



# LIBRARY Michigan State University

This is to certify that the

dissertation entitled

THE INFLUENCE OF WORK VERSUS NON-WORK RESPONSES ON RATINGS OF EXPERIENCE-BASED STRUCTURED INTERVIEWS

presented by

Kevin E. Plamondon

has been accepted towards fulfillment of the requirements for

PH.D. degree in Philosophy

Fack Chaut

Major professor

Date 8/15/02

MSU is an Affirmative Action/Equal Opportunity Institution

0-12771

# PLACE IN RETURN BOX to remove this checkout from your record. TO AVOID FINES return on or before date due. MAY BE RECALLED with earlier due date if requested.

DATE DUE	Γ	DATE DUE		DATE DUE
	+			_
	+		+	
			$\downarrow$	
	T			
	+		T	
	$\downarrow$		+	
	-+			
	-+			
				11/00 c/CIRC/DateDue.p85-p

# THE INFLUENCE OF WORK VERSUS NON-WORK RESPONSES ON RATINGS OF EXPERIENCE-BASED STRUCTURED INTERVIEWS

By

Kevin E. Plamondon

#### A DISSERTATION

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

### DOCTOR OF PHILOSOPHY

Department of Psychology

#### ABSTRACT

#### THE INFLUENCE OF WORK VERSUS NON-WORK RESPONSES ON RATINGS OF EXPERIENCE-BASED STRUCTURED INTERVIEWS

by

#### Kevin E. Plamondon

This study examined whether trained raters are biased towards interview responses that occur in a work setting using laboratory data collected from 258 college students. It was hypothesized that 1) responses occurring in a work setting would be rated more positively than those occurring in a non-work setting, 2) responses occurring in a non-work setting would invoke greater halo error (as indicated by lower variance in interview ratings), 3) rater bias towards work responses would influence ideal candidate prototypes, and ideal candidate prototypes would moderate the relationship between responses and interview ratings such that 4) work responses would be rated more positively and 5) yield greater halo error as ideal candidate prototypes favored work responses. Data were analyzed using repeated measures analysis of covariance and multiple regression. Results supported hypothesis 1 and partially supported hypothesis 2; there was evidence that raters favored work responses and that non-work responses had an impact on halo error. Hypotheses 3, 4, and 5 were not supported. The implications of these findings and explanations for the non-significant results are discussed. Copyright by KEVIN E. PLAMONDON 2000

.

## TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	'ii
INTRODUCTION	. 1
Interview History: The Early Years	.3
Schmitt's 1976 Review	4
Arvey and Campion's 1982 Review	5
Structured Interviews and Meta-Analysis	.5
Interview Research in the 1990's	.7
The Influence of Work versus Non-work Responses	.9
Information Use	10
Interviewer Stereotypes	1
Feldman's Model	2
Fiske's Model of Stereotyping	4
Implications of Cognitive Processing for the Structured Interview	16
Interview Validity and Applicant Experience	
Interview validity and Applicant Experience	. /
METHODS	23
Participants2	23
Development of the Interview Materials2	25
Design2	26
Procedure2	26
Training2	27
Stimulus Materials	28
Interview Reponses	28
The Job	28
Candidate Experience2	28
Measures2	29
Interview ratings	29
Halo error	29
Hiring decision	30
Adjective checklist	30
Rater bias measure	31
Interview experience	31
Resume manipulation check	31
Practice ratings	32

Analyses	
RESULTS	
Manipulation Check	
Confederate & Order Effects	
Confederate & order effects for interview ratings	
Confederate & order effects for halo error	
Test of Hypotheses	
Hypothesis 1	52
Hiring Decision	55
Hypothesis 2	57
Hypothesis 3	60
Hypothesis 4	64
Hypothesis 5	67
DISCUSSION	70
Discussion of Hypotheses	70
Hunothesis 1	
Hypothesis 2	
Hypothesis 2 Hypothesis 3	74
Hypotheses 4 and 5	
Study Limitations	
College student sample	
Laboratory setting and study design	
Effect size	
Base rate	80
Validity	80
Future Research	
Conclusion	
REFERENCES	
ADDENIDIY A: Dower Analysis	02
AFFENDIX A. Fower Analysis	
APPENDIX B: Training Manual	94
APPENDIX C: Video Transcripts	102
APPENDIX D: Resumes	114
APPENDIX E: Measures	117
APPENDIX F: Informed Consent & Debriefing Forms	128

# LIST OF TABLES

Table 1. Summary of Campion et al.'s (1997) findings	8
Table 2. Study Conditions	24
Table 3. Descriptive statistics and scale intercorrelations	35
Table 4. Dimension Intercorrelations for Interview 1	36
Table 5. Dimension Intercorrelations for Interview 2	
Table 6. Observed means of interview ratings	38
Table 7. Observed means of rating standard deviations	39
Table 8. Tests of Within-Subjects Effects for Interview Ratings	42
Table 9. Tests of Between-Subjects Effects for Interview Ratings	45
Table 10. Test of within subject effects for halo error	47
Table 11. Between subject effects for halo error	47
Table 12. Logistic regression results for hiring decision.	56
Table 13. Regression results for Hypothesis 3 mediated model	62
Table 14. Regression results for moderated model using rater bias	63
Table 15. Regression results for rater bias measure	65
Table 16. Test of moderated model for adjective checklist ratings and interview ra	tings 68
Table 17. Test of moderated model for adjective checklist ratings and halo error	69

.

## LIST OF FIGURES

Figure 1. Hypothesized model of rater stereotypes	17
Figure 2. Graph of hypotheses	22
Figure 3. Confederate by order interaction for interview ratings	43
Figure 4. Confederate by response setting interaction for interview ratings	46
Figure 5. Confederate by order interaction for halo error	48
Figure 6. Three-way interaction for response settings and confederate	49
Figure 7. Three-way interaction for response settings and confederate	50
Figure 8. Response setting 1 by order interaction for interview ratings	53
Figure 9. Response setting 2 by order interaction for interview ratings	55
Figure 10. Response setting by order interaction for halo error	58
Figure 11. Response settings by order interaction for halo error	59

#### INTRODUCTION

The employment interview is a widely used assessment technique that has become a core component of many selection procedures (Eder & Harris, 1999). For the purpose of this study, the employment interview will be defined as, "an interviewerapplicant exchange of information in which the interviewer inquires into the applicant's work-related knowledge, skills, and abilities; motivation; values; and reliability with the overall staffing goals of attracting, selecting, and retaining a highly competent and productive workforce (Eder & Harris, 1999)."

Although this definition includes a wide array of interviews, this paper will focus on the structured interview and more specifically, the experience-based structured interview. The experience-based structured interview asks applicants to describe their past experiences using a standard set of interview questions. Research indicates that of all the employment interview techniques, structured interviews yield the highest reliabilities and validities (Huffcutt & Arthur, 1994). These findings are attributed to the fact that structured interviews limit the information processing load placed on interviewers, focus attention on job relevant qualifications, and reduce the influence of irrelevant information on ratings (Campion, Palmer, & Campion, 1997).

The purpose of this study is to expand upon the structured interview research by examining a potential rater bias that may be influencing experience-based interview ratings. In an experience-based interview, applicants are asked to describe past experiences that demonstrate job relevant capabilities. In responding to these types of questions, applicants may refer to experiences that occurred in a work (e.g., current job)

or non-work setting (e.g., school, volunteer organization, athletic team). Despite this potential variability in responses, no published research has examined whether raters are biased in favor of behaviors occurring in a work setting versus those occurring in a nonwork setting. This paper will examine whether response setting (work or non-work) in experience-based structured interviews influences interview ratings. Throughout this paper, the terms rater bias, response setting, and ideal candidate prototypes will be used frequently. Unless otherwise stated, rater bias in this paper refers to tendencies among raters to favor or be biased towards responses occurring in a work or non-work setting. This distinction is mentioned because the interview literature has examined a number of other rater biases that influence ratings; these biases will not be examined in this study. Response setting will also be used frequently. In this paper it will refer to whether interview responses include behaviors that occurred in a work or non-work environment or setting. Finally, the term ideal candidate prototype will be used. This term will be used as it is in the interview literature to refer to cognitive schemas held by raters of the ideal candidate for a job.

This paper will review the trends in the interview literature from 1949 to 1999 to establish the history that led to the development of the structured interview and to define the current state of the structured interview research. It will then discuss the interview research on information use and stereotyping to provide a theoretical support for how response setting (work versus non-work) could influence interview ratings. This discussion will include a review of two theories of stereotype formation –Feldman's (1981) cognitive processing model and Fiske, Neuberg, Beattie, and Milberg's (1987) impression formation model. The paper will then propose a model and the five

hypotheses tested in this study, the methods used in the experiment, as well as the results of the study. A discussion of these results, their implications, and potential limitations will conclude the paper.

#### Interview History: The Early Years

Although published studies of the interview date back to 1911, it was not until 1949 that the first comprehensive review of the employment interview research was published (Wagner, 1949). Wagner was concerned with the psychometric properties of the technique, noting moderate reliabilities (.57) and low validities (.27). He concluded that the interview is useful when a rough screening is needed, the applicant pool is too small to justify more expensive procedures, and the traits being assessed are conducive to the interview format (e.g., interpersonal skills). Wagner also recommended focusing on job related traits and using standardized interview approaches.

Several years after Wagner's article, Mayfield (1964) conducted another comprehensive literature review in which he too noted the poor reliability and validity of interviews. Despite Wagner's recommendations and fifteen years of research, there seemed to be little improvement in the psychometric properties of the interview. Mayfield concluded that researchers should partition the individual factors influencing ratings and examine the unique effects of each. He called for a micro-analytic approach that focused on the decision making process of the interview –a recommendation echoed in a review by Ulrich and Trumbo (1965). Around this time, Webster (1964) and a team of researchers were focusing their attention on specific factors influencing the decision making process of the interview.

Webster and colleagues answered many important questions regarding the cognitive processes influencing the interview (see Schmitt's 1976 review below). However, their research raised concerns that the field had become *too* specific. In a 1969 review, Wright argued for a more holistic approach that examined the interview as an integrated whole –a view that was also shared by Schmitt (1976). Both authors felt that the interview research needed an integrating framework. Schmitt's (1976) article reviewed the existing research and built a foundation for conducting more comprehensive investigations of the selection interview.

#### Schmitt's 1976 Review

Schmitt's (1976) review reflected much of Webster's (1964) influential work on information processing. Factors such as the impact of negative vs. positive information, the temporal placement of information, interviewer stereotypes, the availability of job analysis information, the influence of visual cues, interviewer-interviewee similarities, contrast effects among interviewees, and interview structure all seemed to be influencing ratings. Unfortunately, there seemed to be more evidence for the poor reliability and low validities of the interview than information on how to improve the psychometric properties of the technique. Schmitt recommended 1) structuring the interview process to improve reliability, 2) focusing the interview on job relevant information, 3) training interviewers to avoid many of the rating errors identified by previous research, 4) tailoring the interview to the desired outcome (selection vs. recruitment), and 5) using an interview only when it matched the demands of the job (e.g., to assess motivation or interpersonal skills). Years later, these recommendations would characterize many of the qualities associated with the structured interview.

#### Arvey and Campion's 1982 Review

In 1982 Arvey and Campion conducted yet another review of the interview literature. They examined research on the reliability and validity of interviews and integrated Schmitt's (1976) review with subsequent research on decision-making. Arvey and Campion's (1982) article was one of the first positive reviews of the interview literature to date. In the years since Schmitt's (1976) article, the psychometric outlook for the interview was improving. Specifically, using multiple interviewers and job analysis information improved both the reliability and validity of the interview. In addition, much of the decision-making errors commonly associated with the interview (i.e., contrast effects, primacy-recency effects, first impressions, and personal biases) could be diminished by more carefully structuring the rating task and having raters make specific predictions of job behavior. Interviewer training was also found to have positive effects on reducing rating errors and improving accuracy.

#### Structured Interviews and Meta-Analysis

In the years that followed Arvey and Campion's (1982) article, two major developments greatly influenced the interview literature and addressed many of the concerns raised by Wagner (1949), Schmitt (1976), and Arvey and Campion (1982). The developments include the introduction of structured interview techniques (Janz, Hellervick, & Gilmore, 1986; Latham, Saari, Pursell, & Campion, 1980) and the publication of three interview meta-analyses (McDaniel, Whetzel, Schmidt, & Maurer, 1994; Wiesner & Cronshaw, 1988; Wright, Litchtenfels, & Pursell, 1989).

During the early and mid-1980's two types of structured interviews emerged –the behavioral description interview and the situational interview. Both techniques are

similar in that they have structured procedures and are developed using job analysis information. The crucial difference between the two approaches is the types of questions asked. The behavior description interview asks applicants to describe past behavior and is based on the belief that past behavior is the best predictor of future performance (Janz et al., 1986). A sample question for a bank teller position would be, "Balancing the cash bag is always bottom line for a cashier position, but bags can't always balance. Tell me about the time your experience helped you discover why your bag didn't balance" (Janz et al., 1986). Studies of the behavior description interview found moderate uncorrected validity coefficients (.48 to .54; Janz, 1982; 1989; Orpen, 1985). Although the studies used small samples sizes (n<20) and the ratings often yielded low reliabilities (.46; Janz, 1989; 1982; Orpen, 1985), the initial results were positive.

Unlike the behavioral description interview, which focuses on past behavior, the situational interview focuses on future behavior (Latham et al., 1980). This approach is based on goal setting theory. The belief is that intentions are linked to goals and goals lead to behavior. Applicants are presented with a hypothetical, job relevant situation and asked to describe what they *would* do in the situation. A sample question is "Your spouse and two teen-age children are sick in bed with a cold. There are no relatives or friends available to look in on them. Your shift starts in 3 hours. What would you do?" (Latham et al., 1980). Responses to these questions are scored using a highly structured scoring guide. A series of studies using situational interviews found them to have acceptable reliabilities (.67 to .84) and moderate validities (.30 to .46; Latham et al., 1980; Latham & Saari, 1984; Weekley & Gier, 1987).

These data supported the beneficial effects of structuring the interview process – counteracting nearly 50 years of criticism and concern. The findings were crucial for establishing the interview as an effective selection technique. Equally as influential was the publication of several meta-analyses of the interview research. Wiesner and Cronshaw (1988), Wright et al. (1989), and McDaniel et al. (1994) found the interview to be a valid technique regardless of the situation and found that structured interviews consistently yield higher reliabilities and validities than unstructured interviews. McDaniel et al.'s (1994) study reported an average interrater reliability of .84 for job related, structured interviews. Likewise, the average, uncorrected interview validity reported by each of the three meta-analyses was in the mid 20's to low 30's; when corrected, validity reached as high as .63 for structured interviews (Wiesner & Cronshaw, 1988). The development of structured interview techniques and the publication of interview meta-analyses supported the reliability and validity of the interview and had a significant impact on establishing the interview as a viable assessment option.

#### Interview Research in the 1990's

The interview literature in the 1990's saw the emergence of two lines of research. One focused on interviewer-applicant interactions, examining issues such as impression management and applicant reactions. As these issues are less relevant for the current study, they will not be addressed. The other line of research focused on better ways to structure the interview process (Eder & Harris, 1999). Campion et al. (1997) published one of the more comprehensive reviews in this area. Concerned that the term "structure" was not adequately defined, they conducted an extensive review of various structuring techniques.

The review examined the influence of 1) using job analysis information, 2) asking the same set of questions, 3) limiting prompting questions, 4) question format, 5) interview length, 6) controlling ancillary information, 7) allowing candidates to ask questions, 8) rating each answer separately, 9) using anchored rating scales, 10) taking notes, 11) having multiple interviewers, 12) using the same interviewer, 13) controlling discussion among interviewers, 14) training interviewers, and 15) statistically combining interviewer ratings. Table 1 lists the effects of each factor on reliability and validity. As a whole, these various structuring techniques tend to focus attention on job relevant factors, limit information processing load, and reduce random variability thus minimizing rating errors and improving reliability and validity.

Content	Reliability	Validity
Job analysis		+
Same questions	+	+
Limiting prompting	+	0
Question format	+	+
Longer interview	+	+
Control ancillary information	+	0
Limiting candidate questions	+	0
Evaluation	Reliability	Validity
Rate each answer/Multiple scales	+	+
Anchored rating scales	+	
Detailed notes	+	+
Multiple interviewers	+	0
Use the same interviewers	+	0
Control interviewer discussion		
Training	+	+
Statistical weighting of information	+	+

 Table 1. Summary of Campion et al.'s (1997) findings

+ = positive effect, 0 = mixed results, "blank" = no definitive conclusion or effect

Campion et al.'s (1997) article is cited because it provides a comprehensive summary of the structured interview research to date. The interview has progressed from

being a technique with poor psychometric properties (Wagner, 1949) to being an

effective measurement tool primarily due to the use of structure (Campion et al., 1997; Conway, Jako, & Goodman, et al., 1995; Huffcutt & Arthur, 1994; McDaniel et al., 1994; Wiesner & Cronshaw, 1988). Developing interviews using job analysis information, having interviewers follow a standardized interview protocol (e.g., standard questions, tailored rating scales, limited follow-up questions, etc.), and providing extensive interview training have reduced 1) contamination from irrelevant information, 2) unreliability due to idiosyncratic variability, 3) cognitive processing load placed on raters, and 4) information processing errors (Schmitt, 1976). As a result, structured interviews yield an average reliability of .84 (McDaniel et al., 1994) and corrected validities as high as .63 (Wiesner & Cronshaw, 1994).

The structured interview used in this study incorporated most of these techniques to maximize the measurement properties of the interview process. For example, raters 1) made separate ratings for each dimension assessed by the interview, 2) used anchored rating scales that describe low, medium, and high levels of performance, 3) took detailed notes for each response, 4) did not discuss candidates before making their ratings, and 5) received training on the job demands, interview materials, and interview/evaluation process. By incorporating these procedures it was possible to examine the effects of work and non-work responses on structured interview ratings.

#### The Influence of Work versus Non-work Responses

To examine the effect of response setting on ratings, it is important to first identify the system or process that is likely to be affected. This paper will propose a model in which rater bias favoring work experience will influence rater's ideal candidate prototype and this prototype will moderate the relationship between response setting and

ratings of the response. This model is based largely on the interview research examining information use (Doughtery, Ebert, & Callender, et al., 1986; Graves & Karren, 1992) and rater stereotypes (Dipboye, 1982; Guion, 1987; Rowe, 1984, 1989). The following section of this paper will review research on cognitive processes with a specific focus on information use and stereotypes.

#### Information Use

Information use refers to the cognitive process raters employ when they attend to, weight, and integrate information to evaluate candidates. Studies have repeatedly shown that various types of information can have a significant influence on ratings. Moreover, different information will have different levels of influence. For example, Hollman (1972) found that interviewers accurately weight negative information but have a tendency to under weight positive information. As a result, negative information has a greater influence on overall ratings.

Other studies have found individual differences in the importance assigned to various pieces of information. For example, Hakel, Dobmeyer, and Dunnette (1970) found that students weight applicant experience, academic achievement, and interest significantly different from professional interviewers. Professional interviewers seem to focus almost exclusively on academic performance (accounting for 47% of the variance in ratings) while students appear to use both academic performance (24% of variance) and work experience (15% of variance) when evaluating applicants. Still other studies have found significant individual differences in cue utilization (Valenzi & Andrews, 1973). When provided with five pieces of candidate information, experienced interviewers used different combinations of information and assigned substantially

different ratings. Likewise, policy-capturing studies of interviews have found that raters weight information differently, yielding significant inter-individual differences in ratings (Dipboye, 1982; Dougherty et al., 1986; Graves & Karren, 1992).

The purpose of describing these studies is to demonstrate the effects information can have on ratings. There is substantial evidence that raters attend to and weight information differently. Raters often have unique approaches when assigning ratings and these differences affect both reliability and validity (Dougherty et al., 1986; Zedeck, Tziner, & Middlestadt, 1983). One explanation for differences in information use is the presence of rater stereotypes or ideal candidate prototypes.

#### Interviewer Stereotypes

An interviewer stereotype refers to any predisposition on the part of the interviewer to favor certain applicants or applicant characteristics over others (Arvey & Campion, 1982; Schmitt, 1976). For example, it has been found that applicants are given higher ratings when their gender fits the gender stereotype of the position to which they are applying (Shaw, 1972; Cohen & Bunker, 1975; Cash, Gillen, & Burns, 1977). Raters have pre-conceptions of the type of person they think fits the job opening, and this pre-conception then influences the interview and rating process (Dipboye, 1985).

According to Dipboye, interviewers can form initial impressions of an applicant early in the interview process that subsequently influences all other components of the interview (Dipboye, 1982; 1985; Macan & Dipboye, 1990). It has been hypothesized that interviewers develop an ideal candidate prototype (Guion, 1987; Rowe, 1984, 1989) that is used as a template for evaluating applicant qualifications (Motowidlo, 1986). Research has shown that this prototype can affect attention, information search, and recall such that

the interviewer only looks for and remembers information consistent with the prototype (Dipboye, 1982; Macan & Dipboye, 1990). Thus the cognitive processes invoked during an interview can have a significant impact on ratings. More specifically, ideal candidate prototypes held by raters can influence the way they interpret and evaluate applicant responses in an interview setting. To understand how this phenomenon occurs, it is important to understand how information is processed.

#### Feldman's Model

In 1981, Feldman examined the cognitive processes involved in the performance appraisal rating process. Feldman's theory is discussed because of the similarities between ratings on interviews and performance appraisals. According to Feldman's model, appraisals involve a cyclical, four-stage process of attention, categorization, recall, and information integration. Underlying these stages are cognitive processes that can occur automatically or in a controlled fashion. It is the combination of stages with cognitive process that explain how performance appraisal ratings are made. The cognitive process is important because it permeates all aspects of the rating process. Automatic processes are ones that are so familiar or routine that they occur with little effort or awareness. One such process as described by Feldman would be to notice height or gender. People notice or attend to these features but may not do so in a conscious manner. Because of the efficiency associated with automatic processing, it is the default approach. When information cannot be easily categorized, however, individuals must use more effortful, controlled processing.

Controlled processing occurs when a person is aware of their perceptions, observations, or thought process. For example, telling individuals to identify every

vowel in a list of words will require them to actively attend to each letter. This active processing results in a greater attention to attributes of the target. According to Feldman, people naturally attend to stimulus information (stage one of the rating process). If the information fits with a pre-existing prototype or cognitive schema it is noted and stored automatically. If however the information is unexpected, controlled processes are engaged to understand and interpret the information. Closely tied to attention is the categorization process.

Categorization refers to the manner in which individuals process and store large amounts of information despite limited processing capacity. During initial contact with an employee or applicant the rater will notice characteristics about that person. The extent to which the information resembles an existing category, schema, or prototype will determine how and "where" that information is stored. If the person fits neatly into a category, this process is automatic. Because categorization requires little cognitive processing, it is more likely that the observer or rater will associate stereotypic attributes of the *category* to the target than if the target was categorized using controlled processing.

For example, an observer might automatically categorize a man in a white lab coat as a doctor. Subsequently, the observer is likely to assume that this man carries a stethoscope even though the observer has no data on which to base this assumption. Seeing a lab coat automatically invokes the doctor schema and attributes of that schema become associated with the target. It is through this process that stereotypes are believed to influence ratings. Targets are associated with a pre-existing category and stereotypic qualities of the category are then associated with the target.

As processing becomes more controlled, attributes drawn from the category tend to have less influence on the assumptions made regarding the target. For example, a man might be wearing a white lab coat but also be carrying a test tube. The lab coat is associated with the doctor schema while the test tube is not. In this situation, the observer must actively attend to characteristics of the target to make a categorization – i.e., use controlled processing. This effortful processing results in greater accuracy in recall (i.e., the individual uses test tubes not a stethoscope). The model has an added layer of complexity due to context. Context is believed to influence the salience of target characteristics and their influence on categorization and recall. If everyone in a group is wearing a white lab coat and carrying test tubes, gender, race, height or some other distinct characteristic of targets may be more salient and have a greater effect on cognitive processing.

Feldman's (1981) model provides a process for how targets are perceived, classified, and categorized. This process may occur automatically due to characteristics of the target that fit with a pre-existing prototype or may involve controlled cognitions in which novel or conflicting qualities are integrated to reach a final judgment. Categorization is important because it then influences all future cognitions regarding the target –namely recall and information integration. According to Feldman, the level of processing (automatic vs controlled) used when the initial categorization was made will affect the amount of influence the category has on recall.

#### Fiske's Model of Stereotyping

Similar to Feldman's theory is a two-stage model of social judgments developed by Fiske and colleagues (Fiske, 1982; Fiske et al., 1987; Fiske & Pavelchak, 1986). In

the first stage of the model, the perceiver tries to categorize or classify the target through a four-step process. The steps are 1) initial categorization, 2) category confirmation, 3) recategorization, and 4) piecemeal integration. According to Fiske et al.(1987) impressions are formed about a target at the moment of first contact –i.e., initial categorization. This categorization invokes a social category or stereotype in the mind of the perceiver. Category confirmation occurs if subsequent information is consistent with the initial categorization. The perceiver tries to determine if characteristics of the target actually fit with the exemplar or prototype in their mind –which Feldman might call automatic processing. When information does not fit the prototype, recategorization occurs as the perceiver tries to match attributes of the target person with characteristics of other prototypes. If this process fails to yield an acceptable categorization, the actual characteristics of the target person are stored in memory independent of a social category. This process is very similar to Feldman's controlled processing concept and is referred to by Fiske as piecemeal integration.

The second stage of the judgment process occurs when the perceiver accesses encoded information regarding the target. This cognitive process is believed to occur on a continuum. At one extreme is "category-based" or heuristic processing (Fiske & Pavelchek, 1986) which is similar to Feldman's (1981) automatic processing. Judgments are made quickly and efficiently reflecting characteristics of the category more so than the individual. The other extreme involves "feature-based" processing (Fiske & Pavechak, 1986) or controlled processing (Feldman, 1981) in which the evaluator carefully processes all available information to make a judgment.

Similar to Feldman's model, the categorization process occurring in stage one has a strong influence on the type of information that is retrieved in stage two. If the target fits a social category in stage one, the perceiver is more likely to recall qualities and make judgments based on the category (category based processing) rather than the target. When category based processing occurs, it is difficult for the perceiver to distinguish between characteristics of the target and characteristics of the category to which the target was assigned. In other words, the target is given attributes based on a stereotype rather than his or her behaviors or characteristics. Alternatively, if piecemeal integration occurred, the processing is more deliberate or controlled. Under these circumstances, the perceiver is more likely to recall specific information or features regarding the target and use that information to make a judgment -feature based processing. Depth of processing is important for appraisals and evaluations because it is directly related to rating accuracy. More active cognitive processing leads to raters paying closer attention to the target, resulting in more accurate ratings (Erber & Fiske, 1984; Favero & Ilgen, 1989; Neuberg & Fiske, 1987).

#### Implications of Cognitive Processing for the Structured Interview

The literature on information use and interviewer stereotypes provides a process by which interviewer pre-conceptions, stereotypes, or ideal candidate prototypes influence interview ratings. Fortunately, many of the techniques used to design and conduct structured interviews invoke controlled processing thus yielding more accurate ratings (Neuberg & Fiske, 1987; Favero & Ilgen, 1989). Nevertheless, it is not clear whether structured interviews reduce errors that may result from rater biases in favor of work versus non-work responses.

The subsequent section of this paper will argue that raters are biased in favor of work experience over non-work experience. This bias then influences the ideal candidate prototype that raters hold. As a result, raters expect the ideal candidate to discuss work experiences during the interview. This prototype acts as a moderator between the setting of the response given and the ratings assigned. Other factors such as characteristics of the job as well as the applicant's level of work experience will also influence this relationship as depicted in Figure 1.



Figure 1. Hypothesized model of rater stereotypes

#### Interview Validity and Applicant Experience

A crucial component of the model presented in Figure 1 is rater biases towards experiences occurring in a work environment. It has been suggested that interview validity is moderated by applicant experience level. Campion, Campion, and Hudson, (1994) and Pulakos and Schmitt (1995) examined the experience-based structured interview using experienced employees. The mean tenure for Campion et al.'s participants was 21.7 years (SD=8.29) and Pulakos and Schmitt's participants had 1 to 6 years of experience with the organization. In their review of structuring techniques, Campion et al. (1997) proposed a moderating effect of applicant experience such that experience-based interviews are more valid with experienced applicants. While this hypothesis has not been researched, it does raise the question of whether applicant experience is an important factor in structured interview ratings.

According to a meta-analysis by Quinones, Ford, and Teachout (1995), job experience does relate to job performance. However, they found that the number of experiences and the tasks that were performed were the strongest predictors of future performance. Thus the true relationship between experience and job performance may lie in the frequency and specificity of job relevant behaviors –not in the setting (work or non-work) in which the behaviors occurred.

The assumption underlying the experience-based interview is that abilities will generalize across situations. In other words, validity *should* be a result of the *ability* being demonstrated by the behavior (Binning & Barrett, 1989) –not the setting in which the behavior occurred. If, however, the type of experience described in a response (work vs non-work) inaccurately influences ratings due to rater bias and ideal candidate prototypes, this validity assumption may be inaccurate. It is quite possible that raters use job experience as a proxy measure of job relevant knowledge, skills, and abilities. Rather than focusing on the behavior being described in an interview, they may be focusing on the setting in which the behavior occurred.

Although there are no studies that examine stereotypes regarding work versus non-work responses, there is considerable evidence for the influence of context on

behavior (Mischel, 1968; 1977; Murtha, Kanfer, & Ackerman, 1996; Pervin, 1989; Schmit, Ryan, Stierwalt, & Powell, 1995; Stewart, 1996; Weiss and Adler, 1984). People have a tendency to behave differently in different situations or contexts –irrespective of effectiveness. A work setting may invoke different types of behaviors than a non-work setting. Moreover, there is evidence that considering behavior in context can lead to more valid prediction of future performance (Schmitt et al., 1995). Thus how someone behaves at work may be a better predictor of future job performance than how he or she behaves in other settings.

The point of citing these studies is to reiterate the important role of context when evaluating behavior. This study will not examine validity. However, it will examine whether raters are biased in favor of work versus non-work responses and whether that bias influences ratings. If ratings differ significantly, subsequent studies can examine whether those differences are valid predictors of future job performance or rating error. It is hypothesized that raters will prefer responses involving a work setting because behaviors at work will be perceived as being more job relevant and better indicators of future job performance. Thus, raters will assign higher ratings to responses involving a work setting regardless of the effectiveness or job relevance of the behavior described.

HYPOTHESIS 1: Interviewers will rate responses describing behaviors that occurred in a work setting more positively than responses describing behaviors that occurred in a non-work setting.

Drawing on Dipboye (1982; 1985), Feldman (1981), and Fiske et al.'s (1987) research, it is expected that rater bias will result in greater halo error for work responses. Halo in this study will be operationalized as variance in ratings across the ten rating

dimensions. It is important to note that rating variance itself may not be an error; it may reflect true relationships across performance dimensions. What is of interest in this study is whether rating variance is significantly different across work and non-work responses and is in the hypothesized direction.

It is expected that responses occurring in a non-work setting will be inconsistent with ideal candidate prototypes and therefore invoke controlled processing. This controlled processing should lead to more careful evaluation of responses and greater variability across ratings –i.e., less halo error. Greater variability will presumably be caused by "true" differences in response effectiveness as opposed to rater biases or stereotypes that would reduce rating variability. Conversely, work responses will invoke automatic processing because responses will fit with the ideal candidate prototype and rater biases favoring work experience. The result will be less variability in the ratings of work responses because ratings will reflect biases or stereotypes rather than differences in the effectiveness of responses.

HYPOTHESIS 2: Work responses will yield greater halo error as demonstrated by lower variance in ratings across the ten rating dimensions.

Although the two hypotheses above describe main effects for response setting, the true relationship between setting and interview ratings is expected to be moderated by rater biases towards work responses and ideal candidate prototypes. Specifically, rater biases are expected to influence ideal candidate prototypes, which will then moderate the relationship between response setting and interview ratings. Research has suggested that interviewers make hiring decisions by comparing applicants to prototypes of an ideal candidate (Dipboye, 1985; Hakel, Hollmann, & Dunnette, 1970; Webster, 1964). It is

expected that this prototype will be directly influenced by rater biases towards work responses. It is through ideal candidate prototypes that rater biases are expected to influence the ratings of work and non-work responses. Therefore, ideal candidate prototypes are expected to mediate the effects of rater biases on interview ratings.

HYPOTHESIS 3: The influence of rater biases on ratings will be mediated through raters' ideal candidate prototypes.

Ideal candidate prototypes are expected to moderate the relationship between response setting and interview ratings such that mean differences and halo error of ratings will be more pronounced as the ideal candidate prototype favors work responses.

HYPOTHESIS 4: Ideal candidate prototypes will interact with response setting such that mean differences favoring work responses will be greater as the prototype increasingly favors work responses.

HYPOTHESIS 5: Ideal candidate prototypes will interact with response setting such that halo error will be greater for work responses as the prototype increasingly favors work responses (see Figure 2 for a graph of the hypotheses).

As depicted in the model presented in Figure 1, the relationship between rater bias, ideal candidate prototypes, response setting, and interview ratings are likely to be influenced by two additional factors –characteristics of the job and the work experience of the applicant. To minimize complexity, these two factors were held constant throughout the study. The job was described as a management position, and both candidates had equivalent work experience.



Figure 2. Graph of hypotheses

#### METHODS

The study was a 2X2X2X2 factorial design with three between subject factors and one within subject factor. The between subject factors were 1) confederate (e.g., confederate 1 or 2), 2) the response setting (work or non-work) for the first interview (i.e., the first interview viewed), and 3) the response setting for the second interview (i.e., the second interview viewed). The within subject factor was ratings for the first and second interviews. The 16 study conditions that resulted are shown in Table 2. As a point of clarification, it is important to distinguish the confederate variable from the interview variable. In all of the analyses and discussion that follow, confederate 1 and 2 refers to the two individuals portrayed on the videotaped interviews as the job applicants. Alternatively, interview 1 and 2 or the first and second interview refers to the first interview watched (interview 1) or the second interview watched (interview 2). Prior to evaluating the interviews, undergraduate participants received training on how to rate structured interview responses. The details of the study are outlined in the following paragraphs.

#### **Participants**

Two hundred sixty-seven undergraduates participated in this study in return for psychology research credits. Two cases were removed from the data set because participants knew the confederates. An additional 7 were removed for assigning extreme interview ratings. The variance of the extreme ratings was three or more standard deviations above the mean, which is close to assigning only 1 and 5 ratings on a 5 point scale. As interview responses contained a range of effectiveness levels, the variances

seemed extreme and were excluded from the data set. The result was a total sample of 258. Eighty percent of participants were 18 to 20 years old, 70% were in their first or second year of college, 84% were Caucasian, and 79% were female. Eighty-seven percent had participated in 2 to 4 job interviews, 15% had interviewed someone else for a job, 65% had worked between 2 and 4 part-time jobs, and 55% of participants had held jobs for 4 to 6 years of their life.

Confederate viewed first	Response Setting for interview 1	Response Setting for interview 2	Rating
1	work	work	interview 1 interview 2
		non-work	interview 1 interview 2
-	non-work	work	interview 1 interview 2
		non-work	interview 1 interview 2
2	work	work	interview 1 interview 2
non-work		non-work	interview 1 interview 2
	non-work	work	interview 1 interview 2
		non-work	interview 1 interview 2

Table 2. Study Conditions

There has been considerable debate in the literature that college students may not be an appropriate sample for examining interview ratings. In fact, there is evidence that college students are more lenient than professional interviewers (Bernstein, Hakel, & Harlan, 1975) and may not be appropriate for making *hiring* decisions (Hakel, Ohnesorge, & Dunette, 1970). However, the literature has consistently found similarities in information processing and decision making across professional and student raters (Arvey & Campion, 1982; Bernstein et al., 1975; Dipboye, Fromkin, & Wibach, 1975; McGovern, Jones, & Morris, 1979). As the current study examines information processing and decision making, there is research to support the use of a college student sample for this study. In addition, the demographic data indicate that the majority of participants were familiar with job interviews and had held jobs themselves. Thus, there is evidence that using a college student sample for this study should yield results comparable to those that would be found with professional interviewers.

#### **Development of the Interview Materials**

The experience-based structured interview used in this study was adapted from one in use at an auto parts manufacturing plant. The interview materials were designed in the same manner as the interviews used by Motowidlo et al. (1992) and Pulakos and Schmitt (1995). Development began with a thorough job analysis. The analysis was a multi-stage process consisting of a review of all training materials, job observation, critical incident development, and focus group meetings with job incumbents. The result was a comprehensive list of task, knowledge, skill, and ability statements needed on day one of the job. These statements were organized into ten dimensions.

Using the job analysis information, two or three experience-based interview questions were developed for each dimension. The questions asked applicants to describe past behavior that demonstrated abilities necessary for the management position. Job incumbents then reviewed the questions and wrote sample responses of varying degrees of effectiveness based on their experiences or the experiences of co-workers. These responses were used to develop behavior summary scales that detailed the characteristics of a poor response, moderate response, and excellent response. In this manner, ten anchored rating scales were developed. Participants used these scales to rate

the videotaped interview responses presented in this study (see Appendix B for materials).

#### <u>Design</u>

All participants received the same interview training, which reviewed the interview and rating guidelines. Participants then viewed and evaluated two videotaped interviews. The interviews either consisted of two candidates describing work experiences, two describing non-work experiences, or one describing work and one describing non-work experiences. The order of the confederates and response settings were counterbalanced across condition to test for order or confederate effects. Ratings were made immediately after each interview, and study measures were collected after the interview rating process.

#### Procedure

Upon entering the laboratory, participants were given a brief description of the study and asked to review and complete the informed consent form shown in Appendix F. Participants received a 20-minute training on the job and interview rating process. They then practiced rating two written interview responses and discussed their ratings as a group. Following the training, participants reviewed the resume for the first interview and answered three questions (i.e., resume manipulation check). They then watched a work or non-work interview from one of the two confederate applicants. During the interview participants were asked to record the situation, action, and result for each response. After listening to all ten interview responses, participants were given an opportunity to review their notes and assign a rating for each interview dimension using anchored rating scales. Participants then reviewed a second resume, answered questions
on the resume, and viewed another work or non-work interview from the other confederate. Participants again took notes during the interview and made their ratings after watching all ten responses.

When all participants had finished evaluating the second interview they were asked to complete two applicant perception checklists (i.e., ideal candidate prototype measures), a rater bias measure, and an item asking who they would hire for the job. Finally, they completed a demographic information form indicating their class standing, race, GPA, age, sex, and ACT score (see Appendix E). Participants were also asked to indicate their level of interview and work experience. All participants were then debriefed (see Appendix F).

# Training

The study began with rater training. All participants were trained in structured interview techniques and rating procedures. The training included instructions on 1) taking notes, 2) assigning ratings, and 3) avoiding rating errors. The experimenter guided participants through each portion of the training manual. Participants were then given the opportunity to practice rating two sample responses and discuss their ratings (the samples contained one work response and one non-work response of equivalent effectiveness levels). Although the training emphasized the importance of focusing on the behaviors being described in the interview, it did not include instruction on how work versus non-work responses should be evaluated. This omission was designed to measure the impact of general structured interview training on the ratings of work and non-work responses. Training required approximately 20 minutes to complete (see Appendix B).

#### Stimulus Materials

To enhance the likelihood of finding an effect, the stimulus materials were designed to maximize differences in response setting across interviews from the same confederate and invoke any rater biases towards work experience that might exist.

Interview Reponses. Each interview contained responses occurring exclusively in a work or non-work setting. To avoid systematic differences in interviews, steps were taken to make comparable videotapes. First, two male confederates were used. Secondly, each confederate recorded two virtually identical interviews with the exception of the setting of the response. For example in the work response, the confederate might describe a time in which he had to coordinate a work team while in the non-work response, he would describe a time in which he coordinated an athletic team. Third, interviews were of similar length, lasting approximately 15 minutes each. Fourth, responses within each interview were written to include a range effectiveness levels. Transcripts are shown in Appendix C.

<u>The Job.</u> It was expected that the level of the job (management or entry) would influence the extent to which raters favor work over non-work responses. For this reason, participants were told that they would be hiring for a management position. All participants were given a description of the organization (an auto parts manufacturing plant) and job requirements (manage a production team of 7-10 people; see Appendix B).

<u>Candidate Experience</u>. Before watching an interview, participants were asked to review the candidate's resume and answer three questions. Each resume contained equivalent experience in an area similar to the job (i.e., 3 years of parts manufacturing at one of the big three auto manufacturers). This information was provided to eliminate

differences in actual work experience across candidates. Controlling for actual work experience should have minimized the likelihood that raters would make erroneous assumptions regarding an applicant's work experience based on his responses (see Appendix D).

# <u>Measures</u>

Participants were asked to make a number of ratings and complete a series of measures throughout the course of the study. The dependent variables were the ratings of the two interviews, the variance of the ratings (i.e., halo error), and a one-item hiring decision. Independent variables included adjective checklist ratings (i.e., ideal candidate prototype ratings) and ratings on the rater bias measure, as well as the three dummy coded variables described previously (confederate, response setting for the first interview, response setting for the second interview). Participant experience being interviewed and interviewing others were entered as covariates, and manipulation checks for the two resumes as well as two practice ratings were collected. Each of these measures is described below.

Interview ratings. After watching a full interview, participants were given time to review their notes and use the anchored ratings scales to rate the interview on the ten dimensions assessed. The rating sheet is shown in Appendix E. Ratings were made independently with no discussion among participants/raters. The mean across dimension ratings was then computed and used as the dependent interview rating variable.

<u>Halo error.</u> Halo error was assessed by comparing the relative variance in ratings across interviews. Halo error occurs when there is low variance in ratings across dimensions due to a single causal factor such as rater bias or ideal candidate prototypes.

Halo error was calculated in this study by computing the standard deviation across interview dimensions for each interview. The result was a standard deviation for ratings of the first and second interview for each study participant. These values served as the dependent variables when examining the influence of response setting on halo error. As there was only one standard deviation per interview per participant, reliability could not be determined.

٠

<u>Hiring decision.</u> After evaluating both interviews, study participants were asked to indicate who they would hire for the job, the candidate in interview 1 or interview 2 (see Appendix E).

Adjective checklist. After rating both interviews, participants were asked to complete a 23-item adjective checklist for each interview. The questionnaire was intended to assess raters' ideal candidate prototypes. Items were drawn from an original list of 99 descriptors found in Hakel et al.'s (1970) job applicant checklist and other adjectives commonly used to describe personality and interpersonal characteristics (e.g., dependable, hard-working, etc.).

The final list of items was selected based on the ratings of 46 undergraduates. After receiving a description of the management position being used in the study, students were asked to rate each of the 99 descriptors on whether they described their ideal candidate for the job. Items that were rated very positively (mean 4.5 or greater) with a low standard deviation (0.50 or less) were retained. The result was a 23-item scale (alpha = 0.90) of descriptors that virtually all of the students agreed described the ideal candidate (see Appendix E).

The scale was said to assess raters' ideal candidate prototypes because 46 participants agreed that the items forming the scale described the ideal candidate for the job described in the study. In addition, the items in the scale were not directly assessed in the interview, thus students would have to infer whether or not an item accurately described a candidate. If raters made category-based or stereotypical attributions about candidates as hypothesized, ratings on the checklists should be more positive (i.e., closer to the ideal candidate profile) for candidates giving work responses than for candidates giving non-work responses.

Rater bias measure. The rater bias measure was a four-item scale designed to assess raters' perceptions of the importance of work experience for job performance. A sample item is, "Work experience is the best predictor of future job performance." High scores on the measure would indicate a positive perception of work experience (see Appendix E).

Interview experience. Experience being interviewed and interviewing others were included as control variables in all of the analyses. Participants were asked to indicate the number of times they had been interviewed for a job and the number of times they had interviewed someone else for a job (see Appendix E).

Resume manipulation check. Before watching a videotaped interview, participants reviewed a resume and completed a three-item questionnaire. The questionnaire asked about candidates' amount and type of work experience on the resume to ensure that participants were familiar with candidates' work experience prior to viewing their interviews (see Appendix E).

<u>Practice ratings.</u> After completing the interview training, participants were asked to review two written interview responses and evaluate their effectiveness. The responses were written to be of equivalent effectiveness levels and involved one nonwork and one work response respectively (see Appendix E).

#### Analyses

Hypotheses 1 and 2 were analyzed using repeated measures analysis of covariance (ANCOVA). The effects of response setting and confederate on mean ratings and rating variance were examined controlling for interview experience. Hypothesis 1 examined the effect of response setting on interview ratings. Ideally, there would be no confederate or order effects allowing the two variables to be dropped from the analyses. Under these circumstances, the strongest support for the hypothesis would be a significant main effect for response setting in which work responses were rated significantly higher than non-work responses. If confederate or order effects were found, hypothesis 1 would be supported if work responses were rated significantly higher than non-work responses.

Hypothesis 2 examined the effect of response setting on halo error (operationalized as within interview rating variance). Similar to hypothesis 1, the strongest support for hypothesis 2 would be a significant main effect for response setting irrespective of confederate and order of presentation. Given that halo error is associated with low rating variance, supporting results would indicate lower rating variance for work responses than non-work responses. Once again, if confederate or order effects were found, hypothesis 2 would be supported if work responses invoke significantly greater error than non-work responses.

Hypothesis 3 proposed that rater bias would relate to ideal candidate prototypes and ideal candidate prototypes would moderate the relationship between response setting and interview ratings. The complexity of the design and use of continuous variables were not conducive to a repeated measures ANCOVA. Therefore, regression equations were used to test the hypothesis for the first and second interview separately, controlling for interview experience. Support for hypothesis 3 would require finding 1) a significant relationship between rater bias and ideal candidate prototypes and 2) a significant interaction between ideal candidate prototypes and response setting on interview ratings. If these relationships were found, more specific analyses could be conducted to test the proposed model.

Hypothesis 4 proposed that ideal candidate prototypes would moderate the relationship between response setting and interview ratings. This hypothesis was tested separately for each interview using moderated regression controlling for interview experience. A significant interaction between response setting and adjective checklist ratings in which higher checklist ratings combined with work responses yielded higher interview ratings would support hypothesis 4.

Hypothesis 5 proposed that ideal candidate prototypes would moderate the relationship between response setting and halo error. This hypothesis was tested separately for each candidate using moderated regression controlling for interview experience. A significant interaction between work setting and adjective checklist ratings in which high adjective checklist scores combined with work responses yield lower variance in ratings would support hypothesis 5.

## RESULTS

The first step in the results was to examine the descriptive statistics for the study data. Descriptive statistics and scale intercorrelations for the interview ratings, adjective checklists, and rater bias measure are listed in Table 3. Scale scores for the interview, checklist, and rater bias measures were formed by computing the mean rating across items. Halo for the first and second interview was computed by taking the standard deviation of ratings across the ten interview dimensions. Means and standard deviations did not indicate floor or ceiling effects in the ratings and the intercorrelations among ratings indicated higher relationships within interviews than across interviews. For example, adjective checklist ratings for interview 1 were more highly correlated with ratings for interview 1 (r = .48) than with ratings for interviews (i.e., interview, halo, and checklist ratings) are indicative of rater effects, which are not unusual given the repeated measures design of the study.

# Manipulation Check

Before watching the interviews, participants were asked to review each candidate's resume and answer three questions on the candidate's work experience. Results indicated that 94% of respondents answered all six of the questions correctly (three questions for each candidate). Therefore, the large majority of participants were at least aware of the fact that both candidates had relevant work experience.

	Mean	SD	1	2	3	4	5	6	7
1 Interview 1 <sup>a</sup>	3.35	0.40	(.65)						
2 Halo Interview 1 <sup>b</sup>	0.90	0.20	-0.14*	n/a					
3 Checklist 1 <sup>a</sup>	3.94	0.42	0.48*	0.00	(.92)				
4 Interview 2 <sup>a</sup>	3.31	0.50	0.32*	0.10	0.09	(.79)			
5 Halo Interview 2 <sup>b</sup>	0.85	0.21	-0.11	0.45*	0.02	-0.12	n/a		
6 Checklist 2 <sup>a</sup>	3.88	0.56	0.20*	0.07	0.33*	0.69*	-0.07	(.95)	
7 Rater Bias <sup>a</sup>	3.56	0.72	0.00	0.04	0.10	0.05	0.045	-0.03	(.75)

Table 3. Descriptive statistics and scale intercorrelations

Note: n/a = not applicable (one item measures).

Internal consistency listed on the diagonal

N = 258

<sup>a</sup>Interview, checklist, and rater bias measures computed using the mean of item ratings.

<sup>b</sup>Halo computed using the standard deviation of interview ratings.

\*p<.05

	Mean	SD	1	2	3	4	5	6	7	8	9
1 Problem Solving	3.69	0.67									
2 Leading	3.56	0.79	0.24*								
3 Interpersonal	3.06	0.84	0.27*	0.00							
4 Training Others	2.57	0.84	0.09	0.22*	-0.01						
5 Flexibility	2.88	0.95	0.23*	0.10	0.21*	0.17*					
6 Continuous	4.26	0.71	0.15*	0.14*	0.12*	0.00	0.18*				
7 Quality	3.31	0.92	0.14*	0.20*	0.01	0.27*	0.01	0.07			
8 Safety	2.90	0.88	0.20*	0.21*	0.08	0.23*	0.27*	0.22*	0.28*		
9 Planning	3.14	0.71	0.20*	0.12	0.14*	0.05	0.16*	0.09	0.11	0.29*	
10 Communicating	4.11	0.73	0.15*	0.16*	0.19*	0.19*	0.21*	0.12*	0.23*	0.30*	0.17*
· · · · · · · · · · · · · · · · · · ·											

Table 4. Dimension Intercorrelations for Interview 1

\*p<.05.

	Mean	SD	1	2	3	4	5	6	7	8	9
1 Problem Solving	3.75	0.80									
2 Leading	3.58	0.82	0.30*								
3 Interpersonal	2.98	0.87	0.23*	0.24*							
4 Training Others	2.59	0.88	0.13*	0.32*	0.16*						
5 Flexibility	2.84	0.89	0.27*	0.30*	0.27*	0.33*					
6 Continuous	4.05	0.79	0.32*	0.19*	0.25*	0.19*	0.21*				
7 Quality	3.26	1.07	0.21*	0.26*	0.15*	0.37*	0.20*	0.25*			
8 Safety	2.95	0.81	0.20*	0.36*	0.24*	0.29*	0.24*	0.25*	0.40*		
9 Planning	3.11	0.73	0.28*	0.49*	0.34*	0.30*	0.34*	0.26*	0.33*	0.43*	
10 Communicating	4.00	0.75	0.27*	0.25*	0.26*	0.26*	0.26*	0.31*	0.35*	0.35* (	0.36*

Table 5. Dimension Intercorrelations for Interview 2

\*p<.05.

 Table 6. Observed means of interview ratings

			Order				
			Interview 1		Interview 2		-
	Response	Response					
Confederate viewed first	Setting 1 <sup>a</sup>	Setting 2 <sup>b</sup>	Mean	SD	Mean	SD	N
1	non-work	non-work	3.30	0.35	3.13	0.49	26
		work	3.31	0.47	3.34	0.50	30
	work	non-work	3.54	0.39	2.97	0.47	27
		work	3.41	0.38	3.25	0.49	34
2	non-work	non-work	3.35	0.32	3.05	0.48	53
		work	3.24	0.39	3.29	0.49	64
	work	non-work	3.48	0.47	3.43	0.42	38
		work	3.23	0.34	3.56	0.47	34

<sup>a</sup>Response Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Response Setting 2: setting, work or non-work of the second interview viewed.

			Order				
			Interview 1		Interview 2		-
	Response	Response					
Confederate viewed first	Setting 1 <sup>ª</sup>	Setting 2 <sup>b</sup>	Mean	SD	Mean	SD	N
1	non-work	non-work	0.92	0.21	0.90	0.30	26
		work	0.87	0.20	0.89	0.15	30
	work	non-work	0.81	0.16	0.87	0.21	27
		work	0.92	0.19	0.89	0.18	34
2	non-work	non-work	0.86	0.20	0.76	0.17	38
		work ·	0.93	0.20	0.86	0.24	34
	work	non-work	0.91	0.26	0.87	0.20	27
		work	0.97	0.18	0.83	0.20	37

 Table 7. Observed means of rating standard deviations

<sup>a</sup>Response Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Response Setting 2: setting, work or non-work of the second interview viewed.

Reliabilities for the checklists, rater bias measure, and ratings for the second interview were within an acceptable range (.79 - .95) but the reliability for ratings of the first interview were slightly lower (.65). Correlations among ratings for the first interview were also low (.00 to .30) and moderate for the second interview (.13 to .49) as shown in Tables 4 and 5. These results could be indicative of the multi-dimensionality of the interview ratings or a result of counterbalancing. A factor analysis of the ratings did not produce clearly interpretable factors. Given that individual interview ratings are often aggregated to form a single composite, the decision was made to examine the mean rating across all ten interview dimensions. Observed cell means for interview ratings are presented in Table 6. Observed cell means for halo error (i.e., the standard deviation of ratings) are presented in Table 7.

#### Confederate & Order Effects

Dummy coded variables for confederate and response setting were created to test the proposed hypotheses. However, given the use of two confederates and the repeated measures design of the study, it was important to first test for confederate and order effects in the data. If no effects were found, the confederate variable could be dropped from the analyses and response setting could be collapsed across order. Of most concern would be finding significant, crossing interactions between response setting and confederate, response setting and order, or response setting, confederate, and order. This finding would mean that work responses could be rated significantly higher or significantly lower than non-work responses depending upon the confederate or order of presentation. Alternatively, significant main effects for the confederate variable or an interaction between confederate and order would indicate confederate and confederate by

order effects respectively, making it necessary to include the confederate and order variables in the analysis of the hypotheses.

The data were analyzed using repeated measures ANCOVA controlling for interview experience. Box's M test was non-significant indicating non-significant differences in covariance matrices across conditions; given that only two time periods were examined, the repeated measures ANCOVA assumption of sphericity was met.

Confederate & order effects for interview ratings. As can be seen from the within subject results shown in Table 8, confederate significantly interacted with order indicating a confederate by order effect. The marginal means are graphed in Figure 3. They indicate that within subject ratings favored confederate 1, but tended to favor him more when he was viewed second, while confederate 2 was rated less favorably when viewed second. Although these results were not intentional, they do not pose a threat to the proposed hypotheses. The within subject interaction does not cross indicating a similar pattern of ratings for the first and second interview. More importantly, the within subject results do not indicate confederate by setting interactions. Therefore, the tendency to favor confederate 1 occurs across order as well as work and non-work responses when examined within subject.

Source	Туре Ш SS	df	Mean Square	F	Sig.	Eta <sup>2</sup>
Order	0.05	1	0.05	0.39	0.53	0.00
Order*Interview Experience (covariate)	0.06	1	0.06	0.52	0.47	0.00
Order*Interviewed Others (covariate)	0.22	1	0.22	1.90	0.17	0.01
Order*Confederate	3.37	1	3.37	28.55	0.00	0.11
Order*Setting 1 <sup>a</sup>	1.74	1	1.74	14.78	0.00	0.06
Order*Setting 2 <sup>b</sup>	1.84	1	1.84	15.61	0.00	0.06
Order*Confederate*Setting 1	0.07	1	0.07	0.60	0.44	0.00
Order*Confederate*Setting 2	0.12	1	0.12	1.05	0.31	0.00
Order*Setting1*Setting 2	<b>0.02</b> ·	1	0.02	0.13	0.72	0.00
Order*Confederate*Setting 1*Setting 2	0.24	1	0.24	2.01	0.16	0.01
Error(Interview 1 & 2)	28.65	243	0.12			

Table 8. Tests of Within-Subjects Effects for Interview Ratings

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed.



Figure 3. Confederate by order interaction for interview ratings

Conversely, the between subject results, shown in Table 9, indicate a significant interaction between confederate and response setting for the second interview. The "<" shape of the graph (Figure 4) indicates that when a work response is given for the second interview, interview ratings (collapsed across order) are equivalent for both confederates. However, when the second interview is a non-work response, confederate 2 is rated higher than confederate 1 and higher than ratings for work responses. The concern is that confederate 2's non-work response is rated significantly higher than the work responses contradicting hypothesis 1 and hypothesis 4. Three univariate ANCOVA analyses were run to test whether the means depicted in Figure 4 were significantly different from one another.

The first analysis examined only the non-work responses to see if the means for confederate 1 and 2 were significantly different from one another. Results indicated that

the difference was significant (F=7.90, df=1, 114, p<.01). When examined between subjects, the second confederate's non-work response is rated more positively than the first confederate's non-work response. Although these results were not intended, they do not pose a problem for the interpretation of the hypotheses. The critical factor is whether non-work responses are rated more positively than work responses, which would contradict hypothesis 1.

Separate analyses for each confederate were conducted to test whether ratings for work responses differed significantly from ratings of non-work responses. Results were non-significant (confederate 1: F=2.14, df=1,113, p>.10; confederate 2: F=2.24, df=1,132, p>.10). The means for work and non-work responses within confederate were not significantly different from one another. Therefore, neither the significant difference across confederates for non-work responses nor the significant interaction between confederate and response setting should pose a problem when testing the proposed hypotheses. Nevertheless, the confederate variable was included in all of the analyses that follow.

Order effects were also found for response setting, however, these effects relate to hypothesis 1 and will be addressed in the section that follows.

<u>Confederate & order effects for halo error.</u> Halo error was assessed by computing the standard deviation of ratings for each interview and comparing them across interviews; lower standard deviations indicate greater halo error (cell means are shown in Table 7). Before testing the proposed hypotheses, it was important to test the halo measure for order and confederate effects.

Source	Туре Ш SS	df	Mean Square	F	Sig.	Eta <sup>2</sup>
Intercept	549.10	1	549.10	2152.08	0.00	0.90
Interview Experience (covariate)	0.33	1	0.33	1.30	0.26	0.01
Interviewed Others (covariate)	0.13	1	0.13	0.52	0.47	0.00
Confederate	1.14	1	1.14	4.47	0.04	0.02
Setting 1 <sup>a</sup>	0.00	1	0.00	0.01	0.92	0.00
Setting 2 <sup>b</sup>	0.00	1	0.00	0.00	0.96	0.00
Confederate*Setting 1	0.10	1	0.10	0.38	0.54	0.00
Confederate*Setting 2	1.13	1	1.13	4.42	0.04	0.02
Setting 1*Setting 2	0.31	1	0.31	1.23	0.27	0.01
Confederate*Setting 1* Setting 2	0.30	1	0.30	1.17	0.28	0.00
Error	62.00	243	0.26			

 Table 9. Tests of Between-Subjects Effects for Interview Ratings

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed.



Figure 4. Confederate by response setting interaction for interview ratings Results for the repeated measures ANCOVA are presented in Tables 10-11. As can be seen for the within subject results presented in Table 10, confederate interacted significantly with order. The graph of this effect formed a flat "X" shape as shown in Figure 5. When confederate 1 was viewed first, halo error was equivalent across interview ratings –yielding a virtually straight line on the graph. However, when confederate 2 was viewed first, interview 1 had less halo (i.e., more variance) and interview 2 had more halo (i.e., less variance) than when confederate 1 was viewed first. Separate analyses were run to test for the significance of these differences.

Within subject differences across the first and second interview were tested using repeated measures ANCOVA. When confederate 1 was viewed first, halo was non-significantly different across the two interviews (F=1.29, df=1, 114, p>.10). However, when confederate 2 was viewed first, the differences in halo were significant (F=7.38, df=1,133, p<.05). When participants viewed confederate 2's interviews first, their ratings

for the second interview (confederate 1's interviews) yielded significantly more halo error. This finding however, does not affect the interpretation of the proposed hypotheses. What is important is whether the between subject ratings for confederate 2 yield significantly less halo error when viewed first, but significantly more when viewed second when compared to confederate 1 (i.e., whether the crossing interaction depicted in Figure 5 is significant between subjects).

Source	Туре Ш SS	df	Mean Square	F	Sig.
Halo	0.03	1	0.03	1.48	0.23
Halo*Interview Experience (covariate)	0.00	1	0.00	0.22	0.64
Halo*Interviewed Others (covariate)	0.03	1	0.03	1.46	0.23
Halo*Setting 1 <sup>a</sup>	0.00	1	0.00	0.16	0.69
Halo*Setting 2 <sup>b</sup>	0.02	1	0.02	0.79	0.37
Halo*Confederate	0.27	1	0.27	12.48	0.00
Halo*Setting 1*Setting 2	0.12	1	0.12	5.62	0.02
Halo*Setting 1*Confederate	0.00	1	0.00	0.14	0.71
Halo*Setting2*Confederate	0.00	1	0.00	0.02	0.88
Halo*Setting1*Setting2*Confederate	0.00	1	0.00	0.00	0.99
Error(Halo)	5.29	243	0.02		

Table 10. Test of within subject effects for halo error

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed.

	Type III		Mean		
Source	SS	df	Square	F	Sig.
Intercept	40.61	1	40.61	656.87	0.00
Interview Experience (covariate)	0.05	1	0.05	0.88	0.35
Interviewed Others (covariate)	0.01	1	0.01	0.19	0.66
Setting 1 <sup>a</sup>	0.01	1	0.01	0.20	0.66
Setting 2 <sup>b</sup>	0.11	1	0.11	1.83	0.18
Confederate	0.02	1	0.02	0.30	0.58
Setting1*Setting2	0.01	1	0.01	0.14	0.71
Setting1*Confederate	0.15	1	0.15	2.41	0.12
Setting2*Confederate	0.03	1	0.03	0.42	0.52
Setting1*Setting2*Confederate	0.25	1	0.25	4.12	0.04
Error	15.02	243	0.06		

Table 11. Between subject effects for halo error

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed.



# Figure 5. Confederate by order interaction for halo error

The between subject differences were tested using univariate ANCOVA. The differences across confederates were non-significant for interview 1 (F=1.65, df=1, 249, p>.10), but significant for interview 2 (F=5.90, df=1, 249, p<.05). Therefore, although the lines in Figure 5 cross, the means are non-significantly different for interview 1. It is only when viewed second, that ratings for confederate 1 have significantly less variance or more halo error than confederate 2. When combined with the within subject ratings, these results indicate that when confederate 1 is viewed second, he induces significantly more halo error in ratings (as indicated by lower variance) than confederate 2. This finding would only pose a problem for the halo error hypotheses (hypotheses 2 and 5) if the differences varied across response setting. As Table 10 and 11 demonstrate, no such differences exist.

The repeated measures ANCOVA also identified a significant within subject interaction for order and response setting, but this interaction relates to hypothesis 2 and will be addressed in that section.

Finally, a significant between subject, three-way interaction was found for response setting 1, response setting 2, and confederate. Marginal means are graphed in Figures 6 and 7. The "<" shaped graph depicted in Figure 6 indicates that non-work responses invoked greater halo error (i.e., less rating variance) when confederate 2 is viewed first, contrary to hypothesis 2. Univariate ANCOVA analyses were run to identify where significant differences in halo error existed.



Figure 6. Three-way interaction for response settings and confederate

The differences across confederates reflect the significant confederate by order interaction discussed previously. When two work responses are viewed, the differences in halo error across confederates are non-significant (F=0.00, df=1, 67, p>.10). However, when two non-work responses are viewed, there is significantly more halo error when



<u>Figure 7.</u> Three-way interaction for response settings and confederate confederate 2 is viewed first and confederate 1 is viewed second (F=4.88, df=1, 60, p<.05). As before, this difference does not affect the proposed hypotheses. However, the differences across response settings could be problematic.

As one might expect, the minor differences in halo error across work-work interview pairs and non-work-non-work interview pairs were non-significant when confederate 1 was viewed first (F=.04, df=1, 56, p>.10); but significant when confederate 2 was viewed first (F=5.71, df=1, 71, p<.05). Non-work-non-work response pairs invoked greater halo error (less rating variance) than work-work response pairs, contrary to hypothesis 2. This finding has implications for the halo error hypotheses and will be addressed after examining the results presented in Figure 7.

Figure 7 is a ">" shaped graph depicting the halo error for work-non-work and non-work-work interview pairs. Mean differences were tested for significance using univariate ANCOVA. Results were non-significant. Halo error across interview pairs (i.e., work-non-work and non-work-work pairs) was not significant regardless of whether confederate 1 was viewed first (F=.39, df=1, 113, p>.10) or confederate 2 was viewed first (F=2.33, df=1, 132, p>.10). Likewise the difference in halo when confederate 1 was viewed first was non-significantly different from when confederate 2 was viewed first for the non-work-work pairs (F=.11, df=1, 60, p>.10) and the work-non-work pairs (F=1.67, df=1, 50, p>.10).

As a whole the findings depicted in Figures 6 and 7 do not support hypothesis 2 and pose potential problems for hypothesis 5. Reasons for this finding will be addressed in the discussion section. One explanation is that the effect of response setting on halo error is not as strong as hypothesized or may be in the opposite direction. Another explanation based on Feldman's stereotyping research is that context minimizes the effects of response setting when similar settings are paired (i.e., work-work or non-worknon-work responses). It is quite possible that under these circumstances, other factors such as characteristics of the confederate influence halo error more so than response setting. Feldman might argue that when two work or two non-work responses are viewed, response setting is not salient because it is a common feature across applicants. It is only when work and non-work responses are paired that response setting has an influence on halo -though not to a significant level. Given this potential explanation the decision was made to proceed with the analyses. As the following results will indicate, the significant differences in halo error occur only when ratings are collapsed across order.

# Test of Hypotheses

<u>Hypothesis 1.</u> Hypothesis 1 stated: interviewers will rate responses describing behaviors that occurred in a work setting more positively than responses describing behaviors that occurred in a non-work setting. As confederate and order effects were found, it was not appropriate to only examine the main effects of response setting in the absence of the confederate and order variables. Nevertheless, the data would support hypothesis 1 if ratings were significantly higher for work responses. The analyses for this hypothesis were run previously when testing for order and confederate effects (Tables 8-9).

Results indicated that response setting of the first interview interacted significantly with order as did response setting of the second interview –demonstrating order by setting effects. Marginal means for the interactions are graphed in Figures 8 and 9 respectively. Both graphs are virtually identical to one another, forming an "X" shaped interaction.

The direction of the means are consistent with hypothesis 1. However, it was important to test the means for significant differences. A between subjects, univariate ANCOVA was used to test the differences across response setting for each interview separately. Because the hypothesis was directional, an alpha of .10 was used in evaluating significance. Within subject differences were tested by selecting the appropriate cases in the data set and using a repeated measures ANCOVA. Results for Figures 8 and 9 are presented separately.



Figure 8. Response setting 1 by order interaction for interview ratings The first analysis for Figure 8 tested whether mean differences across response setting for the first interview were significant. Results indicated that they were; work responses were rated significantly higher than non-work responses consistent with hypothesis 1 (F=3.87, df=1, 253, p<.10). Therefore, when order is removed from the analyses (by examining only the first interview watched) and between subject ratings are examined, ratings are significantly higher for work responses.

The second analysis examined whether ratings for the second interview differed significantly based on the response setting of the first interview. Again, the results were significant (F=4.77, df=1, 253, p<.10). When a non-work response was viewed first, the second interview was rated significantly higher, regardless of response context; and when a work response was rated first, the second interview was rated significantly lower, regardless of response context. These results reflect the order effect described earlier. It appears that the first interview influences how the second interview is rated, but the results do support hypothesis 1.

The within subject differences for Figure 8 were tested using a repeated measures ANCOVA. Both analyses yielded non-significant results. Whether a participant viewed a non-work response (F=1.29, df=1, 125, p>.10) or work response first (F=2.15, df = 1, 122, p>.10), his or her ratings across the two interviews were non-significantly different. Overall, Figure 8 indicates a preference for work responses consistent with hypothesis 1. However, this effect is only found between subjects and is dependent on order of presentation.

Similar analyses were conducted on the data presented in Figure 9. Between subject analyses for response setting indicated non-significant differences for ratings of the second interview (F=2.615, df = 1, 253, p>.10). Therefore, work responses when viewed second were not rated significantly higher than non-work responses. There was however, a significant difference for interview 1 (F=5.87, df = 1, 253, p<.10). When the second interview consisted of a work response it was rated significantly higher than interview 1, and when it contained a non-work response, significantly lower than interview 1. These results are consistent with hypothesis 1.

Within subject analyses for response setting yielded non-significant results for Figure 9. When participants viewed a non-work response second, ratings did not differ significantly from those of the first interview (F=.721, df=1, 115, p>.10), nor did they when a work response was viewed second (F=.00, df=1, 132, p>.10). Nevertheless, the results of the between and within subject analyses for Figure 9 are consistent with hypothesis 1; ratings favor work responses.



Figure 9. Response setting 2 by order interaction for interview ratings

Based on the results of the repeated measure ANCOVA presented in Tables 8-9 and Figures 8-9, hypothesis 1 was supported. Between subject ratings of work responses were .23 standard deviation units (d) higher for the first interview and .22 higher for the second interview, which is a large enough difference to influence hiring decisions.

<u>Hiring Decision</u>. Though not addressed in the hypotheses, all participants were asked which of the two candidates they would hire for the job. The question was a forced-choice, dichotomous item in which participants had to choose either the candidate in interview 1 or 2. The dependent variable hiring decision was coded 0 if the candidate in the first interview was hired and 1 if the candidate in the second interview was hired. The independent variables for response setting were coded 0 for non-work and 1 for work. As can be seen from Table 12, results are consistent with hypothesis 1.

Table 12. Logistic regression results for hiring decision.

	В	Exp(B)
Interview Experience (covariate)	-0.11	0.90
Interviewed Others (covariate)	0.41*	1.50
Confederate	-0.31	0.74
Setting 1 <sup>a</sup>	-1.22*	3.40
Setting 2 <sup>b</sup>	1.13*	0.32
Constant	-0.29	0.75
-2 log likelihood: 310.85 $\chi$	<sup>2</sup> Block: 35.17*	$\chi^2$ Model: 38.74*

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed.

<sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed.

\*p<.05

The negative B weight for setting 1 indicates that the candidate in the second interview was more likely to be hired when a non-work response was given in the first interview. Likewise, the positive B weight for setting 2 indicates that the candidate in the first interview was more likely to be hired when a non-work response was given in the second interview. Interaction terms for setting 1, setting 2, and confederate were also examined. A significant interaction between settings 1 and 2 favoring work responses would indicate that raters, when presented with work and non-work responses, were more likely to hire candidates providing work responses; none of the two or three-way interactions were significant.

It was interesting to note however, that when asked to chose between a candidate giving work responses and one giving non-work response (i.e., contrasting response settings were paired), raters were three times more likely to hire the candidate who gave the work response (91/31). Although not without qualifiers, the results of the hiring decision analysis lend further support to hypothesis 1 and the influence of work responses on interview ratings.

<u>Hypothesis 2.</u> Hypothesis 2 stated: Work responses will yield greater halo error as demonstrated by lower variance in ratings across the ten rating dimensions. This hypothesis was tested using repeated measures ANCOVA; the analyses were presented previously in Tables 10-11. The confederate and order effects required the inclusion of the confederate and order variables in the analyses. As a result, it was not appropriate to examine only the main effects of response setting on halo error. However, if work responses yielded greater halo error than non-work responses despite confederate and order effects hypothesis 2 would be supported. It is important to reiterate that the focus

of the analyses is on the *relative* variance across ratings for non-work versus work responses, as variance alone may reflect true relationships in the interview dimensions. Results indicated a three-way interaction for response setting of interview 1, interview 2, and order. Marginal means are graphed in Figures 10 and 11.



Figure 10. Response setting by order interaction for halo error

Figure 10 depicts two virtually parallel lines indicating that non-work responses invoke greater halo error than work responses, contrary to hypothesis 2. These results are similar to those found for the three-way, between subject interaction for response setting 1, response setting 2, and confederate, which were described previously when analyzing confederate and order effects for halo error. Univariate ANCOVA analyses of the between subject effects of response setting indicated that the mean differences in halo error between work and non-work responses for interview 1 (F=.255, df=1, 135, p>.10) and interview 2 (F=1.59, df=1, 135, p>.10) were non-significant. Likewise, the within subject repeated measures ANCOVA indicated non-significant differences in halo for work-work pairs (F = .691, df=1, 68, p>.10) and non-work-non-work pairs (F=1.50, df=1, 60, p>.10). Therefore, the mean differences in halo depicted in Figure 10 are non-significantly different from one another.



# Figure 11. Response settings by order interaction for halo error

The ">" shaped graph depicted in Figure 11 was tested for significant differences in halo error. Between subjects univariate ANCOVA analysis of this difference (across response setting) indicated that it was non-significant for interview 1 (F=1.00, df=1, 118, p>.10) as well as interview 2 (F=.121, df=1, 118, p>.10). The within subject differences in halo error for interview 1 and 2 were also examined. Halo error for the non-workwork pairs (F=.013, df=1, 61, p>.10) and work-non-work pairs (F=.16, df=1,51, p>.10) were non-significant. Therefore, the mean differences in halo graphed in Figure 11, while consistent with hypothesis 2, were non-significantly different from one another. Overall, the results for hypothesis 2 indicate an order effect that is being offset by work-non-work interview pairs. For all but one of the interview pairs, halo error is non-significantly greater for the second interview. However, non-work responses when viewed after work responses limit the increase in halo error resulting from order effects – hence the significant interaction with order. This finding provides some clarity on the significant three-way, between subject interaction found for response setting 1, 2, and confederate –which seemed to contradict hypothesis 2. When halo error is examined for each interview separately, differences in halo error are non-significant. It is only when ratings are collapsed across order that differences reach a significant level. Therefore, it does not appear from these data that non-work responses invoke greater halo error as the between subject analyses suggested. However, the results only partially support hypothesis 2. While non-work responses limit halo error when viewed after work responses, the data do not indicate that work responses invoke significantly greater halo error or that non-work responses invoke significantly greater halo error or that non-work responses invoke significantly greater halo

<u>Hypothesis 3.</u> Hypothesis 3 stated: the influence of rater bias on ratings will be mediated through raters' ideal candidate prototypes. This hypothesis was tested using hierarchical regression. Support for hypothesis 3 would require finding 1) a significant relationship between rater bias and ideal candidate prototypes and 2) a significant interaction between ideal candidate prototypes and response setting on interview ratings. As two ratings were collected (one for each interview), separate analyses were conducted for each.

The first step was to regress the rater bias measure onto the ideal candidate prototype measure (i.e., adjective checklist ratings) for each interview controlling for

interview experience. Results were non-significant (Table 13); the two measures were not related to one another. Therefore, hypothesis 3 could not be supported.

An alternative model was tested in which rater bias moderated the relationship between response setting and interview ratings. To test this model, ratings for interviews 1 and 2 respectively, were regressed onto the rater bias measure, confederate variable, and response setting variable in step 1, the three 2-way interactions in step 2, and the 3way interaction in step 3. Results for the rater bias measure and its interactions were also non-significant (Table 14).

A final model was tested in which interview ratings were regressed onto the rater bias measure, the confederate variable, the two work setting variables, and all of the 2way, 3-way, and 4-way interaction terms. In this manner, it would be possible to see if rater bias moderated ratings when work responses were paired with non-work responses. Results were again non-significant (Table 15). According to these analyses, raters' biases towards work experience did not relate to their stereotypes of candidates or moderate the ratings of interview responses. Hypothesis 3 was not supported nor were the alternative models tested.

	Adjective checklist ratings	Adjective checklist ratings		
	for Interview 1	for Interview 2		
	Beta	Beta		
Interview experience	-0.08	-0.14*		
Interviewed others	-0.13*	0.01		
	$R^2 = .027*$	$R^2 = .019$		
Interview experience	-0.08	-0.14*		
Interviewed others	-0.12	0.01		
Rater Bias	0.09	-0.03		
	$R^2 = .035 \qquad \triangle R^2 = .008$	$R^2 = .02 \qquad \triangle R^2 = .001$		

# Table 13. Regression results for Hypothesis 3 mediated model

\*p<.05

•

.
Interview 1		Interview 2		
	Beta	L	Beta	
Interview Experience	-0.02	2 Interview Experience	-0.09	
Interviewed Others	-0.0	Interviewed Others	0.06	
R <sup>2</sup> =.001	$\Delta R^2 = .001$	$R^2 = .009$	$\Delta R^2 = .009$	
Interview Experience	-0.0	3 Interview Experience	-0.07	
Interviewed Others	-0.02	2 Interviewed Others	0.07	
Rater Bias	-0.0	Rater Bias	0.03	
Confederate	-0.09	O Confederate	0.25*	
Setting 1 <sup>a</sup>	0.12	Setting 2 <sup>b</sup>	0.11	
$R^2 = .024$	$\Delta R^2 = .023$	$R^2 = .082$	$\Delta R^{2} = .073*$	
Interview Experience	-0.03	3 Interview Experience	-0.08	
Interviewed Others	-0.02	2 Interviewed Others	0.08	
Rater Bias	0.14	Rater Bias	-0.06	
Confederate	0.40	Confederate	0.15	
Setting 1	0.48	Setting 2	0.13	
Confederate*Setting 1	-0.12	2 Confederate*Setting 2	-0.24*	
Rater Bias*Setting 1	-0.30	0 Rater Bias*Setting 2	0.12	
Rater Bias*Confederate	-0.4	3 Rater Bias*Confederate	0.26	
$R^2 = .04$	$\Delta R^2 = .016$	$R^2 = .103$	$\Delta R^2 = .021$	
Interview 1		Interview 2		
Interview Experience	-0.0.	3 Interview Experience	-0.08	
Interviewed Others	-0.02	2 Interviewed Others	0.08	
Rater Bias	0.18	Rater Bias	0.03	
Confederate	0.60	Confederate	0.62	
Setting 1	0.69	Setting 2	0.60	
Confederate*Setting 1	-0.4	3 Confederate*Setting 2	-1.07	
Rater Bias*Setting 1	-0.52	2 Rater Bias*Setting 2	-0.36	
Rater Bias*Confederate	-0.6	5 Rater Bias*Confederate	-0.24	
Rater Bias*Confederate*S	etting 1 0.38	Rater Bias*Confederate*Se	etting 2 0.85	
$R^2 = .042$	$\Delta R^2 = .002$	$R^2 = .111$	$\Delta R^2 = .009$	

Table 14. Regression results for moderated model using rater bias

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed. <sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed. \*p<.05

<u>Hypothesis 4.</u> Hypothesis 4 stated: ideal candidate stereotypes will interact with responses such that mean differences favoring work responses will be greater as stereotypes increasingly favor work responses. This hypothesis was tested separately for each interview using moderated regression controlling for interview experience. Interview ratings were regressed onto adjective checklist ratings, the confederate variable, and the response setting variable in step 1, the three 2-way interactions in step 2, and the 3-way interaction in step 3. A significant interaction between response setting and adjective checklist ratings in which higher checklist ratings combined with work responses yielded higher interview ratings would support hypothesis 4.

Results for the interaction terms were non-significant for both interviews (Table 16); adjective checklist ratings did not moderate the relationship between response setting and interview ratings as hypothesized. Therefore, hypothesis 4 was not supported. However, the main effect for adjective checklist ratings was quite strong for both interviews (beta = .49 for interview 1 and .68 for interview 2).

	Interview 1	Interview 2	
	Beta	Beta	
Interview Experience	-0.02	-0.09 0.06	
Interviewed Others	-0.01		
	$R^2 = .001 \ \Delta R^2 = .001$	$R^2 = .009 \ \Delta R^2 = .009$	
Interview Experience	-0.04	-0.06	
Interviewed Others	-0.02	0.07	
Rater Bias	-0.02	0.03	
Confederate	-0.09	0.24*	
Setting 1 <sup>a</sup>	0.13*	-0.13*	
Setting 2 <sup>b</sup>	-0.16*	0.12	
	$R^2 = .05 \ \Delta R^2 = .049^*$	$R^2 = .10 \ \triangle R^2 = .09^*$	
Interview Experience	-0.05	-0.07	
Interviewed Others	-0.02	0.08	
Rater Bias	0.19	-0.07	
Confederate	0.51	0.11	
Setting 1	0.45	-0.17	
Setting 2	0.21	0.09	
Confederate*Setting 1	-0.10	-0.02	
Confederate*Setting 2	-0.14	-0.22*	
Rater Bias*Setting 2	-0.30	0.17	
Rater Bias*Setting 1	-0.26	0.05	
Rater Bias*Confederate	-0.49	0.29	
	$R^2 = .076 \Delta R^2 = .026$	$R^2 = .12  \Delta R^2 = .02$	
Interview Experience	-0.05	-0.06	
Interviewed Others	-0.02	0.09	
Rater Bias	0.33*	0.10	
Confederate	1.23*	0.98	
Setting 1	0.72	0.23	
Setting 2	0.68	0.59	
Confederate*Setting 1	-0.54	-0.69	
Confederate*Setting 2	-0.97	-1.09*	
Rater Bias*Setting 2	-0.79	-0.34	
Rater Bias*Setting 1	-0.53	-0.36	
Rater Bias*Confederate	-1.24*	-0.61	
Rater Bias*Setting 2*Confederate	0.84	0.87	
Rater Bias*Setting 1*Confederate	0.45	0.69	
	$R^2 = .086  \triangle R^2 = .01$	$R^2 = .134 \ \triangle R^2 = .014$	

## Table 15. Regression results for rater bias measure

Table 15. (cont'd)

	Interview 1	Interview 2	
	Beta	Beta	
Interview Experience	-0.05	-0.07	
Interviewed Others	-0.02	0.09	
Rater Bias	0.33*	0.10	
Confederate	1.23*	0.98	
Setting 1	0.72	0.23	
Setting 2	0.68	0.59	
Confederate*Setting 1	-0.51	-0.64	
Confederate*Setting 2	-1.00	-1.15*	
Rater Bias*Setting 2	-0.79	-0.34	
Rater Bias*Setting 1	-0.53	-0.36	
Rater Bias*Confederate	-1.27*	-0.66	
Rater Bias*Setting 2*Confederate	0.92	1.04	
Rater Bias*Setting 1*Confederate	0.49	0.75	
Rater Bias*Setting 1*Setting 2* Confederate	-0.09	-0.18	
	$R^2 = .089  \triangle R^2 = .002$	$R^2 = .142 \Delta R^2 = .009$	

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed. <sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed. \*p<.05

A mediated model was tested to see if candidate stereotypes, as assessed with the checklist measures, might be mediating the effects of response setting. Significant findings for this model would indicate that response setting invoked a stereotype and through the stereotype influenced interview ratings. Results of the regression analyses were not significant; response setting was not significantly related to checklist ratings and therefore could not mediate the effects of response setting on interview ratings. Hypothesis 4 was not supported.

<u>Hypothesis 5.</u> Hypothesis 5 stated: ideal candidate prototypes will interact with response setting such that halo error will be greater for work responses as the prototype increasingly favors work responses. This hypothesis was tested separately for each interview using moderated regression controlling for interview experience. The standard deviation of interview ratings (i.e., halo error) was regressed onto adjective checklist ratings, the confederate variable, and the response setting variable for each interview in step 1, the three 2-way interactions in step 2, and the 3-way interaction in step 3. A significant interaction between work setting and adjective checklist ratings in which high adjective checklist scores combined with work responses yield lower variance in ratings would support hypothesis 5.

Results for both interviews were non-significant (Table 17). Adjective checklist ratings did not significantly interact with response setting as hypothesized nor did they have a significant main effect on the standard deviation of interview ratings (i.e., halo error). Therefore, hypothesis 5 was not supported.

Interview		Interview 2		
	Beta		Beta	
Interview Experience	-0.03	Interview Experience	-0.08	
Interviewed Others	-0.01 Interviewed Others		0.06	
$R^2 = .001$	$\Delta R^2 = .001$	$R^2 = .007$	$\Delta R^2 = .007$	
Interview Experience	0.01	Interview Experience	0.02	
Interviewed Others	0.05	Interviewed Others	0.06	
Setting 1 <sup>a</sup>	0.09	Setting 2 <sup>b</sup>	0.05	
Checklist Rating	0.49*	Checklist Rating	0.68*	
Confederate	-0.16*	Confederate	0.26*	
$R^2$ =.257	$\Delta R^2 = .256^*$	$R^2 = .54$	$\Delta R^2 = .533*$	
Interview Experience	0.00	Interview Experience	0.01	
Interviewed Others	0.04	Interviewed Others	0.06	
Setting 1	-0.54	Setting 2	-0.38	
Checklist Rating	0.43*	Checklist Rating	0.62*	
Confederate	-0.15	Confederate	0.52	
Confederate*Setting 1	-0.07	Confederate*Setting 2	-0.16*	
Checklist Rating*Setting 1	0.68	Checklist Rating*Setting 2	0.54	
Checklist Rating*Confeder	ate 0.04	Checklist Rating*Confederation	te -0.16	
$R^2$ =.262	$\Delta R^2 = .005$	$R^2$ =.553	$\Delta R^2 = .013$	
Interview Experience	0.00	Interview Experience	0.00	
Interviewed Others	0.04	Interviewed Others	0.06	
Setting 1	-1.06	Setting 2	-0.63	
Checklist Rating	0.40*	Checklist Rating	0.60*	
Confederate	-0.55	Confederate	0.27	
Confederate*Setting 1	0.86	Confederate*Setting 2	0.29	
Checklist Rating*Setting 1	1.21	Checklist Rating*Setting 2	0.80	
Checklist Rating*Confederation	ate 0.44	Checklist Rating*Confederat	te 0.10	
Checklist*Confederate*Set	ting 1 -0.94	Checklist*Confederate*Setti	ng 2 -0.45	
$R^2$ =.265	$\Delta R^2 = .003$	$R^2 = .554$	$\Delta R^2 = .001$	

Table 16. Test of moderated model for adjective checklist ratings and interview ratings

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed. <sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed. \*p<.05

Halo error for interview 1		Halo error for interview 2			
		Beta			Beta
Interview Experience	-	-0.02	Interview Experience	-	-0.05
Interviewed Others		-0.06	Interviewed Others		0.02
$R^2 = .005$	$\Delta R^2 = .00$	)5	$R^2 = .003$	$\Delta R^2 = .00$	3
Interview Experience		-0.02	Interview Experience		-0.07
Interviewed Others		-0.06	Interviewed Others		0.02
Setting 1 <sup>a</sup>		0.04	Setting 2 <sup>b</sup>		0.07
Checklist Rating		-0.02	Checklist Rating		-0.09
Confederate		0.09	Confederate		-0.15*
$R^2 = .013$	$\Delta R^2 = .00$	8	$R^2 = .038$	$\Delta R^2 = .033$	5*
Interview Experience		-0.02	Interview Experience		-0.06
Interviewed Others		-0.07	Interviewed Others		0.03
Setting 1		0.37	Setting 2		0.39
Checklist Rating		-0.05	Checklist Rating		-0.10
Confederate		-0.67	Confederate		-0.64
Confederate*Setting 1		0.16	Confederate*Setting 2		0.05
Checklist Rating*Setting 1		-0.44	Checklist Rating*Setting 2		-0.37
Checklist Rating*Confederation	ate	0.67	Checklist Rating*Confederation	ate	0.46
$R^2 = .027$	$\Delta R^2 = .01$	4	$R^2 = .044$	$\Delta R^2 = .00$	6
Interview Experience	·	-0.02	Interview Experience		-0.07
Interviewed Others		-0.07	Interviewed Others		0.03
Setting 1		-0.13	Setting 2		0.07
Checklist Rating		-0.08	Checklist Rating		-0.14
Confederate		-1.04	Confederate		-0.96
Confederate*Setting 1		1.05	Confederate*Setting 2		0.63
Checklist Rating*Setting 1		0.07	Checklist Rating*Setting 2		-0.03
Checklist Rating*Confederation	ate	1.06	Checklist Rating*Confederation	ate	0.80
Checklist*Confederate*Set	ting 1	-0.90	Checklist*Confederate*Set	ting 2	-0.59
$R^2 = .030$	$\Delta R^2 = .00$	)3	$R^2 = .046$	$\Delta R^2 = .00$	2

Table 17. Test of moderated model for adjective checklist ratings and halo error

<sup>a</sup>Setting 1: setting, work or non-work of the first interview viewed. <sup>b</sup>Setting 2: setting, work or non-work of the second interview viewed.

\*p<.05

### DISCUSSION

Results for the study indicate that response setting does play a significant role in how interview responses are rated. Non-work responses are rated significantly lower than work responses and when viewed after a work response can minimize order effects on halo error. However, the non-significant findings for the rater bias measure and adjective checklist ratings make it difficult to determine the cognitive process by which response setting influences ratings. Furthermore, order and confederate effects added to the complexity of the results, making it difficult to draw definitive conclusions regarding the effects of response setting on interview ratings. The following section will address each hypothesis as well as limitations and implications of the findings.

#### Discussion of Hypotheses

Hypothesis 1. Hypothesis 1 proposed a mean difference in interview ratings due to response setting. It was hypothesized that raters would favor work responses and rate them significantly higher than non-work responses. The data supported this hypothesis. When the between subject effects were examined for the first interview, work responses were rated significantly more positively than non-work responses. There were also significant response setting by order interactions in which the second interview was rated significantly lower when the first interview contained a work response and significantly higher when the first interview contained a non-work response. Likewise, the second interview was rated significantly higher than the first interview when the second interview contained a work response. These results combined with raters' tendency to hire candidates giving work responses provides compelling evidence that the raters in this study were influenced by response setting and favored work responses. Moreover, the

significant differences were found using trained raters, a structured interview, and anchored rating scales after controlling for actual work experience in job candidates and using virtually identical interview scripts across work and non-work responses.

Nevertheless, the findings are not without some limitations, most notably the presence of the confederate and order effects. The data indicate that the two confederates used in the study were not perceived as being equivalent. In addition, the confederate by order effects indicate that raters compared confederates to one another and gauged their ratings accordingly. Taken in isolation, these results could have little impact on the interpretation of the hypotheses. Furthermore, it is unlikely that confederate effects could ever be eliminated from the ratings given the effects of aural and visual cues associated with interviews (Motowidlo & Burnett, 1995). However, they raise the question of whether the structured interview procedures had their intended effect or were followed by raters.

Raters were trained to focus on the content of interviews and not personal characteristics of the respondent, to rate each interview in isolation without comparing them to one another, and to rate interviews using the anchored scales rather than their own interpretation of effectiveness. Had significant within subject differences been found, it could be argued that even after controlling for individual differences in how raters evaluated interviews, results still favored work responses. However, within subject differences were non-significant. Therefore, it is difficult to determine whether the results are in fact attributable to response setting as opposed to inadequate rater training or idiosyncrasies across raters.

One plausible explanation raised previously is that when similar response settings were paired (i.e., work-work or non-work-non-work), setting was not salient and had a minimal impact on ratings (Feldman, 1981). In these conditions, differences across the confederates or order of presentation may have been more salient and influential than response setting. The data are consistent with this theory, but cannot unequivocally support it. Before drawing any definitive conclusions regarding response setting, it will be necessary to identify how consistently and strongly it affects ratings and to better understand the factors leading raters to favor work responses over non-work responses.

These concerns raise issues for subsequent research. Perhaps future studies could examine the effects of response setting using multiple confederates to randomize differences across the interviews. Raters could also be asked to recall whether respondents provided work or non-work responses to see if response setting is less salient when similar settings are paired. It may also be advisable to use professional interviewers to see if their perceptions and ratings of work versus non-work responses differ from those of college students.

<u>Hypothesis 2.</u> Hypothesis 2 proposed greater halo error for work responses. This hypothesis is a more subtle way of examining whether raters hold stereotypes towards the setting of an interview response. While raters may actively increase their ratings of interviews containing work responses, they may be unaware of their tendency to commit halo errors when rating those interviews. In addition, halo error is indicative of stereotypes rather than true differences in the effectiveness of responses. A rater may assign the same mean rating to two interviews but have different variability in ratings across dimensions.

While it could be argued that the data were consistent with hypothesis 2, the results did not conform to the true intent of the hypothesis. It seemed that order effects typically led raters to commit greater halo error when rating the second interview. However, this trend was not found when work-non-work interview pairs were rated. In these instances, rating variance was higher (though non-significantly so) for non-work responses. It seemed that when following a work response, non-work responses abated the effects of order on halo error. Therefore, there is evidence that response setting has a significant effect on halo error. It is difficult however, to separate the effects of response setting from order of presentation (i.e., whether non-work responses only limit halo error when they follow work responses).

The results for hypothesis 2 are also plagued with confederate effects and contradictory findings. Perhaps most concerning is the between subject, three way interaction for response settings 1, 2, and confederate. When two non-work responses are paired and confederate 2 is viewed first, there is significantly greater halo error for non-work responses, which is opposite to the hypothesized effect. These results are tempered by the non-significant findings when ratings for interview 1 and 2 are considered separately (rather than being collapsed into one dependent variable as they are in the between subject analyses). It is also plausible that the salience of response setting is low when similar settings are paired, allowing the confederate to have a greater impact on ratings. Nevertheless, the findings cast doubt on the validity of hypothesis 2, and make it difficult to attribute significant differences in halo error to response setting rather than confederate or order effects. In addition, the presence of order effects calls into question how effectively the structured interview procedures minimized rating errors. It is

possible that the significant differences are due to insufficient training rather than response setting, as suggested for hypothesis 1.

As with hypothesis 1, subsequent studies could include multiple confederates, assess the salience of response setting by asking participants whether responses involved work or non-work settings, and use professional raters. In addition, subsequent studies could allow raters more time to practice rating interviews. The videotaped interviews were the only "live" interviews that participants rated during the study. The practice responses were presented in text version and only covered two of the 10 interview dimensions. Providing additional rating practice may reduce the order effects found in the data.

<u>Hypothesis 3.</u> Hypothesis 3 was not supported nor were any of the alternative models tested. It is difficult to determine why the rater bias measure was unrelated to adjective checklist or interview ratings, especially given the significant effects for response setting. Psychometrically, the scale had adequate internal consistency and variance. In addition, the scale items directly targeted the relationship between work experience and job performance. One explanation is that the scale may have been too broad to capture the relationship between rater biases towards work responses and interview ratings.

The rater bias questions asked raters to indicate the extent to which they thought work experience related to job performance in general. Perhaps the effects of rater biases for work responses on interview ratings are more specific than the scale could capture. For example, raters may favor work responses because the behaviors seem more relevant to a work setting, they expect individuals with work experience to discuss those

experiences during the interview, or they feel that non-work responses are a sign of poor job performance. There could also be a moderating variable influencing the effects of rater bias on interview ratings. After one data collection session a participant said that he favored the candidate giving non-work responses because that candidate was well rounded, having both work and life experience. Perhaps there is an individual difference that affects how raters respond to work and non-work responses that was not measured in this study.

These alternative explanations are raised to reiterate the need for further research before concluding that rater bias has no effect on interview ratings. Given the significant findings found for response setting, subsequent studies could investigate the various reactions raters have to work and non-work responses and develop scales to measure those specific reactions or the individual differences that influence them. In this manner it may be possible to identify specific rater reactions to work and non-work responses and what effects if any they have on ideal candidate prototypes and interview ratings.

Alternatively, the non-significant results could be attributable to the design of the study. Specifically, applicant experience was held constant and the job was limited to a management position. Both decisions could have attenuated the effects of rater biases on ideal candidate prototypes, response setting, and ratings. Perhaps the similarity in work experiences prevented response setting from invoking rater biases and ideal candidate prototypes. Raters may have seen the two candidates' work experience as being equivalent and therefore considered other factors, like characteristics of the confederates, when evaluating responses. Alternatively, the decision to use a management level job may have diminished the effects of response setting on ratings as managers may be seen

as requiring "non-work" related skill to effectively manage others. A follow-up study could include multiple levels of applicant experience, a wider range of experiences in the resumes (i.e., work, non-work, job relevant, and non-job relevant experiences), and multiple job levels (entry-level, management, upper management) to see if candidate experience and characteristics of the job affect whether rater biases influence ratings of work and non-work responses.

<u>Hypotheses 4 and 5.</u> Hypotheses 4 and 5 were not supported. Although adjective checklist ratings were highly correlated with interview ratings, they did not moderate the relationship between response setting and interview ratings as hypothesized nor did they mediate the relationship between response setting and interview ratings. They also had a non-significant relationship with halo error. Given that the items for the checklist were selected because of their high means and low variances, the descriptive statistics for the checklists were not unreasonable nor were the internal consistencies. In addition, the original list of 99 items was compiled from adjectives often used to describe job candidates. Thus one would expect stereotypic attributions made by raters to influence how they rated interview candidates on the checklist items. However, the data did not support this hypothesis. As with the rater bias measure, there are plausible explanations for the non-significant findings and several avenues for future research.

For example, the checklist and order it was presented in the study were designed to avoid biasing raters. The items were very general to avoid demand characteristics and the checklists were presented at the end of the study to avoid priming effects. The lack of significant findings could be attributable to one or both of these factors. The items may have been too general and thus failed to assess the prototypes held by raters.

Perhaps questions should have asked about candidate's ability to perform the job, handle a work schedule, conform to a professional work environment, etc. These types of items might have more effectively captured the perceptions raters have of candidates giving work or non-work responses.

Conversely, placement of the scale in the data collection process could explain the non-significant findings. The high correlation between interview ratings and checklist ratings could indicate that raters' evaluation of responses directly influenced how they evaluated the checklist items. As a result of the study design, the checklist might have become an indicator of interview performance rather than a measure of ideal candidate prototypes. This explanation would account for the significant correlation with interview ratings but non-significant correlation with response setting. Perhaps if the checklist measure had been administered prior to watching the interviews or shortly after each interview, the checklists may have moderated ratings as hypothesized. A subsequent study could improve upon the scale as outlined above and vary when checklist ratings are collected to see if prototypes are better measured before or just after interviews are watched. Using multiple collection points across participants could also identify priming effects caused by the prototype measure.

### **Study Limitations**

Having discussed the results, their implications, and potential limitations, it is important to also address some broader limitations that may influence the generalizability of this study.

<u>College student sample.</u> Demographic data indicated that virtually all of the participants had some exposure to the working world and many had interview experience. In addition, prior interview research would support the use of a college student sample given the cognitive processes that were examined in this study (Arvey & Campion, 1982; Bernstein et al., 1975; Dipboye, Fromkin, & Wibach, 1975; McGovern, Jones, & Morris, 1979). Nevertheless, there are reasons to conclude that college students are not equivalent to professional interviewers and not the most appropriate sample to test the proposed hypotheses, especially given the non-significant results of this study. For example, rater biases and ideal candidate prototypes are a critical component of the proposed model. Perhaps college students lack sufficient interview or work experience to have biases towards work responses or to have well developed prototypes of their ideal candidate.

The adjective checklists and rater bias measures were included to assess whether college students do in fact favor work experience or hold ideal candidate prototypes. Unfortunately, the non-significant correlations among measures made it difficult to explain the influence of response setting on interview ratings. As a result, it cannot be determined whether the student sample was inappropriate or the underlying theory was flawed. Future research could examine the hypotheses using professional interviewers to see if the theory holds when experienced raters are used.

Laboratory setting and study design. The laboratory setting and study design also pose some potential problems for the generalizability of the findings. Numerous steps were taken to maintain realism and promote generalizability. A structured interview, anchored rating scales, and rater training from an actual organization were used; the job

description was accurate; and structured procedures were employed throughout the interviews. However, the significant confederate and order effects as well as non-significant findings could indicate that the structuring techniques used in the study were insufficient, that participants did not follow the procedures, or that design of the study prevented finding any effects.

For example, the training was brief (i.e., 20 minutes) and ratings were made at the individual level with no follow-up questions, panel discussion, or consensus rating process. In addition, the interviews lasted only 15 minutes, were watched back to back, and may have seemed contrived due to the response setting manipulation. Proximity of presentation combined with brief rater training could have enhanced order and confederate effects as well as differences in interview ratings due to response setting. Applicant experience level as well as the job were also held constant and may have attenuated the results. Perhaps subsequent studies could devote more time to rater training and practice, include varying levels of applicant work experience in the resumes, examine different job levels, assess or even increase the salience of response setting for work and non-work responses, and use professional raters. Such changes may yield results consistent with the proposed hypotheses or at least provide more insight on the relationship between responses setting, rater biases, ideal candidate prototypes, and interview ratings.

Effect size. The results should also be evaluated for practical significance. Mean differences across response settings were small (d=.22) despite taking numerous steps to enhance the effect of response setting on ratings. Interviews contained all work or non-work responses, the job was a management position, and the training failed to address

how to rate work and non-work responses. In a field setting where experienced raters are used, candidates provide both work and non-work responses in the same interview, jobs include entry-level as well as management positions, and more time is dedicated to training, the effects of response setting may be non-significant. Alternatively, the choice to hold work experience constant and use college student raters (who may not poses biases or prototypes that favor work responses) may have attenuated the findings. Nevertheless, the effect size found in this study was small.

Base rate. Related to effect size is the issue of base rate. It is not known how often applicants actually give all work or non-work responses or what effect combining work and non-work responses will have on ratings. Effects might counteract one another, become more pronounced, or switch direction (e.g., candidates are seen as being more well-rounded). Subsequent studies could vary the number of work and non-work responses provided in the same interview to examine what effects these combinations have on ratings. It is also possible that applicants to the same job (entry level versus management) provide a similar mix of work and non-work responses thus making the effects of response setting on selection decisions a constant. Future inquiries could examine the frequency of work and non-work responses in the field to determine the practical significance of response setting on ratings and selection decisions.

<u>Validity.</u> A critical factor that this study did not consider is the validity of ratings. Quinones et al. (1995) did find significant relationships between work experience and job performance. The current study attempted to address this issue by holding actual experience constant and by making responses virtually identical across response settings. However, it cannot be determined from the current data whether differences in ratings are

accurate assessments of responses and a valid indicator of future job performance. Perhaps managing a baseball team is less predictive of job performance than managing a production team. It will be important in subsequent research to determine whether differences due to response setting are rater error or valid indicators of job performance.

### Future Research

These potential limitations notwithstanding, the data from this study do indicate significant effects for response setting. In addition, the study conditions were not unlike many interview scenarios. Informal or first stage interviews used to screen applicants as well as daylong, back-to-back interview sessions share many of the characteristics of the current study. Raters may have minimal training, be asked to make independent judgments quickly, and use unstructured or partially structured procedures. Likewise, applicants who are just entering the workforce or re-entering after many years may not be able to provide examples from a work environment, and therefore discuss only non-work examples.

Future research could investigate some of the questions raised in this study. For example, raters could be asked to indicate the setting of responses to determine saliency. Raters could also be asked to recall responses to see if they more actively attend to nonwork responses than work responses and have more accurate recall as Feldman's and Fiske's theories would suggest. Professional raters could be used to see if their biases and ideal candidate prototypes are more pronounced and influential than those of college student raters. Raters could also be given more time to practice their ratings before evaluating responses. In addition, the measure of rater biases and ideal candidate prototypes could be refined to see if more precise instruments would yield results in the

hypothesized direction. Likewise, the study conditions and design could be varied to see if placement of the scales, applicant experience, level of the job, or characteristics of the work and non-work responses influenced the results.

More broadly, subsequent studies could examine the effects of mixing work and non-work responses in the same interview, providing training on how to evaluate work and non-work responses, and using panel interviews with consensus ratings. It is also recommended that the relationship between demographic characteristics and work-nonwork responses be examined. Specifically, future studies could examine whether age, gender, or race are related to the number of work/non-work responses provided in an interview and whether certain applicant groups are rated significantly lower as a result.

This study demonstrates that response setting affects interview ratings. However, it provides insufficient information on the process by which this influence might occur and has limited generalizability to professional raters in a field setting. Rater biases that favor work experience and influence ideal candidate prototypes continues to be a plausible explanation for the significant mean difference in interview ratings. Nevertheless, it will be important to study *why* raters in this study evaluated work responses more positively and whether modifying characteristics of the study or sample will yield results that are more consistent with the proposed theory and hypotheses.

### Conclusion

Perhaps the most obvious conclusion of this study is that further research is needed to determine the unique effects of response setting on interview ratings and the influence of ideal candidate prototypes and rater biases towards work responses. The various non-significant or contrary findings as well as confederate and order effects make

it difficult to determine whether problems exist with the underlying theory presented in this paper, the design of the measures or study, or appropriateness of a college student sample for studying this question. However, the data do provide preliminary evidence that response setting significantly influences interview ratings. The prevalent use of the structured interview combined with the goal of minimizing rating errors makes response setting a worthy topic of future research.

## REFERENCES

.

### REFERENCES

Arvey, R. D., & Campion, J. E. (1982). The employment interview: Legal and psychological aspects. <u>Psychological Bulletin, 86</u>, 736-765.

Bernstein, V., Hakel, M. D., Harlan, A. (1975). The college students as interviewer: A threat to generalizability? Journal of Applied Psychology, 60(2), 266-268.

Binning, J. F., & Barrett, G. V. (1989). Validity of Personnel Decisions: A conceptual analysis of the inferential and evidential bases. <u>Journal of Applied</u> <u>Psychology, 74(3),</u> 478-494.

Campion, M. A., Campion, J. E., & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. <u>Journal of Applied</u> <u>Psychology</u>, 79(6), 998-1002.

Campion, M. A., Palmer, D. K., Campion, J. E. (1997). A review of structure in the selection interview. <u>Personnel Psychlogy</u>, 50(3), 655-702.

Cash, T. F., Gillen, B., & Burns, D. S. (1977). Sexism and "beautyism" in personnel consulting decision making. Journal of Applied Psychology, 62, 301-307.

Cohen, J. (1988). <u>Statistical power analysis for the behavioral sciences.</u> New Jersey: Lawrence Earlbaum Associates. pg 317.

Cohen, S. L., & Bunker, K. A. (1975). Subtle effects of sex role stereotypes on recruiters' hiring decisions. Journal of Applied Psychology, 60, 566-572.

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. Journal of Applied Psychology, 80(5), 565-579.

Dipboye, R. L. (1982). Self-fulfilling prophecies in the selection-recruitment interview. <u>Academy of Management Review</u>, 7(4), 579-586.

Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisals. <u>Academy of Management Review</u>, 10(1), 116-127.

Dipboye, R. L., & Fromkin, H. L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 60, 39-43.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. Journal of Applied Psychology, 71(1), 9-15.

Eber, R. & Fiske, S. T. (1984). Outcome dependency and attention to inconsistent information. Journal of Personality and Social Psychology, 41, 709-726.

Eder, R. W., & Harris, M. M. (1999). Employment interview research: Historical update and introduction. In R. W. Eder, & M. M. Harris (Eds.), <u>The employment</u> interview handbook, 1-27, Newbury Park, CA: Sage.

Favero, J. L., & Ilgen, D. R. (1989). The effects of ratee prototypicality on rater observation and accuracy. Journal of Applied Social Psychology, 19, 932-946.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66(2), 127-148.

Fiske, S. T. (1982). Schema-triggered affects: Applications to social perceptions. In Clark, M. S., & Fiske, S. T. (Eds.), <u>Affect and cognition: The 17<sup>th</sup> annual Carnegie</u> <u>symposium.</u> (pp. 55-78). Hillsdale, NJ: Erlbaum.

Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). categorybased and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. Journal of Experimental Social Psychology, 23, 399-427.

Fiske, S. T., & Pavelchak, M. A., (1986). Category-based versus piecemeal-based affective responses: Developments in schema-triggered affect. In Sorrentino, R. M., & Higgins, E. T. (Eds.), <u>Handbook of motivation and cognition: Foundations of social behavior.</u> (pp. 167-203). New York: Guilford Press.

Graves, L. M., & Karren, R. J. (1992). Interviewer decision process and effectiveness: An experimental policy-capturing investigation. <u>Personnel Psychology</u>, 45, 313-340.

Guion, R. M. (1987). Changing views for personnel selection research. <u>Personnel</u> <u>Psychology, 40,</u> 199-213.

Hakel, M. D., Dobmeyer, T. W., & Dunnette, M. D. (1970). Relative importance of three content dimensions in overall suitability ratings of job applicant's resumes. Journal of Applied Psychology, 54, 65-71.

Hakel, M. D., Hollmann, T. D., & Dunnette, M. D. (1970). Accuracy of interviewers, certified public accountants, and students in identifying the interests of accountants. Journal of Applied Psychology, 54(2), 115-119.

Hakel, M. D., Ohnesorge, J. P., & Dunnette, M. D. (1970). Interviewer evaluations of job applicants' résumés as a function of the qualifications of the immediately preceding applicants: An examination of contrast effects. Journal of Applied <u>Psychology</u>, 54, 27-30. Hollman, T. D. (1972). Employment interviewer's errors in processing positive and negative information. Journal of Applied Psychology, 56, 130-134.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. Journal of Applied Psychology, 79(2), 184-190.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. Journal of Applied Psychology, 67(5), 577-580.

Janz, T., Hellervik, L., & Gilmore, D. C. (1986). <u>Behavior description</u> interviewing: New, accurate, cost effective. Boston: Allyn and Bacon.

Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. In R. W. Eder, & G. R. Ferris (Eds.), <u>The employment interview:</u> <u>Theory, research, and practice.</u> 158-182. Newbury Park, CA: Sage.

Latham, G. P., & Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview. Journal of Applied Psychology, 69, 569-573.

Latham, G. P., & Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. Journal of Applied Psychology, 65, 422-427.

Macan, T. H., & Dipboye, R. L. (1990). The relationship of interviewers' preinterview impressions to selection and recruitment outcomes. <u>Personnel Psychology</u>, <u>43</u>, 745-768.

Mayfield, E. C. (1964). The selection interview: A reevaluation of published research. <u>Personnel Psychology</u>, 17, 239-260.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Mauer, S. D., (1994). The validity of employment interviews: A review and meta-analysis. Journal of Applied Psychology, 79(4), 599-616.

McGovern, T. V., Jones, B. W., & Morris, S. E. (1979). Comparison of professional versus student ratings of job interviewee behavior. <u>Journal of Counseling</u> <u>Psychology, 26</u>, 176-179.

Mischel, W. (1968). Personality and Assessment. New York: Wiley.

Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), <u>Personality at the Crossroads: Current Issues in Interactional</u> <u>Psychology.</u> Hillsdale, NJ: Erlbaum.

Motowidlo, S. J. (1986). Information processing in personnel selection. In Rowland, K. M., & Ferris, G. R. (Eds.), <u>Research in personnel and human resource</u> management (pp. 1-44). Greenwich, CT: JAI.

Motowidlo, S. J., & Burnett, J. R., (1995). Aural and visual sources of validity in structured employment interviews. <u>Organizational Behavior and Human Decision</u> <u>Process, 61(3)</u>, 239-249.

Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., Vaughan, M. J. (1992). Studies of the structured behavioral interview. Journal of Applied Psychology, 77(5), 571-587.

Murtha, T. C., Kanfer, R., & Ackerman, P. L. (1996). Toward an interactionist taxonomy of personality and situations: An integrative situational-dispositional representation of personality traits. Journal of Personality and Social Psychology, 71(1), 193-207.

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influence on impression formation: Outcome dependency, accuracy-driven attention, and individual processing. Journal of Personality and Social Psychology, 53(3), 431-444.

Orpen, C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. <u>Journal of Applied Psychology</u>, 70, 774-776.

Pervin, L. A., (1989). Person, situations, interactions: The history of a controversy and a discussion of theoretical models. Academy of Management Review, 14(3), 350-360.

Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. <u>Personnel Psychology</u>, 48, 289-308.

Quinones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. <u>Personnel</u> <u>Psychology</u>, 48, 887-910

Rowe, P. M. (1984). Decision process in personnel selection. <u>Canadian Journal of</u> <u>Behavioral Science, 16</u>, 326-337.

Rowe, P. M. (1989). Unfavorable information and interview decisions. In Eder, R. W., & Ferris, G. R. (Eds.), <u>The employment interview: Theory, research, and practice</u> (pp. 77-89). Newbury Park, CA: Sage.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., and Powell, A. B. (1995). Frame of reference effects on personality scale scores and criterion related validity. <u>Journal of Applied Psychology</u>, 80(5), 607-620.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. <u>Personnel Psychology</u>, 29, 79-101.

Shaw, S. E. (1972). Differential impact of negative stereotyping in employee selection. <u>Personnel Selection, 25</u>, 333-338.

Stewart, G. (1996), Reward structure as a moderator of the relationship between extraversion and sales performance. Journal of Applied Psychology, 81(6), 619-627.

Ulrich, L., & Trombo, D. (1965). The selection interview since 1949. <u>Psychological Bulletin, 63, 100-116</u>.

Valenzi, E. & Andrews, I. R. (1973). Individual differences in the decision process of employment interviewers. Journal of Applied Psychology, 58, 49-53.

Wagner, R. (1949). The employment interview: A critical summary. <u>Personnel</u> <u>Psychology, 2</u>, 17-46.

Webster, E. C. (1964). <u>Decision making in the employment interview</u>. Montreal: McGill University.

Weekley., J. A., & Gier, J. A. (1987). Reliability and validity of the situational interview for a sales position. Journal of Applied Psychology, 72, 484-487.

Weiss, H. M. & Adler, S. (1984). Personality and organizational behavior. In B. M. Staw & L. L. Cummings (Eds.), <u>Research in Organizational Behavior</u>. Greenwich, Connecticut: JAI Press.

Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. Journal of Occupational Psychology, 61, 275-290.

Wright, O. R., (1969). Summary of research on the selection interview since 1964. <u>Personnel Psychology, 22, 391-413</u>.

Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. Journal of Occupational Psychology, 62, 191-199.

Zedeck, S., Tziner, A., Middlestadt, S. E. (1983). Interviewer validity and reliability: An individual analysis approach. <u>Personnel Psychology</u>, 36, 355-370.

APPENDICES

.

.

# APPENDIX A

Power Analysis

## APPENDIX A

## Power Analysis

A power analysis was conducted to determine the number of subjects needed to attain 0.80 power at the 0.05 level. Hypotheses 4 & 5 propose an interaction between prototypes and responses and will therefore require the most number of subjects to test. Assuming a small effect size (f=.10), 240 subjects will be needed to attain 0.80 power (Cohen, 1988).

# APPENDIX B

Training Manual

•

APPENDIX B

**Training Manual** 

# **Rater Training Manual**

**Team Leader Selection** 

# **Introduction**

Today you will be acting as the director of personnel in an auto parts manufacturing plant. Your task is to hire a new employee for the company. It will be your responsibility to view the interview responses of two applicants and rate their qualifications for the position.

## The Job:

You are hiring a Team Leader to manage a production team of 7-10 people. Team leaders are the first level of management in your company. Job responsibilities include:

- Directing team members
- Resolving team member conflicts
- Training and developing team members
- Managing production and employee schedules

## **Training Manual:**

This training manual contains information on the Team Leader selection interview. The interview was designed to elicit information related to the team leader position. It asks all applicants the same ten questions related to critical demands of the job. This training manual will review:

- 1. The interview
  - interview dimensions
  - interview questions
  - interview rating materials
- 2. The rating process
  - taking notes
  - evaluating responses

## The Interview

## **Interview Dimensions**

The interview will assess capabilities in the following areas:

- 1. Problem Solving / Trouble Shooting
- 2. Leading and Influencing
- 3. Interpersonal Management
- 4. Train, Develop, and Evaluate Employees
- 5. Flexibility/Adaptability/Sensitivity
- 6. Performance Improvement Orientation
- 7. Quality Orientation
- 8. Safety/Cleanliness Orientation
- 9. Planning/Scheduling/Organizing

## 10. Communication

Please take a moment to review each dimension and definition before continuing to the next section of this manual.

# NOTE TO READER: DEFINITIONS PROPRIATERY SEE AUTHOR FOR MATERIALS

## **Interview Questions & Rating Materials**

The interview contains one question for each of the ten dimensions. Each question asks about past experiences that might demonstrate job related abilities. You will be using ten anchored rating scales to evaluate responses. The evaluation forms are organized like the example shown below:

Dimension Title					
Dimension Definition					
	Low		Moderate		High
•	Description of low scoring applicant	•	Description of moderate scoring applicant	•	Description of high scoring applicant
	0	0	3	4	\$

1. Interview Question

- In the top row is the Dimension Title.
- The second row contains the Dimension Definition.
- The bottom 3 columns contain descriptions of Low, Moderate, and High scoring applicants.
- The Interview Question is listed below the rating scale.

Take a few moments to thoroughly review the interview questions and rating scales.

# NOTE TO READER: QUESTIONS AND RATING SCALES PROPRIATERY SEE AUTHOR FOR MATERIALS
# Making Ratings

The rating process is critical to selecting the best applicant. This section will review how to 1) **take notes** and 2) **evaluate responses**.

# Note Taking

- As an interviewer, you must take notes on every response provided by candidates.
- Do <u>not</u> write down every word; record key words, critical facts, and main points.
- Your notes should contain three pieces of information for each response:
  - 1. The Situation: a description of the context or background for the event
  - 2. The Action: a description of the applicant's behavior in the situation
  - 3. The **Result:** a description of the consequences of the applicant's actions.

# **Example**

An example of a complete (but concise) response to a question for the Interpersonal Management dimension would be:

"My friends liked to play practical jokes on each other. Well one day, my friend Mike played a joke on my other friend John. John got very upset and before I knew it, he and Mike were fighting. I didn't want to see the situation get worse, so, I got in between them and tried to calm them down. Once they cooled down, I got them to talk. Mike apologized and they both ended up laughing about the whole incident."

# Appropriate notes for this response would be:

Situation: Two friends got into a fight

Action: Applicant got in between them. Had them calm down. Had them talk.

**Result:** Friend apologized. Both laughed about situation.

# **Evaluating the Applicant**

After each interview, you will use the scales and your notes to evaluate candidate's responses.

# The evaluation process:

- Begin with dimension 1. Problem Solving & Trouble Shooting.
- Review the anchors for low, medium, and high.
- Review <u>all</u> of your notes from the interview.
- Compare your notes of the applicant's responses to the anchors.
- Select the numerical rating that best describes the applicant.
- Mark your rating on the answer sheet provided.
- Repeat the process for the remaining 9 interview dimensions.

# NOTE:

- 1. Use <u>all</u> relevant information from the interview when making your ratings.
- 2. If an applicant's responses match some of the LOW descriptions *and* some of the MEDIUM descriptions, the applicant should receive a rating of "2". Similarly, if an applicant's responses match some of the HIGH and MEDIUM descriptions, the applicant should receive a rating of "4".

# Tips for making more accurate ratings:

- Evaluate each dimension separately. Everyone has both strengths and weaknesses that should be reflected in your ratings.
- Evaluate the applicant's responses only. Avoid being mislead by personal characteristics or mannerisms that are not relevant to the job.
- **Don't compare one applicant to another.** Each applicant's effectiveness should be determined by comparing the applicant's responses to the standard provided on the rating scales –not by comparing applicants to each other.

# APPENDIX C

Video Transcripts

.

.

# APPENDIX C

#### Video Transcripts

# **DIMENSION 1**

#### Non-work 1:

I was organizing a community meeting to get various associations together to discuss some important issues facing the neighborhood. My concern was that once we had set a date to meet, people would forget or would skip the meeting. To avoid this problem I identified a contact person in each association. A few days before the meeting I called each of them to remind them of the date and time we had scheduled. As a result, we had strong attendance and each neighborhood group was represented.

#### Work 1:

I was organizing a meeting at my old job to get various departments together to discuss some important issues facing the company. My concern was that once we had set a date to meet, people would forget or would skip the meeting. To avoid this problem I identified a contact person in each department. A few days before the meeting I called each of them to remind them of the date and time we had scheduled. As a result, we had strong attendance and each department was represented.

## Non-work 2:

In college I volunteered to be part of a technology committee in my residence hall. We were in charge of ordering new computer equipment for the study areas in the building. I was concerned that students wouldn't provide input and we would order computers no one would like. So we held a hall meeting to decide what to order. I asked for volunteers on each floor to tell people about the meeting. We had a large number of students show up and we were able to get everyone's input before placing the final equipment order.

## Work 2:

At my old job I volunteered to be part of a technology committee in the company. The committee was responsible for ordering new computer equipment for the company. I was concerned that employees wouldn't provide input and we would order computers no one would like. So we held a company meeting to decide what to order. I asked for volunteers from each department to tell employees about the meeting. We had a large number of people show up and we were able to get everyone's input before placing the final equipment order.

### Non-work 1:

One Christmas I volunteered at a local shelter to help serve food to the homeless. The shelter usually has about 20 volunteers to help serve food throughout the day. When I arrived in the morning, five people had called in to say they couldn't make it. Rather than trying to have 15 people do the work of 20, I asked everyone to call one or two friends to see if anyone would be willing to come in and help us. We ended up finding quite a few people to help. I made a new schedule based on when people were available and we were able to maintain a full crew of volunteers throughout the entire day.

### Work 1:

One Christmas I agreed to work the holiday. We usually need 20 people in the department to run effectively. When I arrived in the morning, five people had called in to say they couldn't make it. Rather than trying to have 15 people do the work of 20, I asked everyone to call one or two co-workers to see if anyone would be willing to come in and help us. We ended up finding quite a few people to help. I made a new schedule based on when people were available and we were able to maintain a full crew of workers throughout the entire day.

### Non-work 2:

I volunteer at a youth center in my neighborhood. Every year we get local business leaders involved with the community. We ask business people to come in and speak with the children about important issues. My first year at the center I spent countless hours making phone calls and coordinating speakers. As a result I was really stressed out and I had little time to spend with the children. Rather than repeating that process for another year, I recruited some of the high school kids in the program to help me contact potential speakers. I gave each one of them a list of people to contact. They made the phone calls, scheduled dates, and acted as the coordinator for speakers they had scheduled. We were able to get a number of speakers and the program worked well.

### Work 2:

At my current job we try to get our customers more involved with our employees. We ask our customers to send members of their top management staff to come speak with our employees about important issues. My first year on the job I spent countless hours making phone calls and coordinating speakers. As a result I was really stressed out and I had little time to get my other work done. Rather than repeating that process for another year, I recruited some of the senior employees from the company to help me contact potential speakers. I gave each one of them a list of people to contact. They made the phone calls, scheduled dates, and acted as the coordinator for speakers they had scheduled. We were able to get a number of speakers and the program worked well.

#### Non-work 1:

Well, I had been living in a new apartment for about a week or so when I got a knock on my door from one of my neighbors. He was very angry with me because he said that I had been stealing his newspaper every morning. We both subscribed to the same paper so I wasn't sure what was happening. Rather than fight with him, I told him that I subscribed to the paper and that he should call the circulation department at the newspaper if he wasn't receiving his paper. He never mentioned it again after that.

# Work 1:

Well, I had just started a new job and had been working for about a week or so when an employee from another area confronted me. He was very angry with me because he said that I had been stealing his supplies. We all used the same supplies so I wasn't sure what was happening. Rather than fight with him, I told him that I had order my own supplies and that he should call the supplier if he wasn't receiving his order. He never mentioned it again after that.

# Non-work 2:

I play baseball with a team in my neighborhood. About a year ago we had a new guy join the team. He was pretty good I and had no problems with him. Then one day he started giving me a hard time. He thought I was slacking off and not pulling my weight. I didn't want to get into a fight with him about it so I told him that I was doing the best I could and that he should take his complaints to the coach. He never bothered me about it again after that.

# Work 2:

In my current job, we work pretty close together. About a year ago we had a new guy join the department. He was pretty good I and had no problems with him. Then one day he started giving me a hard time. He thought I was slacking off and not pulling my weight. I didn't want to get into a fight with him about it so I told him that I was doing the best I could and that he should take his complaints to the supervisor. He never bothered me about it again after that.

#### Non-work 1:

I play on a softball team and was asked to train a new outfielder. He was eager to learn but had never played on a team before. So I reviewed the basics with him and we practiced a few drills to familiarize him with key aspects of the game. I spent about half a day working with him and then put him into the game. I've always thought that the best way to learn is to play –experience is the best teacher. It took him some time to get use to the game, but he ended up being a good outfielder.

# Work 1:

I was asked to train a new employee at one of my old jobs. He was eager to learn but had never worked a job before. So I reviewed the basics with him and we practiced a few exercises to familiarize him with key aspects of the job. I spent about half a day working with him and then put him to work. I've always thought that the best way to learn is on the job-experience is the best teacher. It took him some time to get use to the job, but he ended up being a good worker.

# Non-work 2:

The last time I trained someone was in high school. I was on the newspaper staff. It was my job to take handwritten articles and layouts and enter them onto the computer. One day the teacher asked me if I could train a new staff member to use the computer. He had little computer experience so I started with very basic operating information. Then I showed him the basic steps to entering information onto the computer. I spent about 4 hours working with him and then gave him some layouts to enter. It took him some time to get use to the computer and learn the software, but after some practice he became really good at it.

# Work 2:

The last time I trained someone was several years ago at one of my first jobs. We would generate print-outs for the rest of the company. It was my responsibility to take handwritten materials and create the computer files. One day my boss asked me if I could train a new employee to use the computer. He had little computer experience so I started with very basic operating information. Then I showed him the basic steps to entering information onto the computer. I spent about 4 hours working with him and then gave him some layouts to enter. It took him some time to get use to the computer and learn the software, but after some practice he became really good at it.

# Non-work 1:

My roommate and I were in the process of moving to a new apartment. He and I get along really well, but we have very different approaches to things. For example, I wanted to get a ton of boxes and pack everything ahead of time. That way on the day of the move we could just load up all of the boxes and go. I told him my idea but he thought that it was a waste of time. He thought it would be easier to get a few boxes and transport a little bit at a time. He wanted to make a few trips to the new apartment each day and that way it wouldn't seem like such a big ordeal. I thought about it and it seemed like more work. So we decided to move our things separately. He moved a little at a time, and I moved my things all at once. It worked out really well and we didn't fight at all during the process.

## Work 1:

A co-worker and I had been re-assigned to a work area and had to move all of our stuff. He and I get along really well, but we have very different approaches to things. For example, I wanted to get a ton of boxes and pack everything ahead of time. That way on the day of the move we could just load up all of the boxes and go. I told him my idea but he thought that it was a waste of time. He thought it would be easier to get a few boxes and transport a little bit at a time. He wanted to make a few trips each day and that way it wouldn't seem like such a big ordeal. I thought about it and it seemed like more work. So we decided to move our things separately. He moved a little at a time, and I moved my things all at once. It worked out really well and we didn't fight at all during the process.

# Non-work 2:

I volunteer at a youth center. Well space is pretty tight so two volunteers have to share a desk. I get along really well with the other volunteers, but the person I shared a desk with was very messy. He likes to leave all of his things out and doesn't clear off the desk when he's done with it. As a result, I had to clean up after him or find another place to work. I talked to him about it, but it didn't help. So I asked the center coordinator if I could share a desk with someone else. I was re-assigned to someone who's much more organized. We both work together well, and I'm much happier with my new assignment.

## Work 2:

My company is in the middle of re-locating, but until then space is pretty tight. As a result two employees have to share a desk. I get along really well with the other employees in my department, but the person I shared a desk with was very messy. He likes to leave all of his things out and doesn't clear off the desk when he's done with it. As a result, I had to clean up after him or find another place to work. I talked to him about it, but it didn't help. So I asked my boss if I could share a desk with someone else. I was re-assigned to someone who's much more organized. We both work together well, and I'm much happier with my new assignment.

#### Non-work 1:

Last year as web publishing became more popular I wanted to learn how to create a web page on the internet. I asked around but no one I knew had any idea what I was talking about. So I checked with the community college to see if they offered any classes. They had a special 2 day seminar that covered all the basics. I read the course description and the class seemed to be exactly what I was looking for. I took the course and with some practice have become pretty good at creating and posting pages on the web. Now that the web is so popular, it has become a great skill for me to have. I'm really glad I took the course when I did.

## Work 1:

Last year web publishing became really popular at my old company. I decided that it would be a good idea for me to learn how to create a web page on the internet. I asked my boss about it but he had no idea what I was talking about. So I checked with the community college to see if they offered any classes. They had a special 2 day seminar that covered all the basics. I read the course description and the class seemed to be exactly what I was looking for. I took the course and with some practice have become pretty good at creating and posting pages on the web. Now my company publishing almost everything on the web and it has become a great skill for me to have. I'm really glad I took the course when I did.

## Non-work 2:

When I first started on the newspaper staff in high school they were in the process of getting a new computer software system for setting up newspaper layouts. Before then, they had always done everything by hand. I really enjoy working with computers so I asked if I could go to a training class to learn how to use the new software. The teacher thought it was a great idea so he sent me. It wasn't easy, but I learned how to use all of the different software features. As a result I was placed in charge of entering all of the computer work for the paper.

## Work 2:

When I started working at one of my old job, they were in the process of getting a new computer software system for generating reports and print-outs. Before then, they had always done everything by hand. I really enjoy working with computers so I asked if I could go to a training class to learn how to use the new software. My boss thought it was a great idea so he sent me. It wasn't easy, but I learned how to use all of the different software features. As a result I was placed in charge of all of the computer work for the department.

#### Non-work 1:

Several months ago, I was asked to submit a newspaper article on a community project that we had started in my neighborhood. A lot of people had helped on the project and it was important to me that the article was accurate and that everyone was recognized. I talked to as many people as I could to get as much input as possible on the article. Then, before I submitted the final draft, I circulated it to the group so that people could give me feedback. I used the feedback to make changes and submitted the article. When it was printed, there were no errors and everyone was happy with what I had written.

#### Work 1:

Several months ago, I was asked to submit an article on a project we had started in my department for our company wide newsletter. The article was about a special project we had started in our department. A lot of people had helped on the project and it was important to me that the article was accurate and that everyone was recognized. I talked to as many people as I could to get as much input as possible on the article. Then, before I submitted the final draft, I circulated it throughout the department so that people could give me feedback. I used the feedback to make changes and submitted the article. When it was printed, there were no errors and everyone was happy with what I had written.

#### Non-work 2:

About a year ago, I was asked to order supplies for the youth center where I volunteer. Everyone in the center was suppose to tell me what they needed and I would fax in the order. Well, we only made the orders once a month so it was important that it was accurate. Before I placed the order, I asked each of the volunteers to review it and add anything I had forgotten. As a result, everyone received their needed supplies.

## Work 2:

About a year ago, I was asked to order supplies for my department where I work. Everyone in my section was suppose to tell me what they needed and I would fax in the order. Well, we only made the orders once a month so it was important that it was accurate. Before I placed the order, I asked each of the unit supervisors to review it and add anything I had forgotten. As a result, everyone received their needed supplies.

# Non-work 1:

In the process of moving to a new apartment, I realized that I needed to put some things into storage. Well the storage space was two stories off the ground. They had a ladder but I didn't trust it. If I fell or dropped something I could really hurt myself or someone else. So I hired two guys from the storage company to pick up and store my things for me. It took them less than an hour, nothing got broken, and no one got hurt.

# Work 1:

A few years back at one of my old jobs, my supervisor asked me if I could put a few files into storage. Well the storage space was two stories off the ground. They had a ladder but I didn't trust it. If I fell or dropped something I could really hurt myself or someone else. So I hired two guys from the storage company to pick up and store the files for me. It took them less than an hour, nothing got broken, and no one got hurt.

# Non-work 2:

One of the activities in the youth center where I volunteer is to go into the community to visit kids and talk with their parents. The center has a van but it would cost too much to get insurance for all the volunteers to drive it. Well, my car is old and unreliable and I was concerned that it might break down during the winter or in a bad area. So instead of taking any chances, I got permission to have one of the other volunteers conduct my site visits for me while I conducted his activities at the center. The arrangement worked out really well for both of us.

# Work 2:

One of the responsibilities of my job is to visit customer sites to tell them about new ideas and get their feedback. We have company vehicles but it would cost too much to get insurance for all of the employees to drive them. Well, my car is old and unreliable and I was concerned that it might break down during the winter or in a bad area. So instead of taking any chances, I got permission to have another employee conduct my site visits for me while I filled in for him at the company. The arrangement worked out really well for both of us.

# Non-work 1:

A few years ago I had the opportunity to travel overseas to visit a friend of mine. At the time I was involved with the community project so I had to decide how to handle my responsibilities. The trip was a great opportunity for me, so I decided to accept the offer. I handed over my community project responsibilities to the other volunteers and told them where they could reach me if problems arose. The trip was a great experience for me and I'm really glad I went.

# Work 1:

A few years ago I had the opportunity to travel overseas for work. At the time I was involved with a project at the company and had to decide how to handle my responsibilities. The trip was a great opportunity for me, so I decided to accept the offer. I handed over my work responsibilities to the other employees on the project and told them where they could reach me if problems arose. The trip was a great experience for me and I'm really glad I went.

# Non-work 2:

After my first year on the newspaper staff, I was asked to be the photographer for the football team. If I took the assignment I would have to reduce my responsibilities with newspaper for a few months. I knew this was a great opportunity, so I took the position. I had a lot of responsibilities with the newspaper, so I distributed them to the rest of the staff and told them to call me if they had questions. The transition worked out really well for me and I really enjoyed working with the football team.

# Work 2:

After my first year at one of my old jobs, I was asked by one of the managers to work on a project in his department. If I took the assignment I would have to leave my department for a few months. I knew this was a great opportunity, so I took the position. I had a lot of responsibilities in my department, so I distributed the work to the rest of my co-workers and told them to call me if they had questions. The transition worked out really well for me and I really enjoyed the assignment.

#### Non-work 1:

Earlier I had mentioned a community project we had started in my neighborhood. Well, before we could get funding for the project, we had to present the idea to the city planners committee. Since no one else wanted to speak, I volunteered to present the ideas to the committee. It was important that I presented the information clearly. I would also have to be very familiar with the project in case they asked questions. To help with the presentation, I prepared overheads and speaking notes that would guide me through each point of the project. Then the week before the presentation, I gave a practice talk to the rest of the community group. They asked questions as if they were the committee and they gave me feedback on my presentation. I used their comments to make a few changes to the talk. By the day of the city planners meeting I felt comfortable with the speech and was ready to answer questions. The committee liked out proposal and decided to fund the project.

## Work 1:

Earlier I had mentioned a new project we had started in my department. Well, before we could get funding for the project, we had to present the idea to the top management team. Since no one else wanted to speak, I volunteered to present the ideas to the team. It was important that I presented the information clearly. I would also have to be very familiar with the project in case they asked questions. To help with the presentation, I prepared overheads and speaking notes that would guide me through each point of the project. Then the week before the presentation, I gave a practice talk to the rest of the department. They asked questions as if they were top management and they gave me feedback on my presentation. I used their comments to make a few changes to the talk. By the day of the manager's meeting, I felt comfortable with the speech and was ready to answer questions. Management liked our proposal and decided to fund the project.

#### Non-work 2:

While I was a volunteer at the youth center, I was asked to visit another center in Ohio to discuss some of our programs and activities. I was excited at the opportunity, but I was also a little nervous about giving a speech. I didn't want to lecture people for an hour but I also wanted to show the other center what we were doing. I thought about it and I decided to give a slide show. I bought a few roles of film and took pictures during many of our activities. I had the film developed into slides and used them to guide my talk. While the slides were up I would talk about the activity, why we thought it was important, and its impact. The center in Ohio was very impressed with the presentation and it was a great opportunity to share ideas with other people. The slides also made it much easier for me to discuss the various programs we had started in our center.

## Work 2:

While I was working on the continuous learning committee in my department, I was asked to visit company in Ohio to discuss some of our programs and activities. I was excited at the opportunity, but I was also a little nervous about giving a speech. I didn't want to lecture people for an hour but I also wanted to show people what we were doing. I thought about it and I decided to give a slide show. I bought a few roles of film and took pictures during many of our activities. I had the film developed into slides and used them to guide my talk. While the slides were up I would talk about the activity, why we thought it was important, and its impact. The company in Ohio was very impressed with the presentation and it was a great opportunity to share ideas with other people. The slides also made it much easier for me to discuss the various programs we had started in our company.

# APPENDIX D

Resumes

# APPENDIX D

#### Resumes

#### Candidate A Background Information Form

<b>`</b>	Name:	
	Address:	
	Phone #:	
	Age:	26
	Gender:	Male
	Education:	BA, Michigan State University (May 2000)

# Work Experience:

- **1997-present:** General Motors Corporation **Job Description:** automobile assembly
- **1996-1997:** Fairfield & Wells Equipment **Job Description:** parts production
- 1993-1996: Barnes & Noble Job Description: cashier

# Candidate B Background Information Form

Name:	
Address:	
Phone #:	
Age:	26
Gender:	Male

Education: BA, Michigan State University (May 2000)

# Work Experience:

- **1997-present:** Ford Motor Company **Job Description:** automobile assembly
- **1996-1997:** Great Lakes Manufacturing **Job Description:** parts production

.

.

• 1993-1996: Walden Books Job Description: cashier

# APPENDIX E

Measures

.

### APPENDIX E

#### Measures

#### Sample Responses

# Leading & Influencing

Answer: My roommates and I were in the process of moving out of a four bedroom house. We all worked and went to school so we had very little time to pack and clean. Rather than wait until the last minute, I made a list of what needed to be done and assigned each person a different set of tasks. Then each day I would leave them reminders about what still needed to be done. Eventually, we were able to pack everything up and clean the house in time for the move.

Rating:\_\_\_\_\_

# Train, Develop, and Evaluate Employees

Answer: We had just hired a new employee in my department. He had done really well in the training but was still getting familiar with the job. I knew he was trying hard, but he kept falling behind on his work. So one day I got permission to work with him and provide additional training on the more difficult aspects of the job. The training seemed to really help because he hasn't fallen behind since.

Rating:\_\_\_\_\_

# Manipulation Check for Candidate Resumes

# Candidate A Background Information Questionnaire

# 1. Which automobile manufacturer has candidate A worked for?

- 1) Toyota
- ② Chevrolet
- **3 General Motors**
- ④ Mercedes Benz

# 2. How many years did candidate A work at Fairfield & Wells Equipment?

- 1) 2 years
- 2 4 years
- 3 6 years
- ④ 8 years

# 3. In what year will candidate A receive his BA?

2004
2002
2000

# Candidate B Background Information Questionnaire

# 1. Which automobile manufacturer has candidate B worked for?

- 1) Toyota
- ② Pontiac
- ③ Honda
- ④ Ford Motor Company

# 2. Which company has candidate B NOT worked for?

- ① Ford Motor Company
- 2 Walden Books
- 3 General Motors
- **④ Great Lakes Manufacturing**

# 3. What University does candidate B attend?

- ① University of Michigan
- ② Michigan State University
- **③ Central Michigan University**

# Interview Answer Sheets

# **Answer Sheet**

# Ratings for: CANDIDATE A

Dimensions	Low		Med		Hi
Problem Solving / Trouble Shooting	1	2	3	4	5
Leading and Influencing	1	2	3	4	5
Interpersonal Management	1	2	3	4	5
Train, Develop, and Evaluate Associates	1	2	3	4	5
Flexibility/Adaptability/Sensitivity	1	2	3	4	5
Performance Improvement Orientation	1	2	3	4	5
Quality Orientation	1	2	3	4	5
Safety/Cleanliness Orientation	1	2	3	4	5
Planning/Scheduling/Organizing	1	2	3	4	5
Communication	1	2	3	4	5

# **Answer Sheet**

# Ratings for: CANDIDATE B

Dimensions	Low		Med		Hi
Problem Solving / Trouble Shooting	1	2	3	4	5
Leading and Influencing	1	2	3	4	5
Interpersonal Management	1	2	3	4	5
Train, Develop, and Evaluate Associates	1	2	3	4	5
Flexibility/Adaptability/Sensitivity	1	2	3	4	5
Performance Improvement Orientation	1	2	3	4	5
Quality Orientation	1	2	3	4	5
Safety/Cleanliness Orientation	1	2	3	4	5
Planning/Scheduling/Organizing	1	2	3	4	5
Communication	1	2	<b>3</b>	4	5

#### Ideal Candidate Prototype Measure

# Impressions of the Candidates

Read each of the words or phrases listed below and indicate how similar (or dissimilar) each one is to describing **candidate A**. Please use the following scale to make your ratings:

- ① Very dissimilar
- **② Dissimilar**
- **3 Neither dissimilar or similar**
- **④ Similar**
- **5 Very Similar**

Descriptor	Dis		N		Sim
1. Approachable	1	2	3	4	6
2. Committed	1	2	3	4	5
3. Confident	1	2	3	4	5
4. Dedicated	1	2	3	4	5
5. Dependable	1	2	3	4	6
6. Determined	1	2	3	4	(5)
7. Disciplined	1	2	3	4	5
8. Educated	1	2	3	4	(5)
9. Efficient	1	2	3	4	(5)
10. Good Listener	1	2	3	4	5
11. Hard-working	1	2	3	4	5
12. Has held jobs involving supervisory responsibilities	1	2	3	4	5
13. Honest	1	2	3	4	5
14. Intelligent	1	2	3	4	5
15. Is a good role model for other employees	1	2	3	4	5
16. Open-minded	1	2	3	4	5
17.Organized	1	2	3	4	5
18. Professional	1	2	3	4	5
19. Qualified	1	2	3	4	5
20. Reliable	1	2	3	4	5
21. Resourceful	1	2	3	4	5
22. Respected by others	1	2	3	4	5
23. Trust worthy	1	2	3	4	5

Read each of the words or phrases listed below and indicate how similar (or dissimilar) each one is to describing **candidate B**. Please use the following scale to make your ratings:

- ① Very dissimilar
- ② Dissimilar
- **③ Neither dissimilar or similar**
- ④ Similar
- **5 Very Similar**

Descriptor	Dis	Τ	N		Sim
1. Approachable	1	2	3	4	5
2. Committed	1	2	3	4	(5)
3. Confident	1	2	3	4	(5)
4. Dedicated	1	2	3	4	5
5. Dependable	1	2	3	4	6
6. Determined	1	2	3	4	5
7. Disciplined	1	2	3	4	(5)
8. Educated	1	2	3	4	(5)
9. Efficient	1	2	3	4	5
10. Good Listener	1	2	3	4	5
11. Hard-working	1	2	3	4	5
12. Has held jobs involving supervisory responsibilities	1	2	3	4	5
13.Honest	1	2	3	4	5
14. Intelligent	1	2	3	4	5
15. Is a good role model for other employees	1	2	3	4	5
16. Open-minded	1	2	3	4	5
17.Organized	1	2	3	4	5
18. Professional	1	2	3	4	5
19. Qualified	$\bigcirc$	2	3	4	5
20. Reliable	1	2	3	4	5
21. Resourceful	1	2	3	4	(5)
22. Respected by others	1	2	3	4	5
23. Trust worthy	1	2	3	4	5

New #	Old #	Descriptors	Mean	SD
1	2	Approachable	4.87	0.34
2	6	Committed	4.86	0.35
3	9	Confident	4.70	0.47
4	16	Dedicated	4.83	0.38
5	18	Dependable	4.91	0.28
6	19	Determined	4.74	0.44
7	21	Disciplined	4.70	0.47
8	27	Educated	4.57	0.50
9	28	Efficient	4.78	0.42
10	36	Good Listener	4.74	0.44
11	37	Hard-working	4.91	0.28
12	92	Has held jobs involving supervisory responsibilities	4.61	0.49
13	40	Honest	4.77	0.42
14	47	Intelligent	4.61	0.49
15	93	Is a good role model for other employees	4.74	0.44
16	55	Open-minded	4.74	0.44
17	56	Organized	4.87	0.34
18	61	Professional	4.61	0.49
19	64	Qualified	4.83	0.38
20	65	Reliable	4.83	0.38
21	66	Resourceful	4.74	0.44
22	67	Respected by others	4.70	0.47
23	79	Trust worthy	4.70	0.47

Ideal candidate ratings provided by participants for the adjective checklist measure

# Hire Question

# 1. If you had to hire a candidate for the team leader position, who would you hire?

- ① Candidate A
- <sup>(2)</sup> Candidate B

# Experience Bias Measure

Read each statement that follows and indicate whether you agree or disagree. Please use the following scale:

- ① Strongly Disagree
- ② Disagree
- **③ Neither Agree or Disagree**
- ④ Agree
- **⑤** Strongly Agree

Question	Dis		Ν		Agr
1. Work experience is the best predictor of future job performance.	1	2	3	٩	6
2. The best employees have prior work experience.	0	2	3	4	6
3. The best training for a job is previous work experience.	0	0	3	٩	\$
4. If given the choice, I would hire someone with work experience over	1	2	3	4	6
someone without work experience.					

# Demographic Information Form

	-		
	Hank:		
		©Sophomore	
		CJunior	
		@Senior	
2. GPA:			
<b>3. Major</b>	Field of	f Study: <sup>(1)</sup> Biology/Chemistry	
		2) Business	
		③ Communications	
		Education Educatio	
		© Psychology	
		Pre-law/Pre-med	
		@ Other:	
4. Race:	1 Africa	an American	
	② Asiar	1	
	<li>3 Hispa</li>	anic	
	White	9	
	Othe	·	
5. Age:			
	- <b>-</b> -		
6. Sex: (	v Femal	e (2) Maie	
7. ACT/S	SAT sco	re:	
			·
8. Do yo	u know	any of the applicants in the video	aped interviews? ①Yes ②No
-			•
Work Ex	<u>(perienc</u>	<del>.</del>	
9. How I	m <mark>any</mark> pa	rt-time jobs have you held?	
40.11	•		
10. How	many t	ull-time jobs have you held?	······
11. For	how ma	ny years of your life have you held	a job (full or part-time)?
12. Have	e you ev	er worked in a production/manufa	cturing plant? ① Yes ② No
	-	•	••
<u>Interviev</u>	w Exper	ience:	
13. How	many ti	imes have <i>you</i> been interviewed fo	or a job?
0   ł	nave nev	er been interviewed for a job	
2 1-	2 times		
33-	4 times		
<b>④</b> 5-	6 times		
\$ 7-	9 times		
<b>6</b> 10	) or more	e times	
14. How	many t	imes have you <i>interviewed some</i> d	<i>ne else</i> for a job?
011	nave nev	er interviewed someone else for a ic	b
(2) 1-	2 times		
33	4 times		
<b>@</b> 5-	6 times		
© 7-	9 times		
® 1(	) or more	9	
6 1		-	

# APPENDIX F

Informed Consent & Debriefing Forms

# APPENDIX F

# Informed Consent & Debriefing Forms

#### Managerial Selection Informed Consent Form

**Purpose of Study:** The purpose of this study is to examine the effects of training on interview ratings.

**Procedures:** You will be asked to review an interview training manual that describes an employment interview used to hire managers in a production plant. After reviewing the manual, you will watch two videotaped interviews and rate the effectiveness of each candidate. The entire process should take approximately two hours.

**Benefits to You:** You will be trained on important interviewing skills that may help you if you ever interview someone or are interviewed for a job.

**Confidentiality:** Your information will be kept confidential. No one outside of the research team will have access to your demographic information or any of the data you provide in this study. None of your individual data will ever be released to the public.

**Questions:** You can contact Kevin Plamondon with any questions about the <u>research project</u> at: <u>plamond3@msu.edu</u> or (517)355-2171. Or you can contact Dr David Wright of the University Committee on Research Involving Human Subjects with questions about <u>your rights as a research participant</u>: <u>ucrihs@msu.edu</u> or (517)355-2180.

**Voluntary Participation:** You are under no obligation to participate in this study. There is NO penalty for choosing not to participate or for choosing to withdraw from this study at any time.

**Informed Consent:** If you have read each of the points above, understand each of the points above, and are willing to participate in this study, please write your name and the date and sign in the spaces provided below. By completing and returning this form you indicate your voluntary agreement to participate in this study.

Print Name

Date

Signature

#### Debriefing Form Managerial Selection Study

Thank you for participating in the Managerial Selection Study.

**Purpose of this Study:** The purpose of this study was to investigate the effects of interview training on the ratings of interview responses.

**Questions:** If you have any questions about this <u>research study</u> please contact Kevin Plamondon at <u>plamond3@msu.edu</u> or (517)355-2171. If you have any questions about your <u>rights as a research participant</u> please contact Dr. David Wright of the University Committee on Research Involving Human Subjects: <u>ucrihs@msu.edu</u> or (517)355-2180.

We thank you in advance for not discussing this research study with any students who may participate in this experiment in the future.

