RICE UNIVERSITY

THE EFFECT OF GROUP DISCUSSION ON EVALUATIONS OF JOB APPLICANTS

by

AMANDA PEEK

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

MASTER OF ARTS

APPROVED, THESIS COMMITTEE:

Robert L. Dipboye
Associate Professor of Psychology
Chairman

Craig A. Anderson
Assistant Professor of Psychology

Jay C. Thomas
Adjunct Assistant Professor of Psychology

William C. Howell
Professor of Psychology

HOUSTON, TEXAS

MAY, 1983
Abstract

The Effect of Group Discussion on Evaluations of Job Applicants

Amanda Peek

Two experiments examined the effect of group discussion on subsequent evaluations of job applicants. The hypothesis was tested that discussion would polarize the evaluations of applicants such that the average postdiscussion evaluation would be more extreme in the same direction as the average of the prediscussion evaluations. The relationship between polarization and quality of the evaluations was also examined. Measures of evaluation quality included interrater reliability and accuracy of recall of applicant and job characteristics. Subject groups were employed under two Discussion conditions: discussion of applicants or discussion of an irrelevant topic. Experiment 1 found a marginal polarization effect and an increase in interrater reliability following discussion of applicants. Experiment 2 also found a marginal polarization effect but no differences between conditions on interrater reliability or recall accuracy. Combined results suggest discussion may serve to polarize evaluations of job applicants, although each individual experiment did not statistically confirm the hypothesis.
Acknowledgements

The author would like to express her gratitude to Robert L. Dipboye, who served as chairman of the thesis committee. His continued expert guidance and support throughout the project was invaluable. The significant contributions to the project made by the other committee members, Craig A. Anderson, Jay C. Thomas, and William C. Howell, were also much appreciated.

The author would like to take this opportunity to thank her family, without whose generous financial and emotional support this program of study would not have been possible. Finally, for the humor and understanding which sustained her throughout this project, the author is indebted to Woody Phillips.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>16</td>
</tr>
<tr>
<td>Method</td>
<td>17</td>
</tr>
<tr>
<td>Results</td>
<td>20</td>
</tr>
<tr>
<td>Discussion</td>
<td>25</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>26</td>
</tr>
<tr>
<td>Method</td>
<td>30</td>
</tr>
<tr>
<td>Results</td>
<td>34</td>
</tr>
<tr>
<td>Discussion</td>
<td>41</td>
</tr>
<tr>
<td>General Discussion</td>
<td>45</td>
</tr>
<tr>
<td>Appendices</td>
<td>50</td>
</tr>
<tr>
<td>References</td>
<td>67</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>22</td>
</tr>
<tr>
<td>Table 2</td>
<td>24</td>
</tr>
<tr>
<td>Table 3</td>
<td>35</td>
</tr>
<tr>
<td>Table 4</td>
<td>37</td>
</tr>
<tr>
<td>Table 5</td>
<td>40</td>
</tr>
</tbody>
</table>
INTRODUCTION

An extensive body of literature has examined variables that affect personnel selection decisions. Dunnette and Borman (1979) classified research on selection into four selection methods that have been examined in the literature. They are the selection interview, biographical information, multiple assessment procedures, and job samples. Of these four methods, perhaps the most commonly used method has been the selection interview.

During the last 34 years, six major reviews of the research on the selection interview have appeared (Arvey & Campion, 1982; Mayfield, 1964; Schmitt, 1976; Ulrich & Trumbo, 1965; Wagner, 1949; Wright, 1969). Dozens of studies were reviewed in each paper, and a great variety of interviewer and applicant variables were shown to affect interviewers' selection decisions. Such variables included the negative or positive nature of information presented by the applicant (Bolster & Springbett, 1961; Hollmann, 1972; Springbett, 1958), visual and verbal cues presented by the applicant (Sigelman, Elias, & Danker-Brown, 1980; Washburn & Hakel, 1973), nonverbal cues presented by the applicant (Imada & Hakel, 1977; McGovern & Tinsley, 1978; Sigelman, Elias, & Danker-Brown, 1980; Tessler & Sushelsky, 1978), and the order of presentation of information provided by the applicant (Blakeney & MacNaughton, 1971; Farr, 1973; Johns, 1975; Peters & Terborg, 1975; Springbett, 1958; Tucker & Rowe, 1979). Interviewer experience (Carlson, 1967), interviewer training (Latham, Wexley, & Purcell, 1975; Vance, Kuhnert, & Farr, 1978; Wexley, Sanders, & Yukl, 1973), and interviewee training (Hollandsworth, Dressel, & Stevens, 1977; Hollandsworth & Sandifer, 1979) were also examined as variables.

Past research on the selection interview has been concerned primarily with the perceptions, evaluations, and decisions of the individual in the evaluator role. This emphasis has been most apparent in studies that have investigated unfair discrimination in the interview. In a major review of this literature, Arvey (1979) pointed to two mechanisms that have been used to explain differential evaluations of minority candidates in the interview: (1) stereotyping, and (2) differential behavior emitted during the interview. In the use of stereotyping as an explanation, the focus has been on stereotypes held by individual interviewers while the explanation of differential behavior of minority applicants focuses on individual applicant characteristics. Dipboye (Note 1) claimed that this focus on individual interviewer and applicant variables in the discrimination literature has provided a sterile conceptualization of the phenomenon underlying the research. To focus on the individual is to assume that the decision maker forms his or her evaluations in relative isolation. The thesis of
Dipboye's paper was that the social context of the selection decision has been ignored as a determinant of unfair discrimination.

Dipboye's argument can be extended to the entire selection decision-making literature. The social context of the selection decision has been ignored in selection interview research outside of studies of unfair discrimination. This presents a serious void in the applied literature, as it is common practice in many companies for evaluators to discuss applicants prior to the final selection decision. For example, recruiters may interview applicants on a college campus and make recommendations regarding an applicant's suitability for the job, but final decisions to invite the applicant for a second interview or hire the applicant are likely to be made by another member of the organization. The hiring member is likely to discuss the applicant with the recruiter and perhaps with other relevant members of the organization. Group discussion of the applicant is likely to have an effect on the final selection decision.

This chain of events is not uncommon. Some kind of discussion of job applicants often occurs prior to the final selection decision at all levels of organizational entry. Nonetheless, the effect of group discussion on evaluations of job applicants has been ignored in previous research. The research to be presented provides an initial investigation into the role of group discussion in evaluation of job applicants.

There are at least two alternative predictions for the effect of group discussion on evaluation of job applicants that could be inferred from past research. One prediction is that group discussion decreases the range of individuals' ratings of the applicants following
discussion. The vast research on conformity demonstrates a powerful tendency for members of a group to move toward agreement following discussion. Individual judgments made following group discussion are shown to converge toward the mean of the individual prediscussion judgments, decreasing the range of postdiscussion judgments from the prediscussion range. This convergence has been widely demonstrated even when no instruction is given for the group to reach a consensus (Brown, 1965).

A second possible effect of group discussion is that individuals shift their evaluations of applicants in the direction of the prediscussion mean evaluation. This effect, labeled the group polarization phenomenon, has been found in several hundred studies on group-induced shifts in individual attitudes and behavior (Lamm & Myers, 1978; Myers & Lamm, 1976). This group shift is usually accompanied by a reduction in range of individual judgments as is found in conformity research, but the postdiscussion shift is not equivalent to conformity findings. While conformity predicts the convergence of individual judgments toward the prediscussion central value, the polarization shift demonstrates a postdiscussion shift to a position more extreme but in the same direction as the prediscussion mean.

**Origins of Group Polarization Research**

Before applying the group polarization phenomenon to personnel selection, it may prove enlightening to provide a brief perspective on the origins and development of this area of research. Research on group polarization derives from an extensive wave of investigations into group risk-taking behavior triggered by a finding of James Stoner in 1961. Stoner examined the notion that groups were more cautious
decision-makers than individuals. The "choice-dilemma" paradigm used in his study was employed in much of the subsequent research. Subjects were presented with items that required choices between various alternatives. Each of the choices required the decision maker to make choices among alternatives varying in the degree of risk they posed for a hypothetical "other". After subjects individually made a number of these choice decisions, they were combined into groups and required to discuss the items until a level of risk acceptable to all members of the group had been determined for each item. Subjects were separated and again individually indicated their risk preference on each item.

Stoner's (1961) findings and those of hundreds of subsequent investigations demonstrated that groups accepted higher levels of risk than individuals. Wallach and Kogan (1965) further demonstrated a change in the individual risk level, following group discussion, such that postdiscussion individual ratings were approximately the same as the group consensus ratings. Consequently, individual choices were found to be riskier following group discussion.

The "risky-shift" effect, as it was labeled, was investigated exhaustively during the 1960's, and the phenomenon was found to be robust. Many excellent summaries of the risky-shift literature have been written (e.g. Cartwright, 1971; Clark, 1971; Dion, Baron, and Miller, 1970; Pruitt, 1971; Vinokur, 1971), and the reader is referred to these for thorough reviews of the research.

In the course of investigation into group risk-taking behavior it was demonstrated that topics eliciting relatively cautious prediscussion responses elicited group-induced shift in the cautious direction following discussion (Chandler & Rabow, 1969; Nordhoy, 1971; Rabow,
Fowler, Bradford, Hofeller, & Shibuya, 1966; Stoner, 1968). Further investigation, with a variety of item types other than choice-dilemma items, served to free the phenomenon from its conceptualization as intrinsically connected with risk-taking behavior. In a significant piece of research, Moscovici and Zavalloni (1969) replicated the risky-shift phenomenon using topics unrelated to risk-taking. They asserted that the risky-shift phenomenon was, "a content-bound exception to the averaging tendency of the group" (p. 125), and suggested the term "group polarization" to refer to the general phenomenon found to explain risky-shift findings.

The group polarization phenomenon predicts the average postdiscussion response will tend to be more extreme in the same direction as the average of the prediscussion responses. It should be noted that predictions about group polarization need not apply to individuals. Fraser, Gouge, and Billig (1971) made a distinction between group shifts, where the group is represented by means, and individual shifts, where all calculations reflect changes in the scores of each individual, one at a time. To employ their example, a three person group using a seven-point scale could have initial scores of 7, 5, and 2 (4 being the neutral point on the scale). After discussion, all members could respond with a score of 5. At the group level, a slight move to extremity (polarization, in this case) occurred in a mean shift from 4.67 to 5.00. Yet at the individual level, two subjects shifted to less extreme responses, and none of the subjects demonstrated a more polarized response (p. 17).

The present investigation is an attempt to extend the group polarization phenomenon to personnel selection decisions. In the
typical selection situation, several applicants are reviewed by several
decision makers within the organization. Decision makers typically
review applications individually, conduct interviews with the
candidates, then meet to discuss their individual impressions of the
applicants. Some kind of decision is then made regarding the
suitability of the candidates for the job. One candidate may be chosen
or candidates may be ranked on their suitability for the job. The final
decision may be made by the whole group or unilaterally made by one
member of the group. The selection task provides an excellent situation
in which group polarization may be examined. The present research
tested the prediction that polarization of decisions regarding
applicants will occur following discussion.

There are two areas of research in the group polarization literature
that seemed relevant to the hypothesis tested in the present research.
One area of research has been concerned with the group polarization
effect in person perception situations such as performance evaluation
and jury decision making. Second, recent research on the assessment
center approach provides additional evidence supportive of the
hypothesis. The relevant studies in these two research areas are
reviewed, following which a rationale is presented for the present
research.

Evidence of Polarization with Person Perception Tasks

The generality of the group polarization effect has been established
across diverse laboratory tasks and in "real-world" situations. Rather
than review the entire polarization literature, the present review
focuses on investigations of polarization using person perception tasks.
Of the entire population of polarization studies conducted across
various tasks and stimulus domains, this domain shares the most similar
task characteristics to the personnel selection situation. For the
reader's reference, the population of polarization findings up to 1978
has been summarized, and the summaries are available elsewhere (Lamm &
Myers, 1978; Myers, 1973; Myers & Lamm, 1976).

In an attempt to demonstrate group polarization in a decision-making
context of what he considered "practical importance", Forgas (1977)
studied the effect of group discussion on judgments of person stimuli.
Four simulated interviews were videotaped with actors impersonating
interviewees. Interviews were of a general nature and were not job
interviews. Subjects rated interviewees on a number of semantic
differential scales. Groups that discussed interviewees provided
significantly more extreme postdiscussion ratings for all four stimuli
in the direction of the mean of the individual judgments, thus
supporting the polarization effect. Groups that were instructed to
focus their discussion on the procedural aspects of the discussion task,
as opposed to on interviewee characteristics, did not show a significant
ratings shift. Although Forgas' task is similar in many ways to the
personnel selection task, it differs from the selection task in that no
hiring decisions were made regarding the interviewees.

Performance evaluation. Other research in the polarization
literature has examined the role of group discussion in evaluation of
real or hypothetical faculty members.

The faculty evaluation task shares many characteristics with the
personnel selection task. Real or hypothetical persons are employed as
stimuli and, in the case of hypothetical persons, paper credentials of
these persons are presented. Subjects are required to evaluate a pool
of stimulus persons presented simultaneously, and, as such, some comparison between stimuli is inevitable prior to making final evaluations.

Andrews and Johnson (1971) required subjects to rate university teachers on the basis of eight favorable, eight neutral, and eight unfavorable descriptive statements provided by the experimenter. As predicted, postdiscussion ratings by groups were more positive than the predisussion ratings when judgments were based on favorable cues, and more negative when based on unfavorable cues. However, Shrewsberry and Johnson (Note 2) failed to replicate these findings in a subsequent study.

Subjects in a study by Myers (1975) were required to rate and recommend increases for three "good faculty" (for which the initial tendency was to rate positively) and three "bad faculty" (elicited predominately negative initial tendencies). Experimental subjects discussed the ratings and pay recommendations following their individual responses. Ratings and pay recommendations were made on an individual basis following discussion. Control subjects rated and made pay recommendations for the faculty, discussed an irrelevant topic, and then provided a second set of ratings and recommendations. As predicted, discussion in the experimental groups significantly polarized the mean judgments of the faculty such that "bad faculty" were rated more negatively and "good faculty" more positively following discussion. One problem for Myers' (1975) findings is that although the polarization of judgments was significantly greater in the experimental than in the control condition, control subjects also polarized in their judgments. Polarization of control groups' judgments might result from a less than
distracting filler task that enabled subjects to think about their initial responses. Polarization of individual judgments has been produced by "thought" alone (Tesser, 1978). As Myers does not provide a description of the control condition's filler task, one can only conjecture about the extent to which subjects could think about the stimulus persons during that time period.

Krapf (1972) employed student subjects' actual professors as stimulus persons and required subjects to rate their professors individually and then to rate them in groups. In support of the polarization effect, groups gave more negative ratings than individuals when the average individual ratings were already somewhat negative and gave more positive ratings when individual ratings were positive.

Jury decision. Examination of the effect of group discussion on jury decisions has provided another subset of person perception research in the area of group polarization. The jury decision task shares characteristics common to the personnel selection task. To begin with, subjects are required to evaluate persons as stimuli. Subjects are required to make a consensus decision, following discussion, regarding the stimulus person. Unlike the personnel selection task, however, subjects evaluate the characteristics of only one stimulus person. The decision-making aspect of the task is thus differentiated from the selection task in that it does not require a comparison among a number of stimulus alternatives.

Experimental evidence of jury decision-making has been obtained only through simulations of the jury process, as actual juries may not be directly observed in the deliberation process. These studies expected to find decisions following jury deliberation predicted by
initial tendencies of individual jurors. Specifically, juror decisions following jury deliberation were predicted to be more extreme in the direction of the average initial inclination of the jurors.

Mock jury decisions have shown a clear polarization effect in several studies. Myers and Kaplan (1976) and Kaplan and Miller (1976) required mock jurors to decide the guilt of traffic violators. Defendants were made to appear guilty or not guilty through initial presentation of stimulus materials which elicited a predominant disposition in line with the defendant guiltiness manipulation. Both studies found that, after discussing the low-guilt defendants, jurors were even more definite in their judgments of innocence and more lenient in recommended punishment. After discussion of high-guilt defendants, jurors polarized toward harsher judgments of guilt and harsher punishment recommendations.

In a series of three experiments, Kaplan (1977) examined polarization of guiltiness judgments and punishment recommendations in an attempted manslaughter case. Two courtroom trials were constructed in which one contained witness testimony incriminating to the defendant while the other contained testimony suggesting the defendant's innocence. Jurors then discussed the trial through exchange of notes with bogus jurors (arguments thought to be from other jurors but actually provided by the experimenter). The content of the notes was varied in the three experiments. In Experiments 1 and 2, discussion produced polarization when subjects received arguments from other "jurors" that were similar in proportion of guiltiness/innocence comments towards the defendant to their own initial arguments. In Experiment 3, redundancy of information in the discussion phase was
varied along with similarity of subject-other juror arguments. Postdiscussion ratings polarized more when shared information was nonredundant than when it was redundant. The effect of similarity of proportion of arguments on polarization was also replicated, but shifts were stronger with nonredundant arguments.

In another experiment investigating the effect of deliberation content on polarization of jury decisions, Kaplan and Miller (1977) manipulated the homogeneity of the order in which facts of a case were heard by the jurors. In half the juries, jurors heard either incriminating or exonerating facts about a defendant in the same (homogeneous) order. In the remaining juries, each juror heard the facts in a different (heterogeneous) order. As in the typical paradigm, jurors rated the defendant's guilt, discussed the case as a jury, and then individually re-rated the defendant. They found jurors discussed and remembered a greater variety of facts under the heterogeneous than under homogeneous conditions. This led to greater polarization of postdiscussion judgments under the heterogeneous condition.

Employing mock juries, Rumsey (1975) found group discussion amplified a tendency for jurors to assign more severe penalties for important crimes than unimportant crimes.

Other jury studies fail to corroborate polarization findings. Izzett and Leginski (1974) found a tendency, prior to group discussion, for mock jurors to impose harsher sentences on unattractive defendants than attractive defendants. Although they expected an amplification of this treatment difference following deliberation, they found a significant shift toward leniency in the unattractive defendant condition and no change in the attractive condition.
A shift toward leniency after discussion, regardless of the direction of the initial judgments, has been found in other jury studies (Davis, Kerr, Atkin, Holt, & Meek, 1975; Gleason & Harris, 1976; Rumsey, 1976; Rumsey, Allgeier, & Castore, 1978). Izzett and Leginski (1974) and Rumsey, Allgeier, and Castore (1978) attempted to reconcile a leniency shift with polarization findings. They argued that initial "harsh" sentences in past research actually were lenient when compared with the midpoint of the specific judgmental scale employed. They claimed that the initial tendency should be defined with respect to the scale midpoint. When initial tendencies were redefined in this manner, both experiments found "harsh" sentencers to be "lenient" as their initial judgments fell on the lenient side of the scale midpoint. Redefining initial judgments this way would cause postdiscussion shifts to leniency in the initial "harsh" judgment conditions to be considered evidence of polarization. However, perhaps a better explanation might be that a socially-valued tendency toward leniency in certain kinds of cases may become amplified during jury deliberation, which might serve to override a polarization effect.

Although evidence presented to this point has been based only on studies using jury simulations, Kalven and Zeisel (1966) provide evidence that polarization may be operating in actual jury decision-making. In The American Jury, data was presented from 225 trials demonstrating that the jurors' first ballot vote, taken prior to deliberation, decided the outcome of the verdict. The initial majority predicted the consensus outcome in almost all cases.

Research on selection decisions. Although no research has been conducted testing group polarization effects in selection tasks, one
study suggested such effects will occur. A recent study by Sackett and Wilson (1982) provided evidence of polarization of individual judgments following discussion of job applicants. They set out to examine the consensus judgment process in the managerial assessment center. Final assessment center ratings of a candidate are determined by a two-step process: individual assessors rate each candidate, then discuss the candidates and attempt to reconcile their differences in prediscussion ratings to arrive at a consensus decision regarding the candidate. Although the consensus discussion may be considered the most central aspect of the assessment center process, Sackett and Wilson noted the absence of research on this phase of the process. Among other questions investigated in the study, Sackett and Wilson examined whether the group consensus judgment could be predicted on the basis of the individual assessors' prediscussion ratings. They found the mean of the individual prediscussion ratings predicted the outcome of the consensus discussion for 93.5% of all ratings made in the assessment center investigated. In examination of the patterns of resolution of disagreements in the consensus discussion, they found extreme ratings had more influence than midrange ratings in determining the final decision. The authors noted the resemblance of their findings to group polarization, although they offered no predictions regarding the consensus decision outcome.

Rationale for Present Research

In review of the pool of studies presented, the extent to which group polarization has been demonstrated in person perception tasks is unclear. The inconsistency in the findings in this task domain is interesting, in light of the robustness of the polarization effect across other task domains. The evidence suggests that the extent to
which polarization occurs may depend on the nature of the task. As the personnel selection task differs from person perception tasks reviewed, in respects previously discussed, it is of interest to explore the extent to which group polarization is demonstrated in personnel selection. Despite the inconsistency of the findings, on the basis of studies demonstrating group polarization and the Sackett and Wilson (1982) research on assessment center evaluations, it is suggested that group polarization will occur in the group evaluation of job applicants. The present studies were specifically designed to test this possibility.

One important aspect of the polarization effect that has been ignored in previous research is the relationship between polarization and the quality of the decision made. Although a vast amount of research comparing group versus individual performance (see Hill, 1982, for a recent review) has addressed the quality of group judgments, polarization of judgments has never been related to their subsequent quality. Does polarization of responses lead to better decisions made following group discussion? Perhaps decisions of a lower quality result from group polarization, or there may be no relationship at all between decision quality and polarization. This issue was examined in the present studies.

The selection literature suggests that the two important dimensions of quality are the reliability and validity of the selection decision. Over 60 years of research on the selection interview has focused on the reliability and validity of selection judgments (Arvey & Campion, 1982). Over the years, reviews of the literature have shown that studies of the selection interview report lower reliabilities than should be
accepted for devices used for prediction (Arvey & Campion, 1982; Mayfield, 1964; Schmitt, 1976; Wagner, 1949; Ulrich & Trumbo, 1965). In the present studies, reliability was assessed as a measure of quality of the selection decisions in an effort to determine the relationship between polarization and decision quality.

A second dimension of quality of the selection decision examined was the accuracy of assessment of applicant characteristics. Experiment 2 explored the extent to which response polarization was accompanied by an increase or decrease in accuracy of evaluators' recall of applicant characteristics. Although it would have been desirable to assess the predictive validity of the selection decision, it was not possible in the present experiment. Although the accuracy of assessment of applicant characteristics may or may not be related to the predictive validity of selection decisions, it seems fairly clear that accuracy of assessment is related to the content validity of the interviewer's decision. Thus, it is assumed that accurate recall of applicant characteristics leads to higher quality decisions by the interviewer.

EXPERIMENT 1

In the present experiment, groups of subjects rated job applicants prior to and following either a group discussion of the applicants (Relevant Discussion condition) or discussion of an irrelevant topic (Irrelevant Discussion condition). It was hypothesized that polarization would be demonstrated on the ratings of applicants in the Relevant Discussion condition such that applicants who were initially rated favorably would be rated more favorably following group discussion; and those low in initial favorability would be rated even
less favorably following discussion. No shift was expected for Irrelevant Discussion groups.

A second issue addressed was the extent of agreement among individuals in Relevant Discussion groups compared to that of Irrelevant Discussion groups. One might expect that a polarization effect would be accompanied by more agreement among raters. It is also possible, however, that polarization effects occur without any corresponding increase in agreement. In order to examine the effect of discussion on rater agreement, or interrater reliability, the intraclass correlation coefficient was computed for ratings on each dimension before and after discussion for both Relevant Discussion and Irrelevant Discussion conditions. The magnitude of the intraclass correlations computed for prediscussion ratings was compared with the magnitude of postdiscussion ratings in both conditions. The magnitude of the postdiscussion intraclass correlations in the Relevant Discussion condition was also compared with the postdiscussion coefficients computed in the Irrelevant Discussion condition.

**Method**

**Subjects and design.** Twenty-two female and 20 male undergraduate volunteers participated in the experiment in three- and four-person groups.

Six experimental and six control groups were formed with the sex composition of groups balanced across conditions. The design was a $2 \times 2 \times 8$ factorial, with the independent variables being Group Discussion (Relevant or Irrelevant), Rating Period (Prediscussion or Postdiscussion), and Applicant Suitability (eight levels from highly suitable to low suitability). Group Discussion was a between-groups
factor while the Rating Period and Applicant Suitability factors constituted within-subjects factors.

Procedure. Upon reporting to a session, each group member received a description of the job of Sales Supervisor Trainee (Appendix A) and a set of eight standard employment application forms filled out by eight hypothetical job applicants. Completed application forms contained responses to questions about educational qualifications, extracurricular activities, work experience, and references (see Appendix B). In an effort to provide a realistic pool of applicants, three female and five male applicants were provided. All applicants were highly qualified for the job. Subjects were instructed briefly on the nature of role-playing, and were asked to imagine themselves employment interviewers evaluating the applicants for the Supervisor Trainee position.

Subjects rated each applicant on four 9-point scales which served as the dependent measures of the experiment. Two other 9-point scales served as checks on the qualifications depicted in the resumes.

In all groups subjects read the job description then rated applicants on the six rating dimensions. Subjects were then provided ranking forms and were requested to rank applicants individually from one to eight in suitability for the job. In the Relevant Discussion condition subjects then discussed each applicant for 20 minutes and were instructed to arrive at a group consensus decision regarding the ranking of the applicants in suitability for the job. Irrelevant Discussion subjects were provided hypothetical personnel cases, unrelated to the personnel selection task, following the ranking task. They were
instructed to read the cases, discuss them as a group for 20 minutes, and generate a group consensus solution for each case situation.

Following the discussion tasks, all subjects were given the same set of eight application forms that they had previously reviewed. They were instructed to again review and rate the applicants on a new set of rating forms containing the six original rating dimensions. Subjects were told that they might or might not choose to change their ratings following the second evaluation. When the second rating period was completed, all materials were collected, and the subjects were debriefed.

Dependent measures. Subjects rated applicants along the following six dimensions (intervals are whole).

1. On the basis of all the information you have seen, how would you evaluate this candidate's qualifications for the job for which he/she has applied? (1 = very poor to 9 = very well)

2. If you were the interviewer, would you invite this candidate to visit your company for a second interview? (1 = definitely would not to 9 = definitely would)

3. If you were the interviewer, would you hire this person? (1 = definitely would hire to 9 = definitely would not hire)

4. How much do you think you would personally like the candidate? (1 = I would definitely like this person to 9 = I would definitely dislike this person)

Responses to the first three items were assumed to reflect subjects' evaluations of the applicant's overall suitability for the job. Coefficient alphas were computed on prediscussion and on postdiscussion responses to the three measures and were .82 and .88, respectively.
Therefore, responses to the three job-relevant items were combined into a composite dependent measure for subsequent analysis.

Two additional items served as checks on the level of qualifications of the resumes:

5. How well do you think this candidate's experience fits the job? (1 = very poor fit to 9 = very good fit)

6. How well do you think this candidate's education and training fits this job? (1 = very poor fit to 9 = very good fit)

Results

All analyses used groups, as opposed to individuals, as the unit of analysis in order to avoid the problem of non-independence of individual scores within groups.

Responses to dependent measures regarding applicants' "experience" and "education and training" served as checks on the assumption that all applicants were perceived as highly qualified for the job. As all applicants' experience was perceived to have at least a good fit to the job (means were greater than 6.70 for each applicant), and a good education and training fit to the job (means greater than 6.90 for each applicant), the assumption was verified.

Polarization. A linear trend analysis was performed to test the hypothesized rating shifts on the high and low suitability applicants. The suitability ranking of the applicant (1 to 8) was computed as the rank-order of the mean suitability ranking for each applicant taken prior to the discussion task within each group. Rankings were averaged within each group. The rank-order (from 1 to 8) of the mean suitability rankings was utilized as eight levels of the Applicant Suitability factor. Application of the specific trend coefficients
(7, 5, 3, 1, -1, -3, -5, -7) to the eight levels of Applicant Suitability tested the hypothesis that the greatest positive rating shift would occur for the most suitable applicant, the greatest negative shift for the least suitable applicant, and less extreme shifts would occur for applicants at the moderate Suitability levels. To confirm the polarization hypothesis, the linear trend of Applicant Suitability must be found only at the Relevant Discussion level of the Group Discussion condition factor. Shift scores on the "composite" and "like" rating dimensions served as the dependent variables. Shift scores were computed as postdiscussion minus prediscussion responses to the rating dimensions.

A 2 X 8 mixed design ANOVA was employed to test the polarization hypothesis. Group Discussion condition (Relevant and Irrelevant) provided the between-subjects factor. Eight levels of Applicant Suitability (linear) provided the within-subjects factor. Polarization would be demonstrated with a significant interaction in which the linear trend of Applicant Suitability would differ for the two levels of Group Discussion condition.

Table 1 presents the mean shift score ratings for both Group Discussion conditions on both rating dimensions. Positive shift scores indicate an increase in favorability of the ratings following discussion. Negative shift scores indicate a postdiscussion decrease in favorability of ratings. Examination of Table 1 shows that obtained mean shifts exhibited the polarization pattern to some degree for the "composite" dimension, while "like" showed few of the shifts predicted by the polarization pattern.
Table 1
Mean Shift-score Ratings Grouped by Applicant Suitability and Group Discussion Condition: Experiment 1

<table>
<thead>
<tr>
<th>Rating Dimension</th>
<th>Applicant Suitability(^a)</th>
<th>Relevant Discussion</th>
<th>Irrelevant Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;composite&quot;</td>
<td>1</td>
<td>.15</td>
<td>-.33</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.38</td>
<td>-.34</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.46</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.22</td>
<td>-.35</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-.22</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-.51</td>
<td>-.65</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>-.85</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-.91</td>
<td>-.58</td>
</tr>
<tr>
<td>&quot;like&quot;</td>
<td>1</td>
<td>.00</td>
<td>-.29</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.14</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.24</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.29</td>
<td>-.24</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-.05</td>
<td>-.33</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.48</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>.28</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>.29</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note. Shift-scores are computed from postdiscussion minus prediscussion ratings. \(n = 6\) per condition.

\(^a\)1 = high suitability to 8 = low suitability.
Linear trend coefficients were applied to the mean ratings for Relevant and Irrelevant Discussion groups to produce the Applicant Suitability(linear) factor within each Group Discussion condition. Univariate tests of the linear trend of Applicant Suitability demonstrated a significant linear trend in the Relevant Discussion condition for the "composite" dimension, $t(5) = 2.86, p < .05$, while no such trend was identified for the "like" dimension. Neither dimension demonstrated linear trends in the Irrelevant Discussion condition. In the test of the polarization hypothesis, the interaction of Applicant Suitability(linear) with Group Discussion condition approached significance for the "composite" dimension, $F(1,10) = 3.69, p = .08$.

Reliability. Intraclass correlation coefficients were computed separately for the "composite" measure and for each individual rating dimension in both Relevant and Irrelevant Discussion conditions. Individual ratings within groups were averaged, and a group score computed. As such, there were six independent rating responses employed in computing the intraclass coefficient for the Relevant Discussion condition and six independent responses for the Irrelevant Discussion condition for each rating period on each dimension.

From Table 2 it appears that interrater reliability increased following discussion of the applicants on three of four rating dimensions and the "composite" measure. Irrelevant Discussion condition groups saw the opposite pattern with a decrease in interrater reliability following discussion on three of four dimensions and the "composite" measure. In comparison of the magnitude of postdiscussion
Table 2
Intraclass Correlations for Relevant Discussion and Irrelevant Discussion Conditions on Pre- and Postdiscussion Ratings:

Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Relevant Discussion</th>
<th>Irrelevant Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>&quot;composite&quot;</td>
<td>.49</td>
<td>.78</td>
</tr>
<tr>
<td>&quot;qualifications&quot;</td>
<td>.50</td>
<td>.67</td>
</tr>
<tr>
<td>&quot;invite&quot;</td>
<td>.22</td>
<td>.62</td>
</tr>
<tr>
<td>&quot;hire&quot;</td>
<td>.94</td>
<td>.80</td>
</tr>
<tr>
<td>&quot;like&quot;</td>
<td>.44</td>
<td>.52</td>
</tr>
</tbody>
</table>
coefficients, Relevant Discussion groups provided more reliable ratings on every dimension.

**Discussion**

The first experiment provided limited support for the prediction that group discussion of job applicants would polarize evaluations of the applicants. While the polarization of ratings was found on the "composite" rating dimension for groups that discussed the applicants, the hypothesized difference in polarization between Discussion condition groups only approached significance. As only six groups were employed in each Discussion condition, these results suggest that a larger sample of groups may be necessary to detect statistically significant differences between conditions.

Polarization on the "like" dimension was not demonstrated in the present study. It was not clear why the job-relevant "composite" dimension demonstrated polarization but not the "like" dimension.

The present study demonstrated that groups that discussed applicants provided more reliable postdiscussion ratings than groups that did not discuss applicants. It should be noted that the high variability of reliability coefficients between rating dimensions in the Relevant Discussion prediscussion ratings was unanticipated. This may indicate some instability of the coefficients, probably as a result of a small sample of mean ratings (n = 6) used to calculate each coefficient. As a result, interrater reliability results should be interpreted cautiously. However, the consistency of the increase in postdiscussion reliability coefficients, across every rating dimension, suggests there may be an advantage to using group decision-making in personnel selection. This finding should prove encouraging to those involved with
assessment centers. As mentioned previously, the assessment center procedure utilizes consensus discussion among assessors as the critical component of the assessment center technology (Sackett & Wilson, 1982). Final ratings of a candidate are the product of the consensus discussion. The results of the present study suggest that discussion of ratings may improve final decisions by improving agreement among raters in their evaluations. Of course, increased agreement may be at the cost of accuracy and further research is needed to explore the extent to which polarization is accompanied by errors in assessment.

A second experiment was conducted to investigate whether the polarization effect would be strengthened with utilization of a larger subject sample. It was hoped that utilization of a larger subject pool would also allow a more accurate assessment of interrater reliability, as a result of less variability in reliability coefficients. The relationship between polarization and errors in assessment was also examined in the second experiment.

EXPERIMENT 2

In the following experiment, groups of subjects rated job applicants prior to and following either a group discussion of the applicants (Relevant Discussion condition) or discussion of an irrelevant topic (Irrelevant Discussion condition). In a departure from the design of Experiment 1, prediscussion evaluations of the six applicants employed were manipulated such that each applicant differed in their level of qualifications (six levels from low to high) for the job. The six levels of applicant qualifications served as six levels of the Applicant Suitability factor. In Experiment 1, levels of Applicant Suitability were not manipulated but were constructed from a ranking of the
applicants provided by subjects. All applicants were evaluated as highly qualified for the job on the prediscussion evaluations. As the ranking of applicants in suitability for the job differed from group to group, no single applicant was consistently ranked the most highly qualified applicant and no single applicant was consistently ranked the least qualified. Providing applicants clearly differentiated in qualifications for the job provided a stronger test of the polarization hypothesis in Experiment 2 than was provided in Experiment 1. Consistent with the predictions tested in Experiment 1, group discussion of applicants was hypothesized to polarize evaluations of applicants such that high suitability applicants would be rated more favorably and applicants low in suitability would be rated less favorably following discussion. No shift was expected for the Irrelevant Discussion condition.

Secondly, it was expected that a polarization effect would be accompanied by higher interrater agreement, as found in Experiment 1. This would be demonstrated by an increase in interrater reliability following discussion of the applicants in the Relevant Discussion condition. Interrater reliability of the Relevant Discussion condition's postdiscussion ratings would be expected to be greater than the postdiscussion ratings of groups that did not discuss the applicants.

Three other issues concerned the interpretation of the polarization and reliability effects found in the first study. One such issue was whether increases in interrater agreement accompanying the polarization effect would lead to increased or decreased errors in assessment. Subjects were requested to accurately recall characteristics of the
applicants and characteristics of the job following the discussion task. On the one hand, it might be expected that an increase in interrater agreement, found to accompany polarization in Experiment 1, might lead to errors in recall if only a subset of applicant and job characteristics had been the focal point of group discussion. On the other hand, if interrater agreement was enhanced by thorough discussion of the pool of applicant and job characteristics, as might be expected with equal input from all evaluators, then polarization might be accompanied by greater accuracy. No hypotheses were offered as to the relationship between polarization and accuracy of recall of applicant and job characteristics.

Another question raised by the findings of the first study was whether increased interrater agreement in ratings of applicants accompanying group discussion could be attributed at least in part to increased agreement on perceptions of important applicant qualities for the job. In the second experiment, subjects were asked to rank-order a set of qualities describing applicants in terms of their importance for the job. It was hypothesized that subjects who discussed job applicants would demonstrate greater consensus regarding a rank-order of the importance of applicant qualities for the job than subjects who did not discuss applicants.

The third issue examined in this study concerned the group process possibly accounting for polarization effects. The group process underlying polarization has been examined in research predicting informational influence as an explanation for group polarization. As an explanation for polarization, the informational influence approach attributes the polarized response changes to individuals' cognitive
learning as a result of exposure to arguments during the discussion. It was hypothesized that arguments would be generated in favor of the initially favored alternative or against the least favored alternative. These persuasive new arguments would be learned and integrated by individuals and would affect their future responses by moving them in the direction of the persuasive arguments. Two types of studies have tested this explanation. Manipulation of the availability of arguments has produced polarization in a few studies (e.g. Burnstein & Vinokur, 1973; Myers, Wong, & Murdoch, 1971). The second type of study provides a content analysis of arguments generated by individuals within groups. A number of studies have conclusively demonstrated that the amount of postdiscussion shift on an opinion item varies directly with the number of arguments about that item occurring during the discussion of the item (Bishop & Myers, 1974; Ebbesen & Bowers, 1974; Vinokur & Burnstein, 1974).

Hoffman's (1979a) method of counting the valence of individual arguments was chosen as a method of examination of the group process underlying polarization. The valence of an adopted solution was measured by computing the difference between the sum of arguments generated by group members in favor of the solution and the sum of arguments generated opposing the solution. The valence-counting technique was chosen as it provides an elegant and parsimonious method of content analysis of arguments generated by individuals within groups. Hoffman and his associates found that the valence of individual arguments was a significant predictor of individuals' influence in a discussion and was a predictor of the group's adoption of a solution (see Hoffman, 1979a, 1979b, for review). Corroborative evidence may be
found in Myers and Bishop (1971). They determined that simple counting of positive and negative comments was as predictive as a complex mathematical model in predicting the polarization response shift.

It was hypothesized that valence measures would predict the polarization response shift such that applicants high in suitability would elicit arguments of positive valence while applicants low in suitability would elicit arguments of negative valence. Specifically, valence measures would be directly related to the hypothesized polarization of ratings such that there would be greatest positive valence for the most suitable applicant, the greatest negative valence for the least suitable applicant, and valence scores for applicants 2 through 6 in suitability would be rank-ordered in magnitude corresponding to applicants' suitability levels.

Method

Subjects and design. Fifty male and 30 female MBA students from two universities participated in the experiment in three- and four-person groups. Eleven experimental and 11 control groups were formed with the sex composition of groups balanced across conditions. The design was a 2 X 2 X 6 factorial, with the independent variables being Group Discussion condition (Relevant or Irrelevant), Rating Period (Prediscussion or Postdiscussion), and Applicant Suitability (six levels from highly suitable to low suitability). Group Discussion condition was a between-groups factor while Rating Period and Applicant Suitability constituted within-subjects factors.

Procedure. The procedure was similar to that of Experiment 1. At the beginning of the study, subjects were asked to imagine themselves personnel managers in attendance at the hypothetical A-1 Corporation's,
"1982 Meeting of Managers". All experimental tasks were embedded in exercises for the meeting, with all materials presented on corporate stationery.

In a second change, subjects received six applications completed by six male applicants, as opposed to the eight applicants of both sexes employed in Experiment 1. Sex of applicants was held constant in an effort to eliminate any potential confounding effects of sex bias.

Attached to application forms was an "interviewer summary sheet", a checklist of 24 interviewee behaviors that might have occurred during the interview. Items were selected from Hakel and Dunnette's (1970) Checklists for Describing Job Applicants. Half of the items were favorable in their description of the behavior and half were unfavorable. Favorability of the items had been determined by Hakel and Dunnette in their research conducted with actual interviewers. Ten items were "checked" for each applicant by a hypothetical interviewer on the basis of an interview with the applicant assumed to have been previously conducted. The proportion of favorable versus unfavorable items checked for each applicant was varied to correspond with the applicant's level of suitability for the job. For instance, nine favorable items and one unfavorable item were checked for the most suitable applicant (a sample is presented in Appendix C).

In order to collect valence measures of group process, all groups were audiotape-recorded during the first discussion task. It will be recalled that during this task, Relevant Discussion condition groups discussed the job applicants and Irrelevant Discussion condition groups discussed an irrelevant topic. Although valence measures were only obtained from Relevant Discussion groups, groups from both conditions
were audiotape-recorded to prevent any potential confound of
tape-recording on postdiscussion measures. All groups were instructed
to attempt to ignore the tape-recording. They were told the recording
was for research purposes only, to be heard only by the experimenters.

Two tasks were added to the procedure employed in Experiment 1.
Following the final rating of the applicants, all groups engaged in a
second discussion task, utilized as a distractor. Subjects were
provided hypothetical personnel cases, instructed to read and discuss
them as a group for 20 minutes, and generate a group consensus solution
for each case situation (a sample case is included in Appendix D).
Groups were not tape-recorded during this task.

Following the distractor task, subjects were given a set of
materials that included an accuracy questionnaire (Appendix E), an
"important characteristics" ranking measure (Appendix F), and a
questionnaire requesting demographic information. When the final
measures were completed, all materials were collected, and the "1982
Meeting of Managers" adjourned. Subjects were debriefed shortly
thereafter.

Dependent measures. Subjects were asked to evaluate applicants on
the same six rating dimensions employed in Experiment 1. A composite
measure of responses to the "qualifications", "invite", and "hire"
dimensions was again formed on the basis of extremely high average
interitem correlations on prediscussion and postdiscussion responses
(coefficient alphas were .98 and .99, respectively).

The accuracy questionnaire consisted of two separate measures: an
applicant accuracy measure and a job accuracy measure. The applicant
accuracy measure was a checklist of characteristics of each applicant.
The measure was constructed such that some characteristics accurately described the applicant while others were false. Subjects were asked to accurately recall which characteristics were descriptive of each applicant. Characteristics were taken directly from the application forms and "interviewer summary sheets" previously examined by each subject. The subject was told the number of correct responses for each applicant in an effort to eliminate over- and under-responding.

The job accuracy measure presented subjects with a checklist of 14 job responsibilities, seven of which were listed as required responsibilities of the Sales Supervisor Trainee in the job description initially given to the subjects. Subjects were asked to recall which seven of the 14 responsibilities were required of the Supervisor Trainee job.

The "important characteristics" ranking measure asked subjects to rank-order from 1 (very important) to 9 (unimportant) nine applicant qualities on their importance for the Supervisor Trainee job. Qualities included "social competence", "high energy", "willingness to transfer", "ability to make accurate decisions under stress", "intelligence", "ambition", "humor", "competitiveness", and "varied job experience".

The valence of each applicant was measured as the sum of arguments generated by group members favorable to the applicant minus the sum of arguments unfavorable to the applicant. Valence measures on each applicant were collected for each Relevant Discussion condition group from audiotapes of the groups' discussions. The number of favorable and unfavorable arguments toward each applicant was coded for each group three independent times by the experimenter. The average of the three
nodings was employed as the valence measure on each applicant for each group.

Results

Preliminary analyses were conducted to determine whether subjects from the two university populations differed in their ratings of job applicants. In no instance were statistically significant differences found. Thus, in all analyses responses were pooled across subject population.

Manipulation check. Analysis of variance of the prediscussion ratings of applicant "experience" and "education and training" verified that the Applicant Suitability manipulation was successful in its intended purpose. A main effect was highly significant for both "experience", \( F(5,390) = 229.70, p < .001 \), and "education and training", \( F(5,390) = 317.50, p < .001 \). Mean ratings for "experience" were 7.68, 7.11, 6.95, 5.52, 3.48, and 2.55 for applicants manipulated as 1 through 6 in suitability for the job. Mean ratings for "education and training" were 8.00, 7.64, 7.24, 5.39, 3.18, and 2.35 for applicants 1 through 6.

Polarization. As in Experiment 1, a linear trend analysis was performed to test the hypothesized rating shifts on the high and low suitability applicants. Trend coefficients \( (5,3,1,-1,-3,-5) \) were applied to the six levels of Applicant Suitability as previously described. Table 3 presents the mean shift score ratings for both Group Discussion conditions on both rating dimensions. Table 3 shows that obtained mean shifts closely resembled the polarization pattern for the "composite" dimension while "like" showed few of the predicted shifts.
Table 3
Mean Shift-score Ratings Grouped by Applicant Suitability
and Group Discussion Condition: Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Applicants</th>
<th>Relevant Discussion</th>
<th>Irrelevant Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension</td>
<td>Suitability a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;composite&quot;</td>
<td>1</td>
<td>.33</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.48</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.55</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.11</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-.05</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-.35</td>
<td>-.17</td>
</tr>
<tr>
<td>&quot;like&quot;</td>
<td>1</td>
<td>-.08</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.14</td>
<td>-.08</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.56</td>
<td>-.02</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.09</td>
<td>-.43</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-.08</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-.41</td>
<td>-.10</td>
</tr>
</tbody>
</table>

Note. Shift-scores are computed from postdiscussion minus prediscussion ratings. n = 11 per condition.

a1 = high suitability to 6 = low suitability.
Univariate tests of the linear trend of Applicant Suitability demonstrated a highly significant linear trend in the Relevant Discussion condition for the "composite" dimension, $t(10) = 4.84, p < .001$, while no such trend was demonstrated for the "like" dimension. Neither dimension demonstrated linear trends in the Irrelevant Discussion condition. In the test of the polarization hypothesis, the interaction of Applicant Suitability(linear) with Group Discussion condition approached significance for the "composite" dimension, $F(1,20) = 4.04, p = .06$.

Reliability. Intraclass correlation coefficients were computed for the "composite" dimension and for individual rating dimensions, "qualifications", "invite", "hire", and "like". Correlations were computed from independent group responses, as described in Experiment 1. Intraclass correlations for prediscussion and postdiscussion ratings in Relevant and Irrelevant Discussion conditions were identical on each rating dimension, and were $.99$, based on $11$ independent responses for each rating period within each condition. Such high intraclass correlations indicated almost perfect interrater agreement in both Group Discussion conditions and both Rating Periods on all dimensions.

Accuracy. To examine the accuracy with which subjects recalled applicant characteristics, a score was computed on the applicant accuracy measure for each subject on each applicant. The score measured the number of correct responses made by the subject. Individual scores were averaged within each group, and group scores used as the unit of analysis. Mean applicant accuracy scores for each applicant and Group Discussion condition are presented in Table 4.
Table 4

Mean Applicant Accuracy Scores Grouped by Applicant Suitability
and Group Discussion Condition

<table>
<thead>
<tr>
<th>Applicant Suitability</th>
<th>Relevant Discussion</th>
<th>Irrelevant Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.25</td>
<td>26.28</td>
</tr>
<tr>
<td>2</td>
<td>25.42</td>
<td>23.89</td>
</tr>
<tr>
<td>3</td>
<td>23.66</td>
<td>21.69</td>
</tr>
<tr>
<td>4</td>
<td>22.83</td>
<td>22.09</td>
</tr>
<tr>
<td>5</td>
<td>23.25</td>
<td>22.05</td>
</tr>
<tr>
<td>6</td>
<td>25.31</td>
<td>26.70</td>
</tr>
</tbody>
</table>

Note. Maximum score = 35. $n = 11$ per condition.

$a1 = high suitability to 6 = low suitability.$
Analysis of variance of applicant accuracy scores revealed a strong main effect of Applicant Suitability, $F(5,100) = 26.90$, $p < .001$. Means on the accuracy scores for applicants 1 through 6 in suitability (from high to low) were 26.27, 24.65, 22.67, 22.46, 22.65, and 25.50, respectively. This ordering of means suggests that highly suitable applicants and those low in suitability were recalled significantly better than moderately suitable applicants.

As the initial interest was in differences between Relevant and Irrelevant Discussion conditions on accuracy of recalling the six applicants, the interaction of Applicant Suitability with Group Discussion condition was also examined. The interaction was not found to be statistically significant. A main effect of Group Discussion was also not significant. It appears that discussion of the applicants had no effect on recall of their characteristics.

Accuracy scores were broken down to examine "differential accuracy" in recall of applicant characteristics. Differential accuracy was examined to determine whether a relationship existed between Applicant Suitability and the type of item (favorable or unfavorable in its description of the applicant) likely to be recalled. The correct number of responses in recall of favorable applicant characteristics was subtracted from the correct number of responses on "unfavorable" items to create a differential accuracy score for each subject. A strong main effect of Applicant Suitability on differential accuracy, $F(5,100) = 59.74$, $p < .001$, demonstrated that unfavorable items were recalled significantly better for highly suitable applicants than low suitability applicants while favorable items were recalled better for applicants low in suitability. Means for applicants 1 through 6 in suitability (1 =
highly suitable) were -1.55, -2.91, 1.65, .18, 3.19, and 3.98, respectively. This effect did not show a significant interaction with Group Discussion condition. The ordering of means was very similar in both conditions.

Job accuracy scores were computed as the number of correct responses for each subject. Individual scores were averaged within each group and group scores used as the unit of analysis. Analysis of variance of job accuracy scores failed to reveal a statistically significant effect of Group Discussion condition. Discussion of the applicants appeared to have no effect on recall of job characteristics.

**Ranking of important characteristics.** The coefficient of concordance was employed to test differences between Group Discussion conditions on the rank-ordering of applicant qualities in importance for the job. The coefficient of concordance is an index of the extent to which individuals agree in their rankings. The coefficient is related to the average intercorrelation between the rankings assigned by individuals (Winer, 1971). It was hypothesized that Relevant Discussion groups would demonstrate more consensus than Irrelevant Discussion groups in their rankings. The Relevant Discussion condition produced a coefficient of .68 while Irrelevant Discussion groups produced a coefficient of .64. Although Relevant Discussion groups demonstrated slightly more agreement than Irrelevant Discussion groups, the difference is too small to be of practical significance.

**Valence.** Table 5 presents valence measures on each applicant for each group within the Relevant Discussion condition, grouped by Applicant Suitability. In examination of the valence measures, it is apparent that the hypothesized relationship between Applicant
Table 5
Valence Measures for each Relevant Discussion Condition Group
Grouped by Applicant Suitability

<table>
<thead>
<tr>
<th>Group</th>
<th>Applicant Suitability</th>
<th>Valence</th>
<th>Group</th>
<th>Applicant Suitability</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-3</td>
<td>7</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td></td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td></td>
<td>5</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td></td>
<td>6</td>
<td>-10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>18</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td></td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-6</td>
<td></td>
<td>4</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-24</td>
<td></td>
<td>5</td>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-17</td>
<td></td>
<td>6</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td></td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-2</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-10</td>
<td></td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-15</td>
<td></td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-15</td>
<td></td>
<td>6</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>18</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-4</td>
<td></td>
<td>4</td>
<td>-5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td></td>
<td>5</td>
<td>-21</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-18</td>
<td></td>
<td>6</td>
<td>-13</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td></td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-2</td>
<td></td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-2</td>
<td></td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-4</td>
<td></td>
<td>6</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Valence measures are computed as the sum of arguments favorable to the applicant minus the sum of unfavorable arguments. \( n = 3 \) or \( 4 \) per group.

\( ^{a}1 = \text{high suitability to 6 = low suitability.} \)
Suitability and valence found mixed support at the group level. While groups 4, 7, 10, and 11 closely resembled the hypothesized relationship, groups 6, 8, and 9 showed greater departure from the predicted pattern. To examine the hypothesized relationship across groups, a linear trend test was conducted to test the effect of Applicant Suitability on valence. Mean valence measures at each level of Applicant Suitability corresponded closely with the predicted linear pattern. Means for Applicants 1 through 6 in suitability (1 = highly suitable) were 7.09, 2.00, 3.45, -4.18, -5.18, and -8.36. Trend coefficients (5, 3, 1, -1, -3, -5) were applied to mean valence measures, and a significant linear trend of Applicant Suitability was demonstrated, $t(10) = 4.68, p < .01$. This indicated, across groups, the highest positive valence was associated with the most suitable applicant, the highest negative valence was associated with the least suitable applicant, and the valence measures for the middle levels of Applicant Suitability closely corresponded with the predicted pattern.

**Discussion**

The present study provided limited support for the prediction that group discussion polarizes evaluations of job applicants. While a significant shift in ratings was found on the "composite" measure for groups that discussed the applicants, the hypothesized difference in polarization between Discussion conditions fell short of statistical significance.

As both Group Discussion conditions demonstrated almost perfectly reliable ratings across all dimensions, neither hypothesis regarding an increase in interrater agreement following discussion was supported. Reliabilities of .99 indicated near perfect agreement among raters on
each dimension. It is apparent that the manipulation of stimuli as six levels of Applicant Suitability was so strong that it served to eliminate variance due to rater differences. This elimination of variance may help explain the weak demonstration of polarization found in the present experiment.

Near perfect agreement among raters on the prediscussion ratings suggests that applicants were perceived in a very similar way by all raters. It may be assumed there was little ambiguity of Relevant Discussion raters' beliefs, prior to discussion, in their judgment of applicants' suitability for the job. Some evidence has been found to suggest that ambiguity of initial beliefs about a stimulus object facilitates polarization of judgments about that object (Tesser & Cowan, 1977). Tesser and Cowan (1977) found greater thought-induced polarization (no group discussion was involved) when subjects' initially inconsistent beliefs were ambiguous than when their beliefs were unambiguous. Lack of ambiguity of initial beliefs about applicants may have left raters with little "fuel" for discussion and, in turn, may have eliminated any effect of discussion on subsequent evaluations.

A study by Kaplan (1977), discussed earlier, may also provide useful insight into weak polarization findings. Kaplan found postdiscussion ratings demonstrated greater polarization when information shared in the discussion was nonredundant than when it was redundant. It is likely that information shared in the discussion of applicants in the present study was redundant, as there was almost perfect interrater agreement prior to discussion on ratings of applicants. If Kaplan's findings generalize to the present personnel selection task, redundant arguments
within the selection groups might have inhibited any potential polarization of judgments.

Intraclass coefficients of .99 indicated large rating variance due to applicant differences and almost no variance due to rater differences. Such small error variance made it extremely difficult to detect any differences between Group Discussion conditions on any dependent variable. Consequently, the lack of differences between Group Discussion conditions on recall of applicant characteristics and recall of job characteristics was attributable to high interrater agreement. It was expected that increased interrater agreement due to group discussion would cause differences among Group Discussion conditions in accuracy of recall of applicant and job characteristics, although the direction was not predicted. As there were no differences in interrater agreement between conditions, it was not surprising that Discussion conditions did not differ in responses to these accuracy measures.

The significant relationships reported between Applicant Suitability and accuracy of recall of applicant characteristics were not predicted, but were very interesting, especially considering the strength of the effects. Overall, highly suitable and low suitability applicants were recalled significantly better than moderate applicants. Further, the type of items recalled was found to be affected by applicant suitability. Unfavorable information was recalled better for highly suitable applicants while applicants low in suitability were recalled better for their favorable characteristics. This is an unusual finding, in light of Gordon's (1970) Differential Accuracy Phenomenon (DAP). The DAP predicts greater rating accuracy when ratee behavior is performed incorrectly than when behavior is correctly performed. Results of the
present study suggest the DAP may depend on raters' initial perceptions of the ratee's abilities. Perhaps if level of ratee overall performance was manipulated in initial research on the DAP, the opposite effect predicted by the DAP would be found for poor performing ratees. Raters may be more accurate when rating correctly performed behavior when they expect ratees to perform poorly.

Contrary to the initial prediction, there were no significant differences between Discussion conditions on a rank-ordering of applicant qualities in their importance for the job. In light of the lack of between-subjects differences on other dependent variables, the similarity of rankings between conditions is not surprising.

As a method of assessing group process, Hoffman's (1979a) valence technique proved moderately successful in predicting the marginal polarization effect found on the "composite" dimension. It was of course difficult to draw conclusions about the relationship between valence and polarization shifts, as polarization shifts were not conclusively demonstrated in this experiment. Although the predicted relationship between valence and Applicant Suitability was not perfect when measured at the group level, mean valence measures closely corresponded to levels of Applicant Suitability. These results suggest that Hoffman's valence technique for measuring group process may be useful in predicting polarization response shifts. If Hoffman's technique is found to be useful in predicting polarization shifts in future research, its simplicity and ease of recording is sure to make it an attractive technique for measurement of group process in polarization research.
In conclusion, the lack of influence of Group Discussion condition on dependent measures appears to be due to almost perfect rater agreement regarding applicants' suitability for the job. It is clear that Applicant Suitability was so strongly manipulated that there was little rater ambiguity regarding beliefs about qualifications and other job relevant characteristics of the applicants, and variance due to rater differences was effectively eliminated.

GENERAL DISCUSSION

The two studies presented set out to examine the effect of group discussion on evaluation of job applicants. Experiment 1 demonstrated a polarizing effect of group discussion on evaluations of job applicants that approached statistical significance. In addition, groups that discussed applicants demonstrated greater interrater agreement, following discussion, than groups that did not discuss applicants.

An attempt at replication of these findings, using a larger sample, failed to demonstrate significant polarization of ratings, although a marginal shift in the predicted direction was again found. The differences in interrater agreement found in Experiment 1 were not replicated. Experiment 2 further attempted to examine the effect of group discussion on accuracy of assessment. None of the three measures employed to examine assessment accuracy showed differences due to group discussion.

Although Experiment 2 failed to demonstrate stronger polarization effects than those found in Experiment 1, insight into the effect of group discussion on applicant evaluation may be gained in consideration of the combined results of both studies. When tests of the polarization hypothesis are combined from the two independent experiments, an overall
probability statement may be made and tested for its statistical significance. Probabilities from the two polarization tests on the "composite" dimension ($p = .08$ and $.06$ in Experiments 1 and 2, respectively) may be combined and this combination distributed as a chi square. The chi square tests the null hypothesis that the combined results of the two polarization tests are no different than zero (Winer, 1971). In the present combination of experiments, the combined results provided statistically significant support for the polarization hypothesis on the "composite" dimension, $\chi^2(4) = 10.69, \ p < .025$.

Of course the statistical test of combined probabilities merely supports that which may be gleaned from a less exacting examination of the data. Two tests of the same hypothesis with results similar in magnitude, but of marginal statistical significance, indicate that the hypothesis requires a more powerful test than provided in either experiment. Future experiments will need to provide a larger number of groups to detect polarization differences between Discussion conditions.

It might be concluded, in examination of results obtained in Experiment 2, that differences between applicants were so strong that they eliminated differences in evaluations of applicants due to the effect of group discussion. The pool of applicants differed in their qualifications for the job from extremely suitable to very low in suitability and no two applicants shared a similar level of qualifications. Choices for all raters, were, as such, very clear cut, and there was little room for disagreement in the groups that discussed applicants. In this light, it is not surprising that group discussion would have little or no effect on subsequent judgments of applicants. This conclusion finds support in the literature, as previously
discussed. Subjects are likely to be more ambiguous in their beliefs about the suitability of applicants when they are from a homogeneous applicant pool. Ambiguity of initial beliefs has been shown to facilitate polarization (Tesser & Cowan, 1977). Further, discussion of a homogeneous applicant pool is likely to contain more varied arguments than discussion of a highly similar applicant pool. Redundancy of arguments has been demonstrated to inhibit polarization of judgments (Kaplan, 1977).

To discuss implications of these findings for personnel selection, it is clear that the homogeneous applicant pool employed in Experiment 1 provided a more realistic sample of applicants than the heterogeneous group evaluated in the second experiment. In an actual selection decision, it is likely that a heterogeneous pool of applicants might be narrowed through elimination of the least suitable, prior to any need for group discussion. It is likely discussion would be necessitated more frequently with an applicant pool narrowed to a group of similar qualifications. Of course, if the selection ratio was very small (i.e., a similar number of applicants to the number of job openings), then discussion of the entire applicant pool might be required.

Practical implications may be derived from findings of polarization of evaluations of similar applicants if, indeed, they appear robust in subsequent research. One implication regards the effect of group discussion on bias in hiring. If a majority of group members hold pre-existing biases against an applicant from a protected group, these initial discriminatory tendencies may be heightened following group discussion. The postdiscussion decision would likely reflect greater bias against the applicant than if group discussion had not occurred.
Conversely, if an applicant from a protected group is interviewed and discussed by group members with a majority of unbiased attitudes then the applicant would stand a greater chance for consideration for the job, following group discussion, than if he or she had been interviewed by one of the few biased group members. Group discussion would serve to attenuate discriminatory tendencies of those biased members.

A second implication from these findings is that it may be possible to predict the outcome of group discussion of applicants from prediscussion ratings. Sackett and Wilson (1982) found the outcome of the consensus discussion predicted by the mean of the individual prediscussion ratings for 93.5% of all ratings made in an assessment center. It may be that discussion of an applicant by group members in order to reach consensus may be unnecessary to the final decision. Sackett and Wilson suggested that the consensus discussion component of assessment center technology may be replaced with a mechanical decision rule combining assessors' initial judgments.

Evidence was presented to suggest group discussion may improve intrarater reliability of the selection decision. Group discussion may also have an effect on the validity of the selection decision. Arvey and Campion (1982) proposed, on the basis of a few related studies (e.g. Landy, 1976; Rothstein & Jackson, 1980), that group discussion might improve validity. Although the evidence they presented to support that contention appears weak to the present reviewer, the relationship between group discussion and validity may certainly be an important topic for future research.

It is clear that implications for personnel selection are numerous if future research provides further support for applicability of the
polarization phenomenon. As personnel selection includes a group process component in many organizations, this should certainly be an important avenue of study for personnel researchers. It is hoped the studies presented here provide some insight as an initial investigation of the effect of group discussion on evaluation of job applicants.
SALES SUPERVISOR TRAINEE JOB DESCRIPTION
SUPERVISOR TRAINEES

A large chain of discount department stores, owned by the A-l Corporation, seeks applicants for the position of department supervisor. After successful completion of a six-month training period, applicants will be assigned to one of 30 different departments (e.g., hardware, lawn and garden, etc.) in a store in the Southwest Region (Texas, Oklahoma, New Mexico). A department supervisor is responsible for 15 to 30 sales personnel and a clerical staff. The supervisor must oversee stock inventories, and the setting-up of window displays, as well as other day-to-day sales operations of a department. The supervisor is responsible for the evaluation, motivation, training, and disciplining of employees in the department. Furthermore, the supervisor must frequently meet with store managers and other department supervisors to decide and interpret store policy. Frequently, it will be required to work nights and weekends. A college degree is required. Applicants should be willing to transfer to stores in other regions. In general, the position requires energy, initiative, ambition, the ability to make accurate decisions under stress, and a high degree of social competence.
APPENDIX B

COMPLETED APPLICATION
Application for Employment

A-I CORPORATION

An Equal Opportunity Employer
<table>
<thead>
<tr>
<th><strong>PERSONAL DATA</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NAME</strong></td>
<td>Burkett Allen Scott</td>
</tr>
<tr>
<td><strong>LAST NAME</strong></td>
<td></td>
</tr>
<tr>
<td><strong>FIRST NAME</strong></td>
<td></td>
</tr>
<tr>
<td><strong>MIDDLE NAME</strong></td>
<td></td>
</tr>
<tr>
<td><strong>DATE</strong></td>
<td></td>
</tr>
<tr>
<td><strong>PRESENT SCHOOL ADDRESS</strong></td>
<td>2200 W. First St., Madison, Wisconsin 53706</td>
</tr>
<tr>
<td><strong>PERMANENT HOME ADDRESS</strong></td>
<td>800 Jeffries Ct., St. Paul, Minnesota 55458</td>
</tr>
<tr>
<td><strong>SEX</strong></td>
<td>Male</td>
</tr>
<tr>
<td><strong>HEIGHT</strong></td>
<td>5'11&quot;</td>
</tr>
<tr>
<td><strong>WEIGHT</strong></td>
<td>183 lbs.</td>
</tr>
<tr>
<td><strong>PHONE</strong></td>
<td>524</td>
</tr>
<tr>
<td><strong>MARITAL STATUS</strong></td>
<td>Single</td>
</tr>
<tr>
<td><strong>NUMBER OF CHILDREN</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>OTHER DEPENDENTS</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>PHYSICAL LIMITATIONS</strong></td>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>EDUCATION</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TYPE OF SCHOOL</strong></td>
<td></td>
</tr>
<tr>
<td><strong>NAME AND LOCATION</strong></td>
<td></td>
</tr>
<tr>
<td><strong>DATES ATTENDED</strong></td>
<td>YEARS COMPLETED</td>
</tr>
<tr>
<td>Prep or High School</td>
<td>St. Andrews Academy St. Paul, Minnesota</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>College or University</td>
<td>U. of Wisconsin Madison, Wisconsin</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SCHOOL</strong></th>
<th><strong>MAJOR FIELD OF STUDY</strong></th>
<th><strong>MINOR FIELD</strong></th>
<th><strong>LIST COURSES AND NUMBER OF CREDITS IN MAJOR FIELD OF STUDY</strong></th>
<th><strong>COURSES OF GREATEST INTEREST</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>College or University</td>
<td>Business Management</td>
<td></td>
<td>Fund. of Mgmt. - 4 Sales &amp; Marketing 8</td>
<td>Management Philosophy Public Relations</td>
</tr>
<tr>
<td>Graduate School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Finance - 8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SCHOLASTIC HONORS (Including scholarships, honorary societies, and publications):**

**CAMPUS ACTIVITIES -- OTHER THAN ATHLETICS (Fraternity or sorority, clubs, class organizations, offices, etc.):**

**ATHLETICS (Specify captaincy; encircle varsity sports):**
- Jr. Varsity Ice Hockey - freshman year

**HOBBIES AND OUTSIDE INTERESTS:**
- Ice sports, photography

**PERCENT OF COLLEGE EXPENSES EARNED?** approximately 60%

**HOW EARNED?** Part-time jobs Summer employment

**APPROXIMATE NUMBER OF HOURS PER WEEK WORKED DURING SCHOOL YEAR?** 20
**Your Three Most Important Jobs:**

<table>
<thead>
<tr>
<th>Name &amp; Address of Employer</th>
<th>Dates</th>
<th>Nature of Your Work</th>
<th>Monthly Salary</th>
<th>Reason for Leaving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dayton's Dept. Store</td>
<td>9/780 - 9/80</td>
<td>Sales</td>
<td>$625</td>
<td>Return to school</td>
</tr>
<tr>
<td>Chandler Emporium</td>
<td>6/79 - 9/79</td>
<td>Sales</td>
<td>$575</td>
<td>Return to school</td>
</tr>
<tr>
<td>U. of Wisconsin</td>
<td>9/79 - 4/80</td>
<td>Record keeping</td>
<td>$300</td>
<td>Another campus job offer</td>
</tr>
</tbody>
</table>

**Work Experience**

**Nature of Your Work**

**Monthly Salary**

**Reason for Leaving**

---

**List Three Faculty Members Who Are Well Acquainted With Your Scholarship and General Ability:**

<table>
<thead>
<tr>
<th>Name</th>
<th>Department</th>
<th>Address</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Morton Heinrich</td>
<td>Business Admin.</td>
<td>U. of Wisconsin</td>
<td>X4200</td>
</tr>
<tr>
<td>Dr. Daniel H. Smith</td>
<td>Business Admin.</td>
<td>U. of Wisconsin</td>
<td>X4212</td>
</tr>
<tr>
<td>Dr. Erika Danby</td>
<td>Economics</td>
<td>U. of Wisconsin</td>
<td>X1454</td>
</tr>
</tbody>
</table>

**Indicate Type or Types of Work in Which You Are Interested:**

Mainly marketing and sales - related opportunities will be considered however

**Do You Have Any Preference Regarding Locations? If so, specify:**

None

**Date Available for Employment:**

Immediately upon notice

**Days of the Week, or Specific Dates, Convenient for Further Interviewing:**

Any time

---

**Signature of Applicant**
APPENDIX C

INTERVIEWER SUMMARY SHEET
INTERVIEW SUMMARY SHEET

Every applicant for the Supervisor Trainee position has been interviewed by an interviewer in the A-l Corporation Personnel Department. Following each interview, the following checklist is completed by the interviewer. The interviewer describes the applicant as accurately as possible by checking which of the items apply to the applicant.

Applicant:  **ALLEN S. BURKETT**

- [ ] Smiled frequently.
- [ ] Asked same question over, or asked question already asked.
- [ ] Has worked in many different occupations.
- [ ] Kept the conversation going.
- [ ] Is over-qualified for the job he's being interviewed for.
- [ ] Interrupted you while you were speaking.
- [ ] Scored high on all screening tests.
- [ ] Slouched in chair.
- [ ] Voice carries well.
- [ ] Appeared ill-at-ease during the entire interview.
- [ ] Spoke forcefully.
- [ ] Speech was concise.
- [ ] Frequently used slang.
- [ ] Says he drives himself steadily (doesn't work by fits and starts).
- [ ] Says he is very independent.
- [ ] Says he likes taking responsibility.
- [ ] Says he would be willing to relocate for his job.
- [ ] Says he does not care about what kind of work he will be doing.
- [ ] Pressed for details about the job.
- [ ] Says he is dissatisfied with the co-workers on his present job.
- [ ] Says he likes to take the lead in group activities.
- [ ] Says he can carry out plans assigned by other people.
- [ ] Says he would rather not take chances or run risks.
- [ ] Says he makes friends easily.
- [ ] Talked too fast.
- [ ] Says he dislikes criticism.
APPENDIX D

PERSONNEL CASE
A-1 CORPORATION CASE IV.

THE "CLOAK AND DAGGER" SUPERVISOR

Mr. Charles Wright is responsible for six managers and approximately four hundred employees at the A-1 Corporation. He is considered the top person in his field, having developed the most copied system of billing in the industry. By employing "top notch" managers, he has been able to organize his various departments with a high degree of success. He established standard procedures for everything, including marking the floor to designate furniture location. All procedures were checked by a separate control system to insure that policies were being followed. Mr. Wright had been successful in improving department performance significantly while substantially reducing costs in five of his six departments. His employee turnover was average for the industry. However, the sixth department seemed to be a complete failure. Mr. Wright believed that the department manager was not capable of carrying out his directives, so he replaced the manager; in fact, several managers had been tried before Neal Bell was selected. Bell's office performs various duties, primarily accounting tasks, and employs sixty people.

When Neal arrived, Mr. Wright gave him a week of indoctrination on his method of operation. Mr. Wright believed everyone should start work at 8 a.m., the start of the office day. This meant that employees should arrive ten minutes before 8 a.m. to open their desks, uncover adding machines, typewriters, etc., and put their work on their desks. Mr. Wright would randomly check departments to insure that everyone was on time. Neal's department always started late by Mr. Wright's standards. Mr. Wright would note any employee who arrived at 8 a.m. or later and would ask Neal to interview him or her. Then Neal was to present Wright the reasons for the employee's lateness as well as evidence that the employee had been interviewed and reminded to be "on time."

At various times during the day, Mr. Wright would inspect his departments. He would talk about department performance and operating problems and would say "hello" to the employees. During these visits he would also inspect some of the completed work of each department, again noting deficiencies and requiring a report on them. Mr. Wright suspected that employees in Neal's department threw away work to maintain the appearance that they were not "backlogged," a condition he attributed to employee inefficiency. He was known to have discharged employees for this reason.

During one of his visits, about a month after Neal had begun work, while the office was out for a coffee break, Mr. Wright selected a wastebasket and, after carefully placing a newspaper on Neal's desk, dumped and searched the contents of the wastebasket. He found evidence that work was being discarded. After noting the employee's name, he returned the wastebasket and contents. That evening he obtained a master key and searched all of Neal's employees' lockers. Some lockers contained work that was not included in the reported backlog. The next morning Mr. Wright said he wanted all of the offending employees discharged immediately. Neal was very much upset and suggested that he would be better able to run his department if Mr. Wright would stop interfering and performing "cloak and dagger" work. Mr. Wright retaliated that all of the rest of his departments ran smoothly using these techniques, and Neal should return to his office and carry out his orders.
When Neal returned to his office, he was greeted by five employees who said they represented the rest of the group and would like to talk to him. The employees wanted to know for whom they worked, Neal Bell or Charles Wright. They complained that for three years they had been "hounded" by Mr. Wright and said the whole group agreed the job was not worth the harassment and, unless something happened to improve the situation, everyone would resign at one time.
APPENDIX E

ACCURACY QUESTIONNAIRE
I. The following series of questions is an assessment of the extent to which you are able to accurately recall characteristics of each job applicant reviewed.

On the basis of information recalled from the applicant's application form and from the interviewer's summary sheet, please indicate which of the following characteristics is descriptive of the applicant. Check "True" if the description fits the applicant. There are ______ "True" statements regarding this applicant.

1. Majored in Sales or Marketing. 
   TRUE
2. Dean's List was not indicated in his application as an academic honor received. 
   ____
3. Has teaching or training experience. 
   ____
4. Was a member of a varsity or junior varsity sports team. 
   ____
5. Earned at least 25% of college expenses. 
   ____
6. Graduated magna cum laude from college. 
   ____
7. College grade point average was below 1.75. 
   ____
8. Held at least one summer job during college. 
   ____
9. Has management experience. 
   ____

The following questions are based on the interviewer summary sheet.

10. Scored high on all screening tests. 
    ____
11. Has worked in many different occupations. 
    ____
12. Kept the conversation going. 
    ____
13. Is overqualified for the job he's being interviewed for. 
    ____
    ____
15. Says he is very independent. 
    ____
16. Says he dislikes criticism.

17. Pressed for details about job.

13. Says he makes friends easily.


20. Says he does not care about what kind of work he will be doing.

21. Asked same question over, or asked question already asked.

22. Frequently used slang.

23. Says he is dissatisfied with the co-workers on his present job.

24. Says he would rather not take chances or run risks.

25. Says he drives himself steadily (doesn't work by fits and starts).

26. Says he likes taking responsibility.

27. Interrupted you while you were speaking.

23. Voice carries well.

29. Says he would be willing to relocate for his job.

30. Spoke forcefully.

31. Says he can carry out plans assigned by other people.

32. Appeared ill-at-ease during the entire interview.

33. Speech was concise.

34. Says he likes to take the lead in group activities.

35. Talked too fast.
II. The following series of questions is an assessment of the extent to which you are able to accurately recall responsibilities of the supervisor trainee job as outlined in the job description.

On the basis of job responsibilities recalled from the Supervisor Trainee job description, please indicate which of the following job responsibilities are required in the Supervisor Trainee job. Check "Yes" if it is required in the job. Seven of the following 14 listed job responsibilities are required in the job and should be checked "Yes".

YES

1. Motivates department employees.
2. Oversees major customers' accounts.
3. Meets with store managers and other department supervisors to decide and interpret store policy.
4. Manages 50 - 60 sales personnel.
5. Trains department employees.
6. Oversees stock inventories.
7. Insures all customer billing information is accurately processed by the computer.
8. Sponsors biannual social functions for department employees and their families.
10. Recommends pay increases and promotions for department employees.
11. Oversees setting up of window displays.
12. Disciplines department employees.
13. Processes department employees' insurance claims.
14. Writes advertisements for department products.
APPENDIX F

"IMPORTANT CHARACTERISTICS" RANKING MEASURE
### III. Rank order from 1 (very important) to 9 (unimportant) applicant qualities on their importance for the Supervisor Trainee job:

<table>
<thead>
<tr>
<th>Quality</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>High energy</td>
<td></td>
</tr>
<tr>
<td>Ability to make accurate decisions</td>
<td></td>
</tr>
<tr>
<td>under stress</td>
<td></td>
</tr>
<tr>
<td>Social competence</td>
<td></td>
</tr>
<tr>
<td>Intelligence</td>
<td></td>
</tr>
<tr>
<td>Ambition</td>
<td></td>
</tr>
<tr>
<td>Humor</td>
<td></td>
</tr>
<tr>
<td>Competitiveness</td>
<td></td>
</tr>
<tr>
<td>Varied job experience</td>
<td></td>
</tr>
<tr>
<td>Willingness to transfer</td>
<td></td>
</tr>
</tbody>
</table>
Reference Notes


References


Bishop, G. D., & Myers, D. G. Informational influence in group discussion. Organizational Behavior and Human Performance, 1974, 12, 92-104.


Hoffman, L. R. Applying experimental research on group problem solving to organizations. *Journal of Applied Behavioral Science, 1979, 375-391*. (b)


Myers, D. G. Summary and bibliography of experiments on group-induced response shift. *Catalog of Selected Documents in Psychology*, 1973, 3, 123.


Simas, K., & McCarrey, M. Impact of recruiter authoritarianism and applicant sex on evaluation and selection decisions in a recruitment interview analogue study. *Journal of Applied Psychology*, 1979, 64, 483-491.


Tucker, D. H., & Rowe, P. M. Relationship between expectancy, causal attributions, and final hiring decisions in the employment interview. *Journal of Applied Psychology*, 1979, 64, 27-34.


