided a rating and then wrote down behaviors that may have only been used to justify the ratings already given. For example, a participant may have given the professor a low rating and then only worked to recall those behaviors that would justify this low rating rather than recall all behaviors they had observed. This, in turn, would lead to a decrease in accuracy for all these individuals regardless of mood.

The fifth area of interest dealt with the purpose of the performance appraisal and how this might effect the entire rating process with respect to the mood of the rater. The results suggest that the purpose of the appraisal did not play a significant role regarding performance appraisal quality. Specifically, purpose did not appear to contribute anything additional to the rating quality beyond that of mood at both time of observing and time of recalling behaviors. One explanation for this finding is that the manipulation of the purpose may not have been strong enough. Specifically, many of the participants in the study may have been more familiar with the concept of an administrative purpose of a performance appraisal in general. Therefore, the concept of using a performance appraisal for any other purpose (e.g., development) could have appeared foreign to them.
Another related explanation is that the raters may have been unable to separate the administrative purpose from the developmental purpose. Specifically, the raters may have assumed that even though it was stated that the appraisal was to be used for developmental purposes, somewhere in the future it would have to also be used for administrative purposes. Therefore, every rater would have completed the appraisal assuming the ratings would be used for some sort of administrative purpose at some point. In either case, though, the purpose of the appraisal did not appear to have any effect on the performance appraisal quality.

The final area of interest dealt with how the rater's personality might covary with their mood to affect performance appraisal quality. The results suggest that the personality of the rater did not play a significant role regarding performance appraisal quality. Specifically, the correlations between emotional stability and mood as well as the correlations between extraversion and mood were significant in the suggested direction. It was this presumed connection between these personality traits and mood states which lead to the inclusion of the personality variables to examine their effect on performance appraisal quality. However, neither emotional stability nor extraversion were shown to significantly influence either the number of behaviors read or the number of behaviors recalled in the suggested fashion. Furthermore, these personality variables did not contribute any additional explanation to the differences in behaviors read or recalled beyond that of what mood at time 1 and mood at time 2 were able to explain. One explanation for this finding is that although the relationships between the personality traits and the mood states were significant, these relationships may have been weak.
relative to other combinations of variables (e.g., temperament and personality). It was the case that the relationship between temperament and these personality traits was much higher than the relationship between the mood states and the personality traits. Therefore, it may not have made sense to include these trait variables with the state variables in examining what effect they had on performance appraisal quality.

Internal Validity

The first issue that was examined with respect to internal validity dealt with the mood manipulation. The manipulation check suggested that the mood the participants experienced was equivalent to that intended in the positive and negative mood conditions during both the first and second sessions (which were the mood conditions from which the hypotheses were derived). Also, in order to increase the participants mood state, the participant's were asked to actually write out both the summary of their mood inducing event as well as the feelings they associated with the event. This was intended to make the intended moods more salient to the participants. Furthermore, the participants were occasionally reminded to continue thinking about the statements and corresponding feelings they wrote down during the mood manipulation section. These are all factors that should have enhanced internal validity.

Another issue that was examined with respect to internal validity dealt with the behaviors to be read. The participants had the opportunity to read an equivalent number of positive and negative behaviors of equivalent extremes (i.e., the positive were quite positive and the negative were quite negative) as was determined in the initial study.
Furthermore, the participants had the opportunity to read an equivalent number of behaviors from the different rating dimensions (i.e., communication, research, and organization skills) which were determined to be adequately exhaustive and mutually exclusive as determined by the initial study (i.e., through both participants sorting of the behaviors and SME judgment). Also, the behaviors to be read were ordered in the exact same order for each participant. Therefore, all participants could read the exact same behaviors. Specifically, if two independent raters each read eleven positive and fifteen negative behaviors, they would have each read the exact same behaviors although the order they read each individual behavior may have differed (i.e., all eleven positive followed by all fifteen for one rater versus some random order for the second rater). Lastly, the piles of positive and negative behaviors were randomly placed on opposite sides of the participants so that handedness of the participants would not be a factor. Once again, these all should have enhanced the internal validity.

One final potential threat to internal validity is related to the occupation of the ratee the participants were observing behaviors of (i.e., a college professor). Although this occupation was chosen so that every participant would be familiar with the occupation and what was both acceptable and unacceptable behavior, the participant’s own recent experiences with their own professors may have biased their ratings. Specifically, if a participant recently had either a very negative or very positive experience with a professor then they may have related this to their ratings in this study. In this instance mood may not have been effecting the behaviors read or recalled but rather their recent experiences with a similar professor.
Construct Validity

The primary area where construct validity needs to be examined in this study is with respect to mood. Specifically, the question of whether it was mood or likability (or even temperament) that was the type of affect that was affecting performance appraisal quality in this study is critical. There were, however, two precautions taken in this study to avoid this potential problem. First of all, the participants never met the actual ratee. Therefore, the whole concept of likability which is normally formed through extended experience with a ratee should not have occurred in this experiment. However, if the participants were able to relate a professor they knew to the hypothetical professor in the experiment, then likeability, rather than mood, could have been the form of affect experienced by the participants in the study. The second precaution was to ask the participants to rate how they were feeling “right now” during the experimental session. Therefore, a mood state, rather than their temperament was the type of affect being measured by the manipulation check in this study which suggests that participants in a positive mood manipulation condition did indeed experience a positive mood and participants in a negative mood manipulation condition did indeed experience a negative mood.

Another area where construct validity could be questioned in this study was with respect to mood manipulation itself. Although the mood manipulation did produce a significant difference between the expected mood conditions, the variance accounted for was not very large. Therefore, the mood states may have not been strong enough to produce the intended results. The mood manipulation chosen was done so as to allow the
experimental session to appear more like the appraisal process. That is, the participants were requested to think of a recent mood which would be similar to the moods the supervisor is experiencing during the appraisal process. Nonetheless, the mood manipulation does suggest that the combination of the Velten (1966) and imagination task may have been a weak manipulation.

Statistical Conclusion Validity

Statistical conclusion validity examines how sensitive a study is in its ability to detect covariation between the independent and dependent variables, as well as how strongly these variables covary (Cook & Campbell, 1979). The basic question asks whether there is something that suggests the conclusions are accurate. The three suggested criteria to be used when evaluating any covariance are statistical power, statistical significance, and effect size (Austin, Boyle, & Lualhati, 1994; Cook & Campbell, 1979).

Violations of the assumptions for analysis of variance (ANOVA) were examined first. The participants were randomly assigned to the conditions in an attempt to obtain independence of the behaviors read and the behaviors recalled. Also, normal probability plots suggested normality of treatment populations. It appeared, therefore, that the assumptions were not violated for ANOVA.

Next, most of the effect sizes obtained were moderate (r=.06) according to Cohen and Cohen (1983). This suggests that for those findings where the independent variables
(e.g., mood at time 1 & time 2) did covary with the dependent variables (e.g. behaviors read and behaviors recalled), the magnitude of these relationships was moderate.

Finally, the statistical power was examined to determine if the study was sufficiently powerful enough to detect any differences that may have been present in the population. As was mentioned above, the effect size was moderate for the most part. Furthermore, the sample size was typically between 35 and 40 (with respect to mood at time 1, mood at time 2, and purpose of the appraisal) and an alpha-level of .05 was used. Therefore, the power of the experiment was approximately .80. It should be noted that a power analysis was calculated prior to conducting an experiment in an attempt to ensure that a power-level of about .80 would be obtained. Therefore, the statistical power was acceptable for this study.

External Validity

The study was designed to generalize to the work setting. The intention was to replicate the fashion in which behaviors are observed in a work setting and then examine any differences in rating quality due to the differences in the behaviors that could be recalled. Since mood is always present in every individual’s life (e.g., specific to the performance appraisal process itself or related to any of a multitude of things/events occurring in an individual’s life) it was believed that it may have an effect on what types of behaviors, and more importantly how many behaviors, would be observed while gathering information to be used in justifying a performance appraisal rating as well as triggering the recall of the behaviors while completing the actual appraisal. Also, since
performance appraisals are no longer solely tied to financial gains/losses (i.e., many different uses), it was believed that different types of appraisals might be affected differently depending on the mood of the rater. Lastly, both extraversion and emotional stability were believed to be related to positive and negative mood respectively, and therefore also have an effect on the overall performance appraisal quality. From this, organizations would be able to predict what types of rating errors might occur depending on a rater's mood state and/or personality trait. Specific types of training could then be implemented to explain how mood could affect performance appraisal quality and what the rater's should be aware of while observing and subsequently recalling these behaviors. Furthermore, the concept of keeping a journal of an employee's behaviors could be implemented in order to rely less on a rater's memory which may be tainted by a specific mood state. This in turn would lead to a more accurate rating, and therefore more accurate compensation, training, and developmental feedback, being obtained despite the negative feelings so many employees experience when dealing with performance appraisals.

The main criticism regarding external validity is that the study was performed in a laboratory setting (Tubbs, 1986). The participants (i.e., undergraduates) may not have been equivalent to supervisors on the job and the hypothetical ratee may not have accurately portrayed an employee in an organization. Furthermore, although participants were told that their responses would be used in the actual rating of the professor, they may not have believed that they would truly have a say, or at least not a significant say, in the final rating of the professor. Also, the participants could not "put a face" on the
professor, although most assumed the professor was male, therefore they did not know who they were rating which is often not the case in an organization. Likewise, the participants were not held accountable for their ratings. Specifically, the raters knew that either a very high or very low rating would not be questioned later on which is often not the case in an organization. Lastly, there was a finite number of behaviors that the raters could observe. There may have not been a high fidelity, therefore, between the laboratory and intended field in this study which could threaten external validity (Ilgen, 1986).

The laboratory setting was used in an attempt to possess maximum control over any confounding variables that would have been difficult to control outside of the laboratory (Ilgen, 1986). Also, due to the studies’ exploratory nature, the laboratory appeared to be a better starting point in establishing internal validity and was therefore used for this study.

**Future Work**

The most important issue regarding future work is to lengthen the observation period. Specifically, more than one observation session is necessary. The first benefit of extending the observation period is that it would increase the number of behaviors available for the rater to view. Therefore, if there was indeed a difference between the number of behaviors observed based upon the mood of the rater, this difference could be detected (i.e., increase of statistical conclusion validity).

Another benefit of increasing the observation session is that the frequency of certain behaviors could be added as another experimental variable. Specifically, patterns
of behaviors could be examined to see which behaviors were observed in subsequent sessions as well as which behaviors were recalled. For example, if a rater were to view certain negative (or positive) behaviors in one session, would these same behaviors be searched for in the next session(s), regardless of mood? Furthermore, would more behaviors be observed in subsequent sessions in an attempt to justify the rater's initial impression of the rater (i.e., almost forming a liking/disliking of the ratee)? With respect to the recall session, would patterns of behaviors be used more often than single behavioral instances in justifying the ratings? Finally, would there be a recency effect when ratings were given?

A third benefit of extending the observation session is that it would closer resemble an actual appraisal process. Specifically, the rater would have multiple opportunities over an extended period of time in which to base their ratings. This in turn would allow the rater to determine whether there were patterns of behaviors the ratee performed versus simple behavioral instances that were not that important when occurring only once. This might allow for larger effect sizes (i.e., statistical conclusion validity) as well as allow the study to be more generalizable to an actual organization (i.e., external validity). Of course the drawback, as was mentioned earlier, would be the question of whether it is still mood after several observational sessions that was affecting the number of behaviors observed and subsequently recalled or if it was now temperament. The concept of when mood becomes temperament could be examined by including these additional observation sessions as well.
Another suggestion for future work would be to add a second ratee to the rating process. This would provide an opportunity to determine whether any of the effects of mood would be strengthened by the behaviors of one of the ratees. For example, if a rater were in a negative mood and observed negative behaviors by a ratee (along with the positive behaviors as would be expected in a controlled process), would the rater's mood become even more negative and therefore affect how the behaviors by a subsequent ratee would be viewed. Similar results might also be expected in the recall session.

Specifically, in the current study if the rater was primarily in a positive mood during time of observing behaviors and during time of recalling behaviors, it was hypothesized that this rater would recall more positive behaviors because these were the behaviors that were more prevalent to them. However, if a second ratee was added, would the rating of the second employee be affected by the rater's positive mood (i.e., carryover effect) regardless of the actual behaviors observed for this second ratee.

A third suggestion would be to have the ratee possess more ambiguous qualities. A substantial amount of the work dealing with the effects of affect on ratings quality has focused on how ambiguous individuals are rated (Isen, 1984; Isen & Daubman, 1984; Kraiger, Billings, & Isen, 1989). Specifically, the question of will a positive individual include an ambiguous individual in their positive ratings is often asked. The inclusion of an ambiguous individual may allow for stronger effects to possible. Furthermore, the automatic versus controlled process concept can be examined more closely with respect to the ambiguous individual.
The final suggestion is to improve several of the measures in an attempt to increase construct validity. Specifically, a better measure of automatic versus controlled process is needed so that it can be determined if raters in a positive or negative mood do observe and subsequently recall behaviors in this fashion. The purpose of the appraisal needs to be strengthened as well to ensure that the raters understand the difference between administrative and developmental appraisals as well as understand that the two purposes need to be separated. Furthermore, a manipulation check that examines the purpose of the appraisal process needs to be included to see how this manipulation is working. Lastly, a more care may be needed in determining what the ideal performance dimensions on the appraisal form may be. This ensure that the rating dimensions are indeed directly tied to specific job-relevant performance dimensions that are important to the job. Furthermore, this may also increase the potential for variability on the rating form that in turn would allow for stronger differences to be detected.

Nonetheless, this was a good starting point for research on mood and performance appraisal quality. It offers some suggestions, albeit weak, that mood at both time of observing and time of recalling behaviors may have some effect on performance appraisal quality. Further research is now needed that will take the findings and suggestions of this study into account so that the quality of performance appraisals within organizations can be improved.


APPENDIX A

EXPERIMENTAL MATERIALS
Survey 1

There are 15 statements in this questionnaire. Read each statement carefully. For each statement write down the number of the response that best represents your usual or general behavior. Make sure your answer is on the correct line. Using the scale below, please respond to each statement as accurately as possible.

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>strongly disagree</td>
<td>neutral</td>
<td>strongly agree</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. ____ I may change from happy to sad and back again several times in a single week.
2. ____ I usually feel quite cheerful.
3. ____ I’m frequently “down in the dumps”.
4. ____ I generally look at the sunny side of life.
5. ____ Compared to my friends, I’m less up and down in my mood states.
6. ____ I’m not often really elated.
7. ____ Sometimes my moods swing back and forth very rapidly.
8. ____ I usually feel as though I’m bubbling over with joy.
9. ____ My moods are quite consistent; they almost never vary.
10. ____ I consider myself a happy person.
11. ____ Compared to my friends, I think less positively about life in general.
12. ____ I’m a very changeable person.
13. ____ I am not as cheerful a most people.
14. ____ I’m not as “moody” as most people I know.
15. ____ My friends often seem to feel I am happy.
How Accurately Can You Describe Yourself?

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly your same age.

Before each trait, please write a number indicating how accurately that trait describes you, using the following scale:

<table>
<thead>
<tr>
<th>Inaccurate</th>
<th>Accurate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely</td>
<td>Slightly</td>
</tr>
<tr>
<td>Very</td>
<td>Moderately</td>
</tr>
<tr>
<td>Moderately</td>
<td>Very</td>
</tr>
<tr>
<td>Slightly</td>
<td>Extremely</td>
</tr>
</tbody>
</table>

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

- Bashful
- Bold
- Careless
- Cold
- Complex
- Cooperative
- Creative
- Deep
- Disorganized
- Efficient
- Energetic
- Envious
- Extraverted
- Fretful
- Harsh
- Imaginative
- Inefficient
- Intellectual
- Jealous
- Kind

- Moody
- Organized
- Philosophical
- Practical
- Quiet
- Relaxed
- Rude
- Shy
- Sloppy
- Sympathetic
- Systematic
- Talkative
- Temperamental
- Touchy
- Uncreative
- Unenvious
- Unintellectual
- Unsympathetic
- Warm
- Withdrawn
Imagine feeling the following statements:

"Sometimes I can’t wait to get out of bed in the morning!"

"This is great! I really do feel good! I am thrilled about things!"

Now, please think of a recent event (or events) in your life that made you feel this way. In a sentence or two summarize the positive event(s) below. Also please summarize the specific positive feelings that occurred as a result of the positive event(s). Once you are finished writing, please continue to think about the event(s) and feelings associated with the event. Feel free to take a minute or two to think about the event before writing.

Positive Event:

Positive Feelings:

Please continue to think about the statements above throughout the rest of the session.
Imagine feeling the following statements:

"Every now and then I feel so tired and gloomy that I'd rather sit than do anything."

"I have too many bad things in my life."

Now, please think of a recent event (or events) in your life that made you feel this way. In a sentence or two summarize the negative event(s) below. Also please summarize the specific negative feelings that occurred as a result of the negative event(s). Once you are finished writing, please continue to think about the event(s) and feelings associated with the event. Feel free to take a minute or two to think about the event before writing.

**Negative Event:**

**Negative Feelings:**

Please continue to think about the statements above throughout the rest of the session.
Think of your most recent psychology class. Please write down several things you remember about your most recent psychology class. Please focus only on the psychological material that was presented to you.
The PANAS

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now. Use the following scale to record your answers.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>very slightly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or not at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a little</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>moderately</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quite a bit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>extremely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. pleasant       11. irritable
2. distressed      12. cheerful
3. excited         13. ashamed
4. upset           14. inspired
5. strong          15. nervous
6. guilty          16. determined
7. scared          17. joyful
8. hostile         18. jittery
9. enthusiastic     19. active
10. proud          20. Afraid
The professor has been at OSU for several years. The professor is currently teaching 1 undergraduate upper level course as well as a graduate level seminar. The professor has a fairly broad range of academic interests and is currently involved in several research projects.
POSITIVE BEHAVIORS

A student approaches the professor and explains that they are unable to meet with the professor during any of the scheduled office hours because they work during the day. The professor tells the student that he/she could come in over the weekend if that would be more manageable for the student.

The professor ends each class by asking if there are any questions/concerns about what was discussed in that day’s lecture. The professor states that he/she wants to provide this time at the end of class to allow the students to ask questions and think for themselves rather than tell them how to think.

The professor states that he/she is readily available to answer any questions the students may have regarding the class. The professor schedules adequate office hours as well as makes him/herself available upon request. The professor also allows the students to contact him/her via e-mail or phone for simple questions.

The professor is nominated for an annual award as an outstanding faculty member. The professor is recognized for his/her outstanding teaching, advising, and service.

A student approaches the professor and explains that he/she could not attend the last class due to illness. The professor states that if the student would like, the professor could photocopy his/her lecture notes from the previous class for the student. The professor also states that he/she could briefly discuss several of the major points from the previous class with the student after class.

The professor uses examples in his/her lecture that are current and apply to the real world. He/she relates the lecture material to events that are either taking place on campus or around Columbus. The professor states that he/she wants to discuss events that may have an impact on the students’ lives.

The professor is busy working on a research project when a student arrives during office hours. The professor promptly stops what he/she is working on and devotes his/her complete attention to the student’s concerns.

The professor has three journal articles accepted where he/she is the first author, and two journal articles accepted where he/she is the second author. All five of the articles are accepted in major journals in the professor’s field.

The professor works hard to ensure that all of his/her graduate advisees have adequate funding. The professor contacts area businesses in an attempt to locate intern opportunities as well as establish working relationships between area businesses and the university.
The professor states clearly what is expected of each student at the beginning of the quarter (i.e., material to be covered, how the students will be evaluated). The professor then addresses any questions regarding the course so that each student will understand what is expected of them.

The professor is always well prepared before each class. The professor arrives before class with organized notes, writes an outline for the class on the chalkboard, and is ready to begin teaching as soon as the bell rings.

A student arrives at the professor's office hours while the professor is meeting with another student. The professor asks if the student would like to come in and take part in the discussion. The professor states that both students could benefit from each other's questions.

A student is meeting with the professor during office hours. The scheduled hours have ended but the student still has several unanswered questions. The professor states that he/she is willing to stay with the student until all of their questions are answered.

The professor is awarded a grant by the U.S. Army that will fund a research project over the next three years. The grant will provide enough money to fund the project as well as two of the professor's research assistants over those three years.

After the first exam, several students tell the professor that they studied for the exam, but they still did poorly. The professor schedules extra office hours for students who are doing poorly in the class. The professor states that these extra sessions have been set-up so that each student has an opportunity to understand all of the course material.

The professor begins each class by asking the students if they have any examples they would like to discuss. The professor states that he/she would rather use examples that are relevant to the students than use examples that are of interest to him/her.

A student approaches the professor and asks if the professor could offer any advice regarding graduate school. The professor states that the student is welcome to meet with him/her at any time regarding questions/concerns they may have other than course material (e.g., graduate school, research opportunities, different areas in psychology).

A student approaches the professor and asks if the professor would be willing to serve as a faculty advisor for a university group. The professor states that he/she has served as a faculty advisor on several occasions during his/her tenure at the university, and that he/she will seriously consider being a faculty advisor for this student's group.
The professor announces that he/she will be out of town during the next scheduled office hours. The professor then allows the students to decide when they would like “make-up” office hours to be held. The professor offers “make-up” office hours both before as well as after the normally scheduled hours so that all students will have adequate time to have any questions answered.

A student asks a question in class. The professor offers an alternative explanation for the concept in question. The professor then asks if the student, as well as several other students, can explain the concept that is confusing. The professor states that he/she does this to ensure that the students are understanding what is being discussed in class.

The professor offers detailed review sheets before each exam. These sheets contain the major points that were discussed in class as well as the common problems/questions that students have had in the past.

The professor states on the first day of class that he/she realizes how boring it can be for the students to listen to one individual for an entire class period so he/she is going to offer the students the opportunity to dictate what is discussed in each class. The professor states that each student will be given an equal opportunity to speak during class.

A student approaches the professor and asks if the professor knows anything about senior honors theses. The professor states that he/she is available for undergraduate students who may be interested in writing an honors thesis. The professor also states that he/she has served as an advisor for several students and states that he/she has genuinely enjoyed helping these students learn more through their research projects.

The professor is selected to be the editor of one of the best journals in his/her field. The professor will be responsible for having the final say in what does/does not get published in the journal. The professor will also be responsible for maintaining the journal’s high standards.

(Please do not read any more of the behaviors in this packet. You should have read an adequate number of positive behaviors by now. Please do not turn over this page.)
NEGATIVE BEHAVIORS

The professor does not attend any of the exams for the class, but rather sends a proctor. Several of the students state that the proctor that the professor sends cannot always answer the questions they have about the exam. The professor tells the students that if they have a question, it is probably because they don't know the material. The professor also tells the students that it is a waste of his/her time to attend the exam.

A student is meeting with the professor during office hours when a colleague of the professor comes in to ask a general question about a research project. The professor tells the student that he/she will have to come back some other time and begins talking to the colleague.

A student sees the professor in the hall and asks if he/she could ask the professor a quick question. The professor responds that the student had ample time to ask the question during office hours and tells the student leave him/her alone.

On several occasions the professor has begun asking personal questions of the students who come to see him/her during office hours. The professor will address several of the students' concerns and will then try to gather information about the students' personal lives.

A student approaches the professor and asks if the professor knows of any faculty members who may be in need of research assistants. The professor responds that finding research outlets for undergraduates is not his/her responsibility and tells the student to go ask someone else.

A student misses a scheduled appointment with the professor. The student apologizes for missing the meeting and tells the professor that he/she would like to set-up another meeting to go over several questions/concerns. The professor tells the student that he/she should not bother asking for help anymore because the student obviously doesn't care about his/her education.

A student is leaving the professor's office hours when another professor arrives. While the student is gathering his/her material, the professor proceeds to tell the new professor how poorly the student is doing in the class. The professor also mentions some of the "easy" topics that the student was having difficulty understanding.

A student asks a question during class to which the professor responds: "That is something you really should know by now. Haven't you been paying attention at all this quarter?"; and continues on with the lecture.
A student arrives late for one of the professor's lectures and quietly takes a seat near the back of the room. The professor asks the student to stand up and state why he/she was late for the class. The professor also asks the student to remain standing and summarize all of the reading that was to be completed for that class.

The professor usually takes at least two weeks to give the students feedback on their exams/quizzes. The professor states that he/she has more important things to do than grade exams and that the students will have to wait.

The professor is participating in an intramural sport. The professor notices that one of his/her students is also participating in the sport. The professor mentions to the student that his/her performance could influence the grade they receive in the class.

A student arrives at the professor's office and notices that the professor has several jokes on his/her wall that contain very offensive language. The student mentions to the professor that some individuals could find the jokes offensive to which the professor responds, "I really don't care what you think, nor do I care if people find those jokes offensive. I think the jokes are funny!"

The professor has shown up to several of the faculty/student mixers intoxicated. The professor is often loud and has caused minor scenes on more than one occasion.

A student approaches the professor and asks if the professor could look over a research topic he/she is interested in pursuing. After reading the proposal, the professor tells the student that he/she has done a terrible job with the literature review. The professor further states that because of this, the research will be a waste of time.

After the first exam, the professor tends to only make eye-contact with those students who are doing well in the class. He/she fields their questions first, and only then will acknowledge the remaining students.

The professor works his/her research assistants very hard in order to publish as much as possible. He/she states that someday the assistants might appreciate what he/she has done for them, but for now, they should shut-up and do the work.

A student approaches a professor after class to set-up a meeting. The professor responds that he/she does not have any time and that the student will simply have to wait.

The professor will often poke fun of or point out trivial mistakes with other faculty's research at the university. The professor states that he/she is trying to provide examples of what poor research looks like. The professor also states that the students should be allowed to know who does the worst research at the university.
The professor gets behind in his/her lecture notes. The professor ends up speaking very quickly through the rest of the lecture in an attempt to cover all of the intended material.

A student meets with the professor to go over the test. The student believes that one of the questions was worded a little awkwardly and that was probably why they responded incorrectly. The professor responds that the test was worded fine and the student should read the questions better next time.

A student arrives at scheduled office hours just as they are ending. The professor states that the student should have come earlier, and that office hours are now over. The professor then proceeds to read the newspaper.

The professor often uses language in the classroom that is politically incorrect (i.e., sexist and sometimes racist).

A student approaches the professor and explains that he/she is writing a research report for another class. The student then asks the professor if the professor has a book they need for the report. The professor states that he/she is a professor, not a library. The professor also states that he/she doesn’t trust students enough to lend them books.

A student tells the professor that he/she will not be able to take an exam during the scheduled time because of an athletic conflict. The professor tells the student that it is wrong that student athletes should be given special privileges. The professor then tells the student that he/she is going to make the make-up exam much more difficult than the exam that will be given to the rest of the class.

(Please do not read any more of the behaviors in this packet. You should have read an adequate number of negative behaviors by now. Please do not turn over this page).
Think of the positive event that you wrote down last session. Please add any additional thoughts about the event below. If nothing has changed with respect to the event that you wrote down last time, simply summarize the event and respective feelings again.

Once again, the statements were:

“Sometimes I can’t wait to get out of bed in the morning!”

“This is great! I really do feel good! I am thrilled about things!”

**Positive Event:**

**Positive Feelings:**

Please continue to think about the statements above throughout the rest of the session.
Think of the negative event that you wrote down last session. Please add any additional thoughts about the event below. If nothing has changed with respect to the event that you wrote down last time, simply summarize the event and respective feelings again.

Once again, the statements were:

"Every now and then I feel so tired and gloomy that I’d rather sit than do anything."

"I have too many bad things in my life."

Negative Event:

Negative Feelings:

Please continue to think about the statements above throughout the rest of the session.
Once again, please think of your most recent psychology class. Please write down several things you remember about your most recent psychology class. Please focus only on the psychological material that was presented to you.
The PANAS

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way right now. Use the following scale to record your answers.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>

139
Performance Appraisal Form

You will now complete a developmental performance appraisal (e.g., helping the employee understand what their strong areas are, and what areas they need improvement in) for the college professor from the previous session. You will be given a rating scale that ranges from 1 (poor) to 5 (good). Please consider the entire range before making your final rating. Some problems in making ratings that you should consider and avoid are: severity: you always use the low end of the scale, regardless of the performance quality; leniency: you always use the high end of the scale, regardless of performance quality; central tendency: you always use the middle of the scale, regardless of performance quality; halo: you make several ratings based on one characteristic (e.g., appearance, gender). Once again, you are encouraged to consider the entire range of the scale and base your final judgment on all of the appraisee’s behaviors. Also, before making a rating, please write down comments to justify the rating you give. You will be rating the professor on three specific factors (communication skills, research skills, & organizational skills) as well as providing an overall rating. Please use the following scale to rate the professor:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>POOR</td>
<td>ADEQUATE</td>
<td>GOOD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Performance is repeatedly below expectations. The employee should be counseled regarding specific steps to be taken to evidence improvement.

2. Performance occasionally below expectations. The employee should be counseled concerning the specific steps to be taken to meet expectations.

3. Performance consistently meets expectations. The employee should be informed of ways in which performance could exceed expectations.

4. Performance occasionally above expectations. The employee should be informed of the strengths and accomplishments observed and ways to continue to improve upon them.

5. Performance repeatedly well above expectations. The employee should be informed of the strengths and accomplishments which have been observed as well as potential for future growth with the organization.
You will now complete an administrative performance appraisal (e.g., promotions, pay raises, or terminations) for the college professor from the previous session. You will be given a rating scale that ranges from 1 (poor) to 5 (good). Please consider the entire range before making your final rating. Some problems in making ratings that you should consider and avoid are: severity: you always use the low end of the scale, regardless of the performance quality; leniency: you always use the high end of the scale, regardless of performance quality; central tendency: you always use the middle of the scale, regardless of performance quality; halo: you make several ratings based on one characteristic (e.g., appearance, gender). Once again, you are encouraged to consider the entire range of the scale and base your final judgment on all of the appraisee’s behaviors. Also, before making a rating, please write down comments to justify the rating you give. You will be rating the professor on three specific factors (communication skills, research skills, & organizational skills) as well as providing an overall rating. Please use the following scale to rate the professor:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>POOR</td>
<td>ADEQUATE</td>
<td>GOOD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Performance is repeatedly below expectations. The employee should be strongly considered for a demotion and a pay decrease.
2. Performance occasionally below expectations. The employee may be considered for a minor demotion or a pay decrease.
3. Performance consistently meets expectations. The employee should receive a “cost of living” pay raise.
4. Performance occasionally above expectations. The employee may be considered for a minor promotion or a pay raise.
5. Performance repeatedly well above expectations. The employee should be strongly considered for a promotion and a pay raise.
1. Communication Skills

Performance criteria:

- Communicates in a logical and clear manner.
- Expresses self effectively when presenting ideas or instructions.

Please write as few/many comments as you like to justify your rating.

Comments:

Rating:_____
2. **Research Skills**

Performance criteria:

- Generates effective research ideas and solutions.
- Keeps informed on up-to-date research ideas and techniques.
- Encourages new researchers.

Please write as few/many comments as you like to justify your rating.

**Comments:**

Rating: ___
3. Organizational Skills

Performance criteria:

• Prioritizes to adapt to varying situations, responsibilities and people.
• Identifies and applies available information and resources.
• Produces results within sufficient time limits.

Please write as few/many comments as you like to justify your rating.

Comments:

Rating: _____
4. Overall performance

Please write as few/many comments as you like to justify your rating.

Comments:

Rating:____
Retrieval Task

Please circle the behaviors listed below that were presented to you in the initial experimental session (i.e., which behaviors did you read in the first session). Please do not look at your notes or your ratings. I realize that you may not have viewed all of the behaviors in a packet; we are only asking you to circle those you remember viewing.

1. A student sees the professor in the hall and asks if he/she could ask the professor a quick question. The professor responds that the student had ample time to ask the question during office hours and tells the student to leave him/her alone.

2. A student approaches the professor and asks if the professor knows of any faculty members who may be in need of research assistants. The professor responds that finding research outlets for undergraduates is not his/her responsibility and tells the student to go ask someone else.

3. The professor is often late to class and is usually quite unorganized. On several occasions, the professor has blamed his/her tardiness on the class and has stated that his/her life would be much easier if he/she didn’t have to worry about teaching this stupid course.

4. The professor does not attend any of the exams for the class, but rather sends a proctor. Several of the students state that the proctor that the professor sends cannot always answer the questions they have about the exam. The professor tells the students that if they have a question, it is probably because they don’t know the material. The professor also tells the students that it is a waste of his/he time to attend the exam.

5. The professor never shows up to his/her office hours. He/she states that he/she only posts office hours because the university forces him/her to. He/she also states that there are far better things that he/she could be doing than answering the tedious questions of students.

6. The professor will often offer sample questions to his/her students before an exam. The professor states that he/she wants the students to know what to expect on each exam and have an opportunity to clear up possible confusions before it is too late.

7. The professor states that he/she is readily available to answer any questions the students may have regarding the class. The professor schedules adequate office hours as well as makes him/herself available upon request. The professor also allows the students to contact him/her via e-mail or phone for simple questions.

8. A graduate advisee approaches the professor and states that they do not have adequate funds to perform a research project that they are interested in. The professor states that he/she does not have any grant money available for them, but offers to help out with money out of his/her own pocket.

9. The professor is heavily recruited by several other universities who would like to offer the professor the position of chairperson of their department.

10. The professor is busy working on a research project when a student arrives during office hours. The professor promptly stops what he/she is working on and devotes his/her complete attention to the student’s concerns.
Task Methods Perception Questionnaire

Please circle a response to each of the 21 questions below.

1. Gender:  Male  Female

2. Have you ever appraised someone else's performance in a work setting?
   Yes  No

3. Have you ever had someone else appraise your performance in a work setting?
   Yes  No

4. How would you describe your mood during the first session (i.e., the session where you read the performance behaviors)?
   1  2  3  4  5  6  7
   Negative  Neutral  Positive

5. To what extent was something in the experiment the cause of your mood during the first session?
   1  2  3  4  5  6  7
   Very little  Very Much

6. How would you describe your mood during the second session (i.e., the session where you completed the performance appraisal)?
   1  2  3  4  5  6  7
   Negative  Neutral  Positive

7. To what extent was something in the experiment the cause of your mood during the second session?
   1  2  3  4  5  6  7
   Very little  Very Much

8. To what extent was your mood different between the first and second sessions?
   1  2  3  4  5  6  7
   Very little  Very Much
9. How "positive" did you find the positive behaviors from the first session?

   1  2  3  4  5  6  7
Very little   Very Much

10. How "negative" did you find the negative behaviors from the first session?

   1  2  3  4  5  6  7
Very little   Very Much

11. To what extent were you provided an adequate number of positive behaviors in the first session?

   1  2  3  4  5  6  7
Very little   Very Much

12. To what extent were you provided an adequate number of negative behaviors in the first session?

   1  2  3  4  5  6  7
Very little   Very Much

13. To what extent did the performance appraisal form allow you to make an adequate rating of the professor (i.e., were all the necessary skills covered in the rating form)?

   1  2  3  4  5  6  7
Very little   Very Much

14. For the communication skills did you:

   A. First give a rating and then justify the rating with comments.
   B. First write down comments and then make a rating based upon the comments.

15. For the research skills did you:

   A. First give a rating and then justify the rating with comments.
   B. First write down comments and then make a rating based upon the comments.

16. For the organizational skills did you:

   A. First give a rating and then justify the rating with comments.
   B. First write down comments and then make a rating based upon the comments.
17. For the overall rating did you:
   A. First give a rating and then justify the rating with comments.
   B. First write down comments and then make a rating based upon the comments.

18. To what extent did you use the notes from the first session to complete the performance appraisal?
   1 2 3 4 5 6 7
   Very little  Very Much

19. To what extent did you use your memory to complete the performance appraisal?
   1 2 3 4 5 6 7
   Very little  Very Much

20. To what extent were your answers to this questionnaire sincere?
   1 2 3 4 5 6 7
   Very Insincere  Very Sincere

21. What gender did you believe the professor to be?
   Male    Female
Rating quality is typically a concern of anyone involved in performance appraisals. Employees, managers, and researchers are all curious about rating quality, especially if objective performance criteria are not available. It is unreasonable to believe that raters will gather and eventually retrieve information on a ratee in a vacuum. Rather, mood should play some role in the gathering and retrieval of information whether the performance criteria are objective or subjective.

The experiment you just took part in was designed to study the effect of affective state (i.e., mood) on the performance appraisal process. Specifically, how does mood of the rater affect the rater’s encoding (and eventual retrieval) process of performance criteria.

The point of the this study is that the “fairness” of performance appraisals is questioned nearly every day by an employee or an employer. If any relationship between mood and the performance appraisal process is discovered, then further research can be performed in an attempt to ensure higher rating quality.

If you wish to read additional references on this topic, one place to start is a book published in 1995 entitled “Understanding performance appraisal: Social, organizational, and goal-based perspectives” by Kevin Murphy and Jeanette Cleveland.

THANK YOU VERY MUCH FOR YOUR HELP

Contact persons for further information:

Robert Calderón. 292-8175. Room 128 Lazenby Hall
James T. Austin. 292-0685. Room 140f Lazenby Hall