INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

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

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Is there judgment bias in the assessment center method?

Hayes, Theodore Laurance, Ph.D.
Rice University, 1990
RICE UNIVERSITY

IS THERE JUDGMENT BIAS IN THE ASSESSMENT CENTER METHOD?

by

THEODORE L. HAYES

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

APPROVED, THESIS COMMITTEE

Barbara B. Gaugler, Director
Assistant Professor of Psychology

William C. Howell
Professor of Psychology

Richard F. Martell
Assistant Professor of Psychology

Stephen L. Klineberg
Professor of Sociology

Houston, Texas

April, 1990
Is There Judgment Bias In
The Assessment Center Method?

by

Theodore L. Hayes

ABSTRACT

Recent analyses of assessment center ratings have demonstrated that assessors who are trained to make dimension-based assessments may instead base their judgments on information other than dimension performance. This study evaluated the effects of enhanced accountability to make justifiable behavioral recordings and evaluations on assessor accuracy. Specifically, it was predicted that enhanced accountability to justify ratings and behavioral observations would lead assessors to make more accurate ratings and observations, as compared to the ratings and observations made by assessors whose personal accountability was not enhanced. Results showed that when accountability was not enhanced, as predicted, assessors relied on extraneous performance information (exercises, personality evaluations) when making their overall ratings. Assessors whose accountability was enhanced used only dimension information when making overall ratings, made more efficient behavioral observations and classifications, and had higher overall rating accuracy than did assessors whose accountability was not enhanced. However, enhanced accountability did not result in significantly different overall confidence of assessors in their decisions as compared to those whose accountability had not been enhanced.
ACKNOWLEDGEMENTS

I would like to thank my committee members -- Barbara Gaugler, Chair; Bill Howell; Rich Martell; and Steve Klineberg -- and Doris Malone, for their insight, tolerance, and guidance during this process. Though the memory for my results may fade, my memory of what I've found out from them about psychology will endure.

I would also like to thank Liz Kotler, Bob Lewis, Sue Parker, Kim Donner, Lee Friedman, R.J. Harvey, Maryalice Citera, and Joan Rentsch for their support and ideas. My student assistant administrators deserve credit for meritorious service. And finally, I'd like to thank my subjects for succumbing to my experimental manipulation.

This dissertation is dedicated to my Mom.
TABLE OF CONTENTS

Introduction .......................................................1
Accountability.................................................5
Accountability in the assessment center.................6
Social judgment issues.......................................8
Categories and schemas in assessment centers..........10
Implicit performance constructs.........................15
Leadership IPC's.............................................17
Employment interview IPC's.................................17
Performance appraisal IPC's.................................19
Personalogical IPC's.........................................20
IPC implications for assessment centers...............22
The present investigation.....................................23
Hypotheses.....................................................23

Method..........................................................24
Overview.........................................................24
Subjects........................................................25
Materials........................................................27
  Video tapes.................................................27
  Assessment center dimensions.........................28
  Target ratings..............................................28
Procedure.......................................................32
Overview.........................................................32
Training session.............................................32
Experimental session.......................................33
Ratings..........................................................35
Results ................................................................. 38
  Aggregation of data .............................................. 38
  Manipulation check .............................................. 43
Hypothesis 1: Accountability effects on rating strategies ............ 50
Hypothesis 2: Accountability effects on assessor confidence ........... 59
Equation 1: Estimation of Brier scores ................................ 64
Hypothesis 3: Accountability effects on rating accuracy ............... 65
Hypothesis 4: Accountability effects on observation and classification accuracy ........................................... 68
  Observation accuracy .......................................... 68
  Classification accuracy ...................................... 70
Generalizability of the results .................................... 73
  Proportion of significant results .............................. 73
  Comparability of effect sizes .................................. 74
  Power analyses ................................................. 79
Discussion .......................................................... 82
  Accountability ................................................... 82
  Assessment center method ..................................... 84
Postscript ............................................................ 95
Conclusions .......................................................... 96
References .......................................................... 97
Appendix A .......................................................... 105
Appendix B .......................................................... 106
Appendix C ......................................................... 116
Appendix D ......................................................... 141
Appendix E ......................................................... 158
Appendix F ......................................................... 179
LIST OF TABLES

Table 1: Differences in demographic characteristics between Rice and University of Houston study participants........................................26

Table 2: Definitions of performance variables assessed in the current study.........................................................29

Table 3: Bivariate intercorrelations among rating variables.................................................................40

Table 4: Means and standard deviations of manipulation check measures..................................................45

Table 5: Canonical discriminant analyses: Predicting experimental group membership from the manipulation dependent measures..................................................49

Table 6: Hierarchical regression analysis for each condition: Prediction of the OAR...............................52

Table 7: Hierarchical regression analysis for each condition: Identification of the incremental nonnormative component in predicting the OAR.............53

Table 8: Rating variable means and standard deviations..57

Table 9: Analyses of covariance for the assessor rating models: Prediction of the OAR using duration of dormitory residency as a covariate.........................60

Table 10: Means and standard deviations of confidence ratings for each dimension and the OAR (by condition)...63

Table 11: Rating accuracy component scores......................67

Table 12: Behavioral observation and classification efficiency mean scores and standard deviations............71

Table 13: Tests of differences for obtained effect sizes between this study and average effect sizes found in the accountability literature..........................77

Table 14: Estimates of power for the obtained accountability effect.................................................80
LIST OF FIGURES

Figure 1: An example of the information processing of assessors who observe and evaluate one candidate across two exercises and four dimensions: Zedeck's (1986) hierarchical assessment center model.............14
INTRODUCTION

This study investigates judgment processes in the assessment center method, an applied human resource management technique. The assessment center method is a structured procedure that integrates multiple, structured job-related situations (exercises) and other tests and measures to evaluate a candidate's job-related skills and abilities for a variety of human resource purposes. Many assessment centers, including the one in this study, incorporate three procedures which in combination differentiate assessment centers from other human resource interventions.

The first unique procedure in the development of assessment centers is the identification of the behavioral clusters, or dimensions, of work performance that are both important and critical to carrying out work duties effectively (Thornton & Byham, 1982). These dimensions are defined so that they are specific, verifiable, and can be rated reliably (Thornton & Byham, 1982).

A second feature is the development of evaluation techniques, or exercises. Exercises are chosen and developed based on their ability to elicit behavior reflecting the target dimensions identified as important for successful performance on the job. One goal of exercise development is to sample behavior relevant to the job requirements without actually recreating the job situation.
Another goal of exercise development is to elicit as many of the relevant dimensions as possible within each exercise. Several types of exercises are used so that assessors can determine the candidate's abilities in each dimension based on the convergence of across-exercise evidence.

The third unique assessment center procedure is the use of multiple trained assessors to make performance judgments. Assessors are rigorously trained to observe, record, and integrate candidate behavior as accurately as possible. By using several trained assessors who share behavioral observations, the method procedurally restrains premature evaluation of performance while increasing the bandwidth of sampled behavior available for evaluation.

The use of a "typical" assessment center may follow these procedures. Job analysts determine the critical dimensions of performance for the target job, and subject matter experts develop exercises to elicit these dimensions. Managers or consultants are recruited and trained to serve as assessors; their training focuses on the importance of making accurate observations the use of observations in making dimension ratings. Assessors observe candidates in the assessment center. At the close of the assessment center, assessors sort their observations into dimensions and pool their observations to make dimension ratings and an overall assessment rating (OAR). The assessment center is completed once assessors arrive at a consensus rating for
each dimension and the OAR for each candidate.

The development and execution of assessment centers is a rigorous and thorough process. However, research has identified several shortcomings of the assessment center method. Exploratory factor analyses have shown that even when assessments are made on many dimensions, only two to five general factors are needed to account for most of the underlying variance in ratings (Sackett & Hakel, 1979). Confirmatory factor analyses have indicated that factor structures of dimension ratings made by different assessors within the same assessment center vary, and that these rating patterns do not reflect the dimensional structures the assessors were trained to use (Russell, 1985). Instead, evidence shows that assessment center ratings are best described by exercise factors rather than dimension factors (Bycio, Alvares, & Hahn, 1987; Sackett & Dreher, 1982; Wollowick & McNamara, 1969). Further investigations using multitrait-multimethod analyses of dimension ratings have indicated that there is high agreement among disparate dimension ratings within exercises, but low or moderate convergent validity for the same ratings across exercises (e.g., Silverman, Dalessio, & Woods, 1986), a sign of method (exercise) bias. Finally, related research (Sackett & Wilson, 1982) has demonstrated that processes such as assessor group dynamics may lead to rating bias.

In summary, though recent meta-analyses of assessment
centers have demonstrated the high criterion-related validity of the OAR for predicting performance and promotability (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Hunter & Hunter, 1984), there is also evidence that assessors may attend to, store, combine, and/or retrieve candidate performance information in ways that were not intended by the developers of assessment centers.

The current study sought to determine the extent to which assessors used both normative (dimension-related) and nonnormative (non-dimension related, e.g., exercise and personallogical) information about candidates in an assessment center. In many assessment centers (such as the current one), assessors are trained carefully to observe and use behavioral information that is salient to prespecified behavioral dimensions. Evidence that assessors have followed a normative assessment center judgment model is presented through a statistical demonstration that assessor decisions about candidates resulted only as a function of that dimension information. Conversely, a typical finding in the literature is that assessors use nonnormative assessment models which contain information other than that which they were trained to use in order to evaluate candidates.

It is the contention of this paper that assessment center development and interpretation may be improved first by determining the types of information that assessors
typically use, and next, through investigation of ways to minimize biased judgments. Uncertainty about whether judgment bias exists or how it is manifested may have implications for the defensibility of the method. It is conceivable that assessment center decisions that have adverse impact may be challenged on the grounds of biased decision-making; indeed, meta-analytic results (Gaugler et al., 1987) have already demonstrated that assessors may judge women and minorities differently than majority group members. Differing judgment standards may lead to lowered criterion-related validity through introduction of rating error. Thus, to ensure that the content and criterion-related validity of the method is sound, research should seek to address judgment issues.

Accountability

One factor that may influence assessors' judgments is their sense of accountability to use information accurately and appropriately when rating candidates (e.g., Tetlock, 1983). Tetlock (1983) defined accountability as "a special type of transmission set in which one anticipates the need not only to communicate one's opinions, but also to defend those opinions against possible counterarguments" (p. 75). People who expect to justify their views may be more vigilant information processors than those who do not need to justify their views.

Tetlock developed the rationale for the importance of
accountability by reasoning that accountability demands should improve the quality of decision-making since social interaction reflect and are facilitated by the judicious application of social skills. In his model (Tetlock, 1985), accountability is built into the social sphere to provide for regularity in performance based on shared norms, roles, etc. Individuals are accountable as role players and social participants for self-regulation and for enforcing the norms of their conduct and the conduct of others. In this way, accountability results in order and predictability in social conduct (Tetlock, 1985).

Accountability In The Assessment Center

The nature of the assessment center process may instill a sense of accountability in assessors for making high quality judgments. Assessors must make careful observations and explain these observations to others. In many assessment centers, when there are discrepancies in ratings, assessors must justify their ratings through use of behavioral evidence. This places a premium on accurate and thorough behavioral observation, classification, and judgment. Also, assessors are typically accountable to the organization either as managers or as external consultants. Organizational expectations about the quality of assessment center judgments may place assessors in the position of feeling some level of accountability.

Even though these characteristics of the assessment
center method would seem to demand accountability, this type of accountability -- standard accountability -- apparently does not lead assessors to use normative judgment strategies. Tetlock (1983, 1985) proposed a number of hypotheses as to why this may be so. He emphasized that accountability leads to improved judgment quality specifically when decision makers: (a) are made to feel that they must justify their judgments to others; and (b) when assessors are unsure of the prior expectations of those to whom they are accountable. Alternatively, accountability to one with known views results in what Tetlock (1985) described as acceptability, or thought and opinion that tends to be congruent with the views of those to whom one is accountable. The standard level of accountability in an assessment center may result in "acceptable" decisions if judgments reflect organizational notions of person/position fit or survival (Klimoski & Brickner, 1987). Also, other demands (such as acquaintance with the candidate or pre-assessment center information) may result in attempts to justify one's decision ("defensive bolstering"; Tetlock, Skitka, & Boettger, 1989; see also Dipboye, 1989).

Tetlock's model and empirical evidence suggest that in order to increase the appropriateness and accuracy of assessor judgments, the level of accountability that assessors feel to justify their ratings and observations to a reviewer whose opinions are unknown should be enhanced.
According to Tetlock's accountability model, enhanced accountability should result in assessors using more appropriate judgment strategies, having greater awareness of their cognitive processes, and in their processing information in a more data-driven, less theory-driven manner than they would have if their accountability had not been enhanced. Research has shown (e.g., Chaiken, 1980; Hagafors & Brehmer, 1983; McAllister, Mitchell, & Beach, 1979; Tetlock & Kim, 1987) that enhanced accountability to justify decisions results in greater thought complexity, better calibrated confidence estimates, and greater prediction accuracy.

In summary, although assessor training and role demands instill some accountability in assessors, assessors still use performance information in a nonnormative manner. Specifically, assessors may use exercise performance information or other performance information (e.g., personalogical) when making ratings. Enhancing the accountability demands on assessors to use only normative, dimension-based information may lead them to make more accurate decisions, as has been found in other investigations of accountability.

Social Judgment Issues

Research in behavioral decision-making and social judgment has widely documented the shortcomings, or heuristics, of human decision-making. Heuristics can impede
correct information processing and result in nonnormative
decision-making (Kahneman, Slovic, & Tversky, 1982; Nisbett &
Ross, 1980; Fiske & Taylor, 1984). Heuristics and biases
have traditionally been studied within an intrapersonal
paradigm. The typical viewpoint in social cognition and
social judgment research is that the interpersonal feedback
process dilutes these biases and that heuristics tend to
cancel each other out under normal circumstances (Nisbett &
Ross, 1980; though see Janis & Mann, 1977). As Fiske &
Taylor (1984) phrased it,

In short, when we succeed strikingly, we do so
because we use normative models and built-in
devices and safeguards to ensure that we
adhere to those models and because we make
judgments collectively by pooling expertise.
Our successes, then, may be successful
precisely because we do not use our intuitive
inferential strategies (p. 282).

Thornton & Byham (1982) also stressed the benefits of
interpersonal feedback in assessment center decision-making.
Even though assessors may carry preconceived notions of
performance and biased processing strategies into the
integration session, idiosyncrasies are canceled out of the
assessment process once dimensional information about
candidates is reviewed and the group provides feedback.
Implicit accountability in the assessment center, then,
results from the reliance on interpersonal feedback to
correct the judgment biases of individual assessors.

However, the factor analytic evidence cited earlier
shows that group process does not lead assessors to make
normative dimension judgments, that is, not in terms of the way assessors have been trained to observe and combine behavioral information. If group interaction and feedback indeed lessen biased intrapersonal "intuitive inferential strategies" (Fiske & Taylor, 1984), then obtained ratings should fit a normative dimension-based model. Since this is not the obtained result, either idiosyncratic processes remain impervious to group process, or the group members themselves may share and reinforce performance stereotypes and nonnormative processing strategies. That is, assessors may share performance stereotypes and categories that are not dimension-based.

**Categories And Schemas In Assessment Centers**

Since assessors seem not to rely on dimension-based judgment strategies, one may wonder what types of performance strategies assessors use, and how these strategies are developed. Klimoski and his colleagues (Klimoski & Strickland, 1977; Klimoski & Brickner, 1987) and Zedeck (1986) have raised these issues in conceptual papers and proposed that assessors rely on prototypes, categories, and schemas to make assessment center judgments.

One may use a category, defined as a class of objects considered equivalent (Fiske & Taylor, 1984), to classify stimuli and make predictions about members of the category. Judges view a stimulus as belonging (or not belonging) to a category by evaluating the stimulus in terms of prototypes,
which are themselves distilled from the instances of the
most typical features of the members of the category (Cantor
In particular, a judge may classify a stimulus into a
category based on family resemblance, or the degree of
overlap between the structural features of the stimulus and
the prototype. A judge may also classify a stimulus into
one or more categories based on cue validity, the
conditional probability that a stimulus attribute describes
individuals within a category given that it also describes
members of other categories. For example, Rosch & Mervis
(1975) showed that prototypes are likely formed by the
family resemblance of constituent individuals, and after the
prototype is formed the stimulus attribute cue validity for
discrimination among categories is maximized.

Assessor judgments within an assessment center may
reflect their use of categories and schemas. A dimension is
developed in an assessment center to encompass a set of
behaviors that either are similar in nature or that require
the same underlying ability for performance. Zedeck (1986)
argued that the taxonomic function of dimensions is
analogous to the function of categories. Assessors will
most likely include behaviors into a dimension that have the
closest match to the dimension's exemplar prototype
behaviors. He noted that the classification process is
probabilistic, as no one particular feature or set of
features is sufficient for inclusion of a behavior into a dimension.

An assessor's probabilistic matching is made more tenuous when behaviors representative of the same dimension vary across exercises. Specifically, behaviors that occur within a particular exercise tend to share same-exercise surface features (e.g., 'behaviors made during an interview'). This similarity could obscure distinctions among dimensions within exercises, thereby reducing dimension classification accuracy. Additionally, assessors may not have the training, ability, or motivation to develop appropriate prototypes for cross-situational dimensions or to make normative inclusion probability estimates. They may instead rely on shortcut strategies or heuristics to classify stimuli. The high within-exercise dimension intercorrelations reported in the literature may reflect the operation of these heuristics (e.g., Silverman et al., 1986).

An assessor's use of classification heuristics may further impede a normative judgment strategy because assessors may have inappropriate confidence that the encoded observations, now included in schemas, are accurate. A schema is "a cognitive structure that represents organized knowledge about a given concept or type of domain" (Fiske & Taylor, 1984, p. 140). Zedeck (1986) claimed that assessors use schemas to make sense out of the candidate's performance
given the goals of the exercise, and that a schema enables the assessor to anticipate related behavior in the future. The relation between categories and schemas is one of level of process: whereas prototypes and categories delineate the encoding of behavior, schemas guide the storage and recall of the behavior.

An assessor may integrate candidate performance across exercises through job behavior schemas (e.g., a management behavior schema). Zedeck (1986) emphasized that assessors with prior knowledge of an exercise's purpose will impose their expectations of performance onto observation and processing of a candidate's actual exercise performance. Assessors use the job behavior schema to help fill in performance information that would be likely to have occurred, though may in fact not have been observed in the exercises. What would be remembered is the assimilated information, not actual performance data.

In summary, Zedeck (1986) concluded that assessors process information about candidates in a hierarchical fashion (Figure 1, after Zedeck, 1986, p. 286). Assessors encode candidate performance into categories or dimensions through prototype matching. Exercise and person schemas subsume and relate relevant dimensions. Assessors rely on job behavior schemas to integrate and weight information encoded in person and exercise schemas, and make dimension ratings.
Figure 1

An example of the information processing of assessors who observe and evaluate one candidate across two exercises and four dimensions: Zedeck's (1986) hierarchical assessment center model

```
    OVERALL MANAGEMENT BEHAVIOR SCHEMA
      
Management behavior schema 1  Management behavior schema 2
                      
SCHEMA:  
Exercise 1  Exercise 2
          
DIMENSION:  
D_1 \ldots D_4  D_1 \ldots D_4
          
BEHAVIORAL INSTANCE:  
b_1, b_2, \ldots b_x  b_1, b_2, \ldots b_y
```

Note. b_x and b_y refer to the xth and yth behaviors exhibited by the candidate within each exercise, respectively. Assessors evaluate behaviors in terms of exercise schemas, scripts, and other stereotypes that have been formed as a result of personal experience and/or belief. Expectations for dimension performance are guided by exercise purpose and goals, ensuring high within-exercise intercorrelation.
Zedeck's account of assessment center processes is limited in that it summarizes empirical results, but doesn't offer boundary conditions or hypotheses for when schema-based judgment would -- or should -- be circumvented. Despite its shortcomings, Zedeck's hierarchical judgment model offers a link between empirical assessment center findings and judgment models in other applied disciplines.

Implicit Performance Constructs

Although the various categories, prototypes, and schemas assessors use have not yet been evaluated, the implicit performance constructs of raters in other areas of applied psychology have been investigated. Researchers have examined the extent and influence of implicit categories in areas such as leadership (Larson, 1982; Lord, Foti, & DeVader, 1984; Phillips, 1984; Rush, Thomas, & Lord, 1977), employment interviews (Hakel, Hollmann, & Dunnette, 1970; Jackson et al., 1980; Jackson, Peacock, & Holden, 1982), performance appraisal (Borman, 1983, 1987), and test validation design (Pulakos, Borman, & Hough, 1988).

These diverse lines of research have many nuances, but they are all investigations of the implicit performance constructs (IPC's) people may use to evaluate one another. IPC's are functional knowledge content structures that decision-makers hold for a given performance domain. IPC's are presumed to vary in content across individuals. In this sense, IPC's are similar to schemas and categories (Borman,
1987). IPC's tend also to be related functionally to other IPC's, and in this sense IPC's have much in common with implicit personality theory (Schneider, 1973; Lay & Jackson, 1969). Thus, analyses of IPC's differ from analyses of more generalized schema phenomena such as selective encoding, memory decay, or biased recall since both the functional nature of the content and the interrelations of IPC's are critical aspects of study.

The explanation of assessor judgment strategies based on IPC's may have several advantages over schema-based interpretations. For instance, Zedeck's schema model is based on probabilistic matching, and doesn't allow for individual variation in schema development. Medin (1989) has criticized similarity-based categorization (underlying Zedeck's model) for its inability to explain category structure and interrelations. Medin suggested that a theory-based explanation of categorization would be more tenable. Medin's model views categorization as based on a theory of functional relatedness among members. It may be that an assessor's categorization is not based on behavioral surface similarity, but on the functional relations among behaviors. The notion of a functional category is equivalent to an IPC.

Zedeck's model also makes several assumptions about schemas which have been challenged by recent categorization research. Briefly, Alba & Hasher (1983) noted that
processes related to retrieval, rather than encoding, may have a more potent influence on memory performance. Their review suggested that schemas may be less important than previously supposed in cognitive functioning.

The defining features of IPC's -- the functional nature of the construct and the relation of constructs to each other -- are more consistent with emerging trends in categorization research than a schema-based approach. This is not to imply that Zedeck's model is "wrong". Instead, more progress may be made in understanding assessors' decision-making strategies by investigating and integrating IPC's that are common to other applied areas of research rather than hypothesizing that assessors use schemas that are unique to the assessment center method.

Leadership IPC's

Lord and his colleagues (Rush et al., 1977; Lord et al., 1984; Foti & Lord, 1987) have theorized that implicit theories of leadership will bias encoding and recall of target stimulus behaviors. They reported that family resemblance of leader descriptors was correlated with both cue validity and prototypicality ratings, indicating a nomographic structure of the category "leader" (Lord et al., 1984). Further, they found that prototypical leader descriptions were more quickly accessed and recognized as leader-related than nonprototypical descriptors.

Employment Interview IPC's
Jackson and his colleagues have demonstrated the congruence of personality profiles and occupational fit for numerous occupations based on empirical self-reports (Seiss & Jackson, 1970). Specifically, raters can understand highly differentiated personality/position fit structures for different occupations and make inferences from these structures. For example, candidates making self-referent personality statements that were more congruent with the empirical personality/position fit described by Seiss & Jackson were rated by both professional interviewers and student subjects as more suitable and as having a greater probability of success on the job than candidates who made less congruent statements.

Jackson and his colleagues have explained these findings by hypothesizing that raters first infer linkages between behavioral statements and appropriate personality traits, and then construct the trait/occupational linkages leading to assessments of suitability, success expectations, etc. Other researchers have reported similar findings for personality/occupational interest congruence and congruence-success linkages (Holland, 1985; Costa, McCrae, & Holland, 1984; Johnson & Hogan, 1981; Hogan & Hogan, 1985).

Even though Jackson's studies were based on laboratory work, field studies of IPC's have tended to support a similar conclusion. Hakel et al. (1970) found differences among groups of interviewers in their descriptions of
accountants when rating accountant interests. Hakel et al. interpreted their results as indicating that differences between groups of evaluators would lead to differential weighting of candidate information within the interview. In summary, interviewer IPC's could be identified, the IPC content varied between groups, and these IPC's were related to expressed interests.

Finally, Dipboye & Macan (1988) proposed a social information processing model of the interview that stresses the role of interviewer IPC's. They hypothesized that interviewers hold many implicit performance theories for various jobs, and that the differential activation of one IPC over another would depend on an interaction of interviewer, interviewee, and role considerations. Activation of an IPC in the Dipboye and Macan model would then direct the conduct of the interview and the types of evidence sought or deemed admissible during the interview.

**Performance Appraisal IPC's**

Feldman (1981; Ilgen & Feldman, 1983) and Cooper (1981) have argued that performance appraisal processes follow IPC's. Performance categories develop "along the lines of the correlations found in nature" (Feldman, 1981, p. 130). Feldman hypothesized that when categorization cannot proceed according to the IPC a judge typically uses, the judge will tend to use attribution principles to explain performance and then make controlled categorization judgments based on
either personal constructs of performance or trait-based categories.

A more direct discussion of the role of IPC's in performance appraisal has been provided by Borman (1983, 1987; Pulakos et al. 1988) in terms of personal constructs. Personal constructs are "category systems for individuals within a 'focus of convenience' ... (which) aid in organizing and simplifying (behavioral) information" (Borman, 1987, p. 309). These constructs reflect ideographic performance qualities that raters find salient in their assessment of others. Such constructs may also indicate the most readily accessible trait categories raters are prone to use when describing others (Bargh, Lombardi, & Higgins, 1988). Interrater agreement among evaluators may result from similarities in these personal work constructs acquired through common norms, roles, etc. within the work place. Constructs generalize acceptably well across observers (Borman, 1987), and Borman and his colleagues have found that performance appraisal categories based on IPC analysis provide adequate and stable criteria for selection test batteries (Pulakos et al., 1988).

Personal logical IPC's

Much personality research has centered on the implicit personality theory debate and the "big five" trait dimensions of personality (e.g., Schneider, Hastorf, & Ellsworth, 1979; McCrae & Costa, 1987). Though earlier
reviewers argued for an illusory correlation interpretation of personality constructs and their interrelations (Schneider, 1973), more recent studies (Jackson, Chan, & Stricker, 1979; White, 1980; Borkenau, 1988; McCrae & Costa, 1987; Mershon & Gorsuch, 1988; Digman, 1990) have provided evidence for personality structure that is cross-cultural, trait-based, and that has clear implications for real-world outcomes (job performance, health status, etc.).

These personalological dimensions are categories in that they distinguish among actions and they tend to reflect the perceived covariance of social performance (Feldman, 1981). There may be a natural tendency for perceivers to evaluate social performance in trait category terms. This could result through automatic personalological category activation during encoding or unintended access of trait-related information during schema formation, integration, and decision-making (Feldman, 1981; Wyer & Srull, 1981; Cooper, 1981).

Krzystofik, Cardy, & Newman (1988) provided an example of typical reliance on traits in a performance appraisal situation. They demonstrated that raters inferred job-related personality characteristics from discrete behavioral events, and that evaluation of inferred personality characteristics had a significant incremental contribution to performance ratings over behavioral information. They concluded that trait categorization is a natural and
automatic process which may be inherent in social information processing, and that trait evaluation seemed to influence evaluations of behavioral categories.

**IPC Implications For Assessment Centers**

The IPC studies cited above have indicated that perceivers tend to process social information and make judgments according to readily identifiable, nomographic patterns (e.g., in terms of leadership, personality, exercise performance, or person/position fit). While IPC's have been investigated in many applied settings, the extent to which assessors may use IPC's in their ratings, and the actual content of these IPC's, is unclear.

One type of IPC may be based on exercise performance evaluation in making OAR's. While this phenomenon is neither "good" nor "bad", use of nonnormative IPC's may indicate the need for more effective training, a refocusing on exercise performance as well as dimension performance in assessment center design, or some other intervention.

Also, it is not clear that assessors use only exercise IPC's; as yet, no research has addressed personality-evaluative IPC's in assessment center evaluations in the manner that Kryzstofiak et al. (1988) did in a performance appraisal study. However, Bray & Howard (1985) provided evidence that self-report personality dispositions may serve as valuable indicators of future performance, thus indicating that personality-based IPC's may be utile
predictors.

The Present Investigation

This study examined the extent of the effect of enhanced accountability on assessor judgments. The IPC's, observations, and classifications of assessors whose accountability was enhanced were compared to the IPC's and behavioral recordings of assessors whose level of accountability was not enhanced. The effect of enhanced accountability was assessed through assessors' use of performance information, rating accuracy, accuracy in behavioral observations, and accuracy in behavioral classifications.

Hypotheses

The following four hypotheses were proposed for investigation of judgment bias factors:

1) Assessor whose accountability for justifying their ratings is not enhanced will make nonnormative ratings that are based on the candidates' personalities, exercise performances, and performance on dimensions. Assessors whose accountability for justifying their ratings is enhanced will make normative ratings that are based on the candidates' performances on dimensions alone;

2) Assessor whose accountability for justifying their ratings is not enhanced will be more confident in their ratings than will assessors whose accountability for justifying their ratings is enhanced;
3) Assessors whose accountability for justifying their ratings is not enhanced will make less accurate dimension ratings than will assessors whose accountability for justifying their ratings is enhanced.

4) Assessors whose accountability for justifying their ratings is not enhanced will make less accurate observations, and will also classify behaviors less accurately, than will assessors whose accountability for justifying their ratings is enhanced.

METHOD

Overview

Students from several undergraduate psychology courses served as assessors in an assessment center simulation to evaluate candidates for a resident hall student assistant position. In an initial session, all assessors were told the purpose of the assessment center and trained how to make appropriate observations and behavioral classifications. In a subsequent session, subjects made written observations of videotaped performances of three student assistant candidates and classified their behavioral observations into dimensions of performance. Assessors then met in groups and shared their observations for each candidate in turn. Subsequently, each assessor made an individual rating of that candidate's across-exercise dimension performance. At the close of the experiment, assessors made exercise and personalogical ratings. Assessors in the enhanced
accountability condition were reminded regularly throughout the experiment that their ratings would be evaluated carefully by the experimenters.

Subjects

Ninety-four subjects from Rice University and the University of Houston volunteered to participate as assessors for course credit. In addition, three students from the University of Houston received cash remuneration instead of credit for their participation. In all, ninety-seven students participated in the study, with 71% from the University of Houston and 29% from Rice University. The average age of the participants was 21.6 yrs.; 50.5% of the participants were women. The average residency in a dormitory or residence hall was 7.8 months (almost one full school year). Forty-nine students participated in the standard accountability condition, while forty-eight participated in the enhanced accountability condition. Table 1 presents the demographic characteristics of the Rice and University of Houston participants.

Results presented in Table 1 indicate that the University of Houston participants were older than the Rice students, but a higher percentage of the Rice students reported having lived in a dormitory or residence hall. Also, the average Rice student had lived for a longer time in a dormitory or residence hall. The possible effects of these differences are reviewed later in the Results section.
Table 1

Differences in demographic characteristics between Rice and University of Houston study participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>School</th>
<th>Mean (SD)</th>
<th>N</th>
<th>t</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years</td>
<td>Rice</td>
<td>19.9 (1.7)</td>
<td>27</td>
<td>-3.6**</td>
<td>90.4^e</td>
</tr>
<tr>
<td></td>
<td>Houston</td>
<td>22.3 (4.7)</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex^b</td>
<td>Rice</td>
<td>1.64 (.49)</td>
<td>28</td>
<td>1.7</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Houston</td>
<td>1.45 (.50)</td>
<td>69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dormitory^c, 86.6^e</td>
<td>Rice</td>
<td>.93 (.26)</td>
<td>28</td>
<td>7.4**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Houston</td>
<td>.36 (.48)</td>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residency^d</td>
<td>Rice</td>
<td>15.7 (9.6)</td>
<td>28</td>
<td>5.9**</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Houston</td>
<td>4.5 (8.0)</td>
<td>68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.  **p<.01.  ^a Number of participants with nonmissing data.  ^b A value of 1 was assigned for men and a value of 2 was assigned for women.  ^c A value of 1 was assigned for having resided in a dormitory or residence hall and a value of 0 was assigned for not having lived under those conditions.  ^d The number of months reported living in a dormitory or residence hall (1 semester was estimated as 4 months).  ^e Estimates for the degrees of freedom were used when the variances were not homogeneous.
Materials

**Video tapes.** Assessors viewed three 10-minute videotaped segments of an assessment center for selecting college residence hall student assistants (SA's). The exercises were based on a study by Gaugler (1987) and included the following: a leaderless group discussion (LGD) among the SA's about limiting student alcohol consumption in the dorms; a one-on-one counseling session (CS) between an SA candidate and student (role-player) with "personal problems"; and an intervention between two quarreling roommate role-players (RC). Tapes were professionally developed and produced, and all performers were compensated actors.

There is some controversy over the use of college students in applied research (e.g., Gordon, Slade, & Schmitt, 1986). There are two mitigating factors of this concern in the present study, though. First, Campbell's (1986) summary of the evidence from field and laboratory studies demonstrated that there is a considerable level of correspondence between the results obtained from these two experimental settings. Second, the SA job was selected because students would be familiar with the SA's duties. This complies with Thornton & Byham's (1982) stipulation that assessors should be very familiar with the job for which the candidate applies. Third, this study investigates a basic psychological process of social interaction, the
evaluation of others in a structured situation.

**Assessment center dimensions.** Four dimensions were chosen for examination within this study by a job analysis and the consensus of a panel of assessment center experts. The panel consisted of three Doctoral students and two faculty members, each of whom had participated in the development and/or evaluation of at least one assessment center.

The four dimensions chosen for evaluation included: Planning and Organizing; Oral Communication; Analysis and Judgment; and Sensitivity. These four dimensions were chosen and worded to be as distinguishable as possible while still reflecting accurately the content of the SA job. Behavioral examples were chosen for each dimension and incorporated as minimal standards of performance on each dimension within each exercise. These minimal standards were chosen to reflect behaviors the candidates should exhibit in order to be considered minimally qualified on the dimension by assessors. Table 2 lists all variables (dimension, personalogical, and exercise) assessed in this study and their definitions.

**Target ratings.** A new panel consisting of the author and two volunteer psychology graduate students served as the experts who provided the target ratings to which the participants' ratings were compared. Each panel member had successfully completed a course on the assessment center
Table 2

Definitions of performance variables assessed in the current study

ASSESSMENT CENTER DIMENSION VARIABLES:

ANALYSIS and JUDGMENT (A/J): The ability to identify problems and possible causes; gathers relevant information and relates data from different sources to solve problems.

ORAL COMMUNICATION (OC): The ability to express oneself effectively in individual or group situations, including gestures and other nonverbal behaviors; conveys thoughts clearly and concisely; does not go off on tangents.

PLANNING and ORGANIZING (P/O): The ability to establish a course of action for self and/or others to accomplish a specific goal; makes proper assignments of people and allocates resources appropriately.

SENSITIVITY (SN): The ability to appraise accurately the needs, feelings, skills, and competencies of others in interpersonal situations and act accordingly.

EXERCISE VARIABLES:

LEADERLESS GROUP DISCUSSION (GD): The candidate's perceived performance in the LGD exercise.

COUNSELING SITUATION (CS): The candidate's perceived performance in the CS exercise.

ROOMMATE CONFLICT (RC): The candidate's perceived performance in the RC exercise.

(Table 2 Continues)
(Table 2 Continued)

PERSONALOGICAL VARIABLES:

NEUROTICISM (NE): The propensity to experience a variety of negative affect, including disturbed (nonpsychotic) thought and behavior that accompanies psychological distress.

EXTRAVERSION (EX): The propensity to enjoy and participate in social gatherings; includes sociability and energy level.

OPENNESS TO EXPERIENCE (OP): The tendency to be curious about, and not prejudiced against, new experiences, ideas, and feelings. Includes intellectual breadth, though only modestly related to intelligence.

AGREEABLENESS (AG): The tendency to have a positive orientation toward other individuals, including trust, compassion, and altruism.

CONSCIENTIOUSNESS (CO): The tendency to apply organization, persistence, and motivation in goal-directed behavior. Includes dependability and fastidiousness.
method, and served as assessors and assessment center
designers. These experts viewed each tape and made detailed
observations and classifications of the performance of each
of the three candidates. For each exercise, the panel
members shared their classified observations and made
independent dimension ratings. Next, the experts shared
their initial dimension ratings and arrived at consensus
ratings for each candidate's dimension and OAR ratings.
These judgments, based on an enhanced viewing opportunity,
served as the "target scores" for derivation of the accuracy
components described below.

There is some debate as to whether the experts provided
"true" scores in the classical test theory sense (Stanley,
Since undergraduate subject participants and not psychology
graduate students provided study data, the target scores in
this study may not be those that preselected and motivated
undergraduates, given an enhanced rating opportunity, would
have provided. In using graduate students as experts, this
study follows established precedent in studies of the
assessment center method, performance appraisal, and
interviews (e.g., Gaugler & Thornton, 1989; Murphy & Balzer,
1986). Neither this study nor the others listed address the
construct validity debate about the "true"- vs. "target"-
score distinction that can only ultimately be answered
empirically. Target scores are included in Appendix A.
Procedure

Overview. All assessors received the same initial training which focused on identifying and making valid behavioral observations. Assessors were then randomly assigned to either the enhanced accountability (manipulated) or standard accountability (control) condition. In the enhanced condition, assessors were informed repeatedly that the experimenters would evaluate the justifiability of each assessor's observations and ratings. In the standard condition, no mention was made of whether or not experimenters would review observations and ratings. Assessors were then reassigned to integration groups where they shared observations and made independent performance judgments.

Training session. In the initial training session, participants were given a description of the SA's job and then were trained in making accurate behavioral observations, in the use and understanding assessment center dimensions, and in evaluation of behavior in terms of minimally acceptable levels of performance. As part of their training in making behavioral observations, assessors were taught how to identify characteristics of good and poor observations and practiced making observations of performance from a five-minute video tape. Individualized feedback on the adequacy of behavioral observations was provided during the subsequent session. Assessors were
encouraged to ask questions to ensure that they had an adequate understanding of the nature and mechanics of the assessment center.

The training session lasted one hour. Though many assessment centers in industry conduct more rigorous and time-consuming training programs, the duration of the training program does not seem to moderate assessment center validity (Gaugler et al., 1987) or assessors' ability to discriminate among dimensions (Dugan, 1988). At the close of training, assessors were assigned randomly to one of several experimental sessions. Appendix B contains all forms and the training protocol used in the training session.

Experimental Session. The experimental session was designated at random in advance as either an enhanced accountability condition or a standard accountability condition. The following outline describes the general sequence of the experimental session.

Upon arrival for the experimental session, assessors were assigned randomly into one of three observation groups to view the video tapes. The presentation order of candidates for each observation group's tapes was unique, though all observation groups saw the same order of exercises (leaderless group, counseling, and roommate conflict).

Trained research assistants, blind to the hypotheses of
the study, served as assessment center administrators. Each administrator was assigned randomly to lead one of the three observation groups. Within each group, the administrator distributed a packet of materials, reviewed the purpose of the study, explained feedback from the training session, and reiterated the principles and importance of making valid observations and classifications.

Assessors then were told to make observations on a particular candidate in the first vignette, the LGD. After the assessors completed viewing the LGD, assessors had 15 minutes in which to classify their observations into the appropriate dimensions of performance. This sequence of observation and classification was followed for the two remaining vignettes, CS and RC.

Upon completion of the observation and classification sequence, assessors were assigned through stratified random sampling to one of several integration groups. Within each group, at least one assessor had made observations on each candidate in each exercise. For example, in each integration group one assessor had observed Candidate A in the LGD, another had observed Candidate A in the CS, and a third had observed Candidate A in the RC.

Since there were three candidates and three vignettes, each observation group was balanced if it was comprised of three assessors. However, it would frequently happen (due to fewer or more participants arriving for the session than
planned) that there would be more than three assessors in an observation group. In these unbalanced cases, there was one observation group that was larger than the others. Two assessors from the larger observation group were then included in one of the integration groups, and the two assessors took turns reporting their classifications for each dimension in the integration session.

**Accountability Manipulation.** For the sessions designated as *enhanced accountability* conditions, assessors were told that the purpose of the experiment was to evaluate the information processing and social judgment factors of assessors in an assessment center. They were told before the observation session that the experimenters would review each assessor's ratings and behavioral classifications to determine the justifiability of these results, with justifiability defined in terms of "match to the observations and ratings made by experts". Assessors then signed an additional consent form for the release of this information "for further study" of assessment center processes. These assessors were also reminded periodically that they should be as circumspect in their observations, classifications, and ratings as possible, because the experimenters would evaluate their data carefully. Finally, before the beginning of the integration session, these assessors were asked to record their phone numbers and a convenient time for the experimenters to contact the
participants if their observations and classifications were deemed "unjustifiable". Assessors were told that they would be contacted within the next two days.

In the standard accountability condition, there was no mention of what would become of the data, nor that anyone would review the data. They were told that the purpose of the study was to evaluate a new system for hiring resident hall student assistants. Assessors in this condition were not asked to complete a data release form. Appendix C contains all forms used in the experimental conditions. Appendix D contains the protocol for the experimental sessions.

One may question the generalizability of these two levels of accountability to functional assessment centers. The standard level of accountability in this experiment may be less than that in some functional assessment centers; however, the levels of accountability in this study are meant to be relative to each other. Even in functional assessment centers, the amount and type of accountability vary. For example, some assessors may need to make reports to the candidate, though others may not; some assessment centers use group consensus for ratings, while others (such as this one) may not.

**Ratings.** Candidates were evaluated sequentially, dimension by dimension. Once observations for each dimension had been shared, each assessor rated independently
the candidate's effectiveness on the dimension based on the minimal standards of performance, and rated their confidence that the dimension rating was correct. Dimension ratings were based on a 1 ("Low performance on the dimension") to 5 ("High performance on the dimension") scale, with 3 anchored as "Acceptable performance on the dimension, not low, not high". Confidence ratings were based on a 0 (0 percent confident) to 100 (100 percent confident) scale. This rating procedure was followed for each dimension (within candidate) in turn. After all four dimensions had been evaluated, each assessor made an independent judgment of the candidate's overall performance in the assessment center, and rated their confidence in this rating.

The integration session was designed so that there was no disclosure of dimension or confidence ratings and no consensus discussion among the assessors. The data of interest were at the level of the individual ratings. Sackett & Wilson (1982) found that, in many cases, individual and group judgments were highly similar (see also Thornton & Byham, 1982), indicating that the actual ratings may have been the same in the two formats. However, the use of group decision data could have resulted in undesirable group rating effects (e.g., group polarization; Sackett & Wilson, 1982). This "inappropriate" nonnormative rating variance was controlled by the examination of only individual-level data.
After evaluating all three candidates, assessors evaluated the personalological characteristics and exercise performances of each candidate. Candidates' personalological characteristics were evaluated through the use of a checklist of forty bipolar adjectives. The checklist was adapted from McCrae & Costa (1985) and described the five prominent personalological trait categories that have been derived through multiple studies (Digman, 1990): Neuroticism; Extraversion; Openness to Experience; Agreeableness; and Conscientiousness. Eight bipolar adjectives were presented for each dimension.

Assessors made evaluations of each candidate's performance in each exercise using a 1 to 5 scale. The scale points corresponded to those for the dimension scales, but no minimal standards for exercise performance were presented. The order in which assessors made personality and exercise evaluations was counterbalanced. Assessors had not been told in advance that they would have to make exercise or personalological evaluations. Finally, assessors completed the manipulation check and demographic characteristics form. All assessors were then thanked and later debriefed.

RESULTS

Aggregation Of Data

For most analyses reported in this study (except Hypothesis 3), each assessor's ratings were averaged across
role player stimuli to arrive at a construct composite rating. For instance, each assessor's rating of "OAR" is the average of the assessor's OAR for Candidate A, Candidate B, and Candidate C.

This aggregation procedure was adopted because analyzing the candidates separately would treble the "number of observations" in this study: instead of ninety-seven "averaged" sets of ratings in the database, there would be three sets of ratings for each of the ninety-seven assessors, for a total of 291 "individual" ratings, thus lowering power (by increasing "family-wise" Type I error) and violating assumptions about the independence of ratings.

Table 3 presents the bivariate correlations among the dimension, exercise, and personalological characteristics variables. The reliabilities of the personalological variables were obtained by determining each of the five coefficient alphas for each personalological characteristic separately, and then averaging the coefficient alphas across candidates for each personalological characteristic.

The average of the bivariate correlations between the exercise ratings and the dimension ratings was estimated through the Fisher r-to-z transformation as .438. Analyses of 95% confidence intervals (Hays, 1981) indicated that the average exercise-dimension correlation was significantly higher than both the average correlation of personalological characteristics and dimensions (.187) and the average
Table 3

Bivariate intercorrelations among rating variables

1: Correlations of all rating variables with dimension variables

<table>
<thead>
<tr>
<th></th>
<th>A/J</th>
<th>OC</th>
<th>P/O</th>
<th>SN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A/J</td>
<td>(.97)(^b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OC</td>
<td>.36</td>
<td>(.98)(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P/O</td>
<td>.51</td>
<td>.42</td>
<td>(.99)(^b)</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>.53</td>
<td>.45</td>
<td>.32</td>
<td>(.97)(^b)</td>
</tr>
</tbody>
</table>

**Exercise variables**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GD</td>
<td>.40</td>
<td>.44</td>
<td>.59</td>
<td>.46</td>
</tr>
<tr>
<td>CS</td>
<td>.52</td>
<td>.39</td>
<td>.44</td>
<td>.38</td>
</tr>
<tr>
<td>RC</td>
<td>.44</td>
<td>.39</td>
<td>.32</td>
<td>.41</td>
</tr>
</tbody>
</table>

**Personalological variables**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NE(^d)</td>
<td>.21</td>
<td>.17</td>
<td>.02</td>
<td>.06</td>
</tr>
<tr>
<td>EX</td>
<td>.17</td>
<td>.28</td>
<td>.12</td>
<td>.29</td>
</tr>
<tr>
<td>OP</td>
<td>.03</td>
<td>.21</td>
<td>.10</td>
<td>.22</td>
</tr>
<tr>
<td>AG</td>
<td>.27</td>
<td>.38</td>
<td>.03</td>
<td>.23</td>
</tr>
<tr>
<td>CO</td>
<td>.20</td>
<td>.36</td>
<td>.17</td>
<td>.22</td>
</tr>
</tbody>
</table>

(Table 3 continues)
(Table 3 continued)

2: Correlations of exercise and personalogical characteristics variables with exercise variables

<table>
<thead>
<tr>
<th></th>
<th>GD</th>
<th>CS</th>
<th>RC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exercise variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GD</td>
<td>.99(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>.43</td>
<td>.94(^b)</td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>.32</td>
<td>.31</td>
<td>.99(^b)</td>
</tr>
</tbody>
</table>

**Personalogical characteristics**

<table>
<thead>
<tr>
<th></th>
<th>NE(^d)</th>
<th>EX</th>
<th>OP</th>
<th>AG</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE(^d)</td>
<td>-.10</td>
<td>.09</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EX</td>
<td>.11</td>
<td>.10</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>.14</td>
<td>.07</td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>.01</td>
<td>.21</td>
<td>.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>.11</td>
<td>.17</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3: Intercorrelations among personalogical characteristics

<table>
<thead>
<tr>
<th></th>
<th>NE</th>
<th>EX</th>
<th>OP</th>
<th>AG</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE(^d)</td>
<td>.79(^c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EX</td>
<td>.25</td>
<td>.85(^c)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>.10</td>
<td>.76</td>
<td>.81(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>.66</td>
<td>.47</td>
<td>.34</td>
<td>.87(^c)</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>.47</td>
<td>.52</td>
<td>.50</td>
<td>.70</td>
<td>.83(^c)</td>
</tr>
</tbody>
</table>

**Note.** For all analyses, N = 97. Correlations exceeding .19 are significant, two-tailed p<.05. A/J: Analysis and judgment; OC: Oral communication; P/O: Planning and organizing; SN: Sensitivity; GD: Leaderless group discussion; CS: Counseling session; RC: Roommate conflict; NE: Neuroticism; EX: Extraversion; OP: Openness to experience; AG: Agreeableness; CO: Conscientiousness. \(^a\) Rating variables are defined in Table 1. \(^b\) Intraclass correlation estimates of reliability. \(^c\) Coefficient alpha
estimates (SPSSx, 1988) for the variable obtained through averaging the estimates of personalogical rating alphas for each candidate. For all analyses, Neuroticism is reverse-scored (see definition).
correlation of the exercise and personalogical ratings (.110). Finally, a 95% confidence interval around the personalogical characteristics-exercise ratings correlation included 0. Thus, it seems likely that the exercise and personalogical characteristic ratings were not related.

Manipulation Check

At the close of the assessment center, all assessors completed a series of questions designed to ascertain whether they correctly perceived the accountability manipulation. Assessors were asked to report: how important they perceived the task to be; whether they thought their observations and classifications were justifiable; how well their ratings and observations matched those of assessors; and how closely they thought the experts would examine their ratings and classifications. These items were similar to items used in previous accountability research (e.g., Chaiken, 1980), and are referred to here as Accountability Manipulation Check 1 (AMC1).

During the course of the study, after conference with members of the author's committee, it was felt that extra manipulation check items were desirable. Therefore, the final sixty-six assessors responded to the AMC1 items plus the following: whether assessors were told that their data would later be scrutinized; how much pressure they felt to perform well; and how much thought they put into the task. The items relating to "thought" and "pressure" constituted a
second manipulation check, Accountability Manipulation Check 2 (AMC2). The item asking whether or not assessors were informed of the manipulation was kept separate of both AMC1 and AMC2 to reflect memory for the manipulation itself, as it is possible that assessors may have remembered being manipulated but not feel particularly accountable.

To summarize, all participants responded to the four items in AMC1; the final sixty-six participants also responded to the two items in AMC2 (about pressure and thought) and the dichotomous manipulation recall item. A final composite measure for only those sixty-six participants who completed all six manipulation check measures is included as AMC3. Mean values for each item in each accountability manipulation check are presented in responded to these items) indicated that the composite dependent variables differed significantly between groups, $F(4, 92) = 2.84, p < .05, \eta^2 = .110$. ($\eta^2$ is a measure of Table 4.

Differences between the two accountability conditions were tested for unit-weighted mean manipulation check measures. These group differences in ratings were tested through multivariate analysis of variance (MANOVA). Although the item and scale means for the two groups seem similar, a MANOVA of the AMC1 items (all participants the variance explained in the centroid scores by group membership, and is calculated as $1 - (\text{Wilks's Lambda})$.) For
Table 4

Means and standard deviations of manipulation check measures

<table>
<thead>
<tr>
<th>DEPENDENT MEASURE</th>
<th>WHOLE GROUP</th>
<th>STANDARD</th>
<th>ENHANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts¹:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>3.11 (.653)</td>
<td>3.13 (.665)</td>
<td>3.09 (.647)</td>
</tr>
<tr>
<td>Justifiable¹:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>3.60 (.702)</td>
<td>3.73 (.730)</td>
<td>3.46 (.651)</td>
</tr>
<tr>
<td>Closely¹:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>3.54 (.867)</td>
<td>3.41 (.788)</td>
<td>3.67 (.930)</td>
</tr>
<tr>
<td>Important¹:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>3.87 (.731)</td>
<td>3.90 (.743)</td>
<td>3.83 (.724)</td>
</tr>
<tr>
<td>AMC¹,³:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>0.00 (1.00)</td>
<td>.03 (1.014)</td>
<td>-.03 (.995)</td>
</tr>
<tr>
<td>Thought²:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>3.63 (.485)</td>
<td>3.58 (.515)</td>
<td>3.69 (.451)</td>
</tr>
<tr>
<td>Pressure²:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>3.00 (.579)</td>
<td>2.92 (.554)</td>
<td>3.09 (.597)</td>
</tr>
<tr>
<td>AMC²,³:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>0.00 (1.00)</td>
<td>-.11 (.994)</td>
<td>.13 (1.01)</td>
</tr>
<tr>
<td>AMC³²,³:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN (SD)</td>
<td>0.00 (1.00)</td>
<td>-.16 (.976)</td>
<td>.16 (1.01)</td>
</tr>
</tbody>
</table>

(Table 4 continues)
(Table 4 continued)

**Told about manipulation**:  

<table>
<thead>
<tr>
<th>MEAN</th>
<th>.606</th>
<th>.323</th>
<th>.857</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SD)</td>
<td>(.492)</td>
<td>(.475)</td>
<td>(.355)</td>
</tr>
</tbody>
</table>

**Note.** SD = Standard deviation.  

1. \( N = 97 \).  
2. \( N = 66 \).  

Results for AMC1, AMC2, and AMC3 reported in Z-score format. All accountability manipulation check items (Justifiable, Experts, Closely, Important, Thought, and Pressure) measured on a 1 (None) to 5 (Extreme) Likert-type scale. Enhanced accountability condition: \( N = 48 \); Standard accountability condition: \( N = 49 \).
the two dependent measures alone in AMC2, the MANOVA composite did not differ significantly between groups, $F(2, 63) = 0.46, \eta^2 = 0.014, p > 0.05$. A MANOVA of the six dependent measures completed by the final sixty-six assessors (AMC3) was statistically significant, $F(6, 59) = 2.57, p < 0.05, \eta^2 = 0.207$.

The statistically significant MANOVA results for AMC1 and AMC3 were further clarified by analyses of the canonical discriminant structure of the variables. Membership in either experimental condition serves as the criterion of interest in canonical discriminant analysis. The observed (manipulation check) variables are combined through linear weighting into a canonical variate. Mean profile values are developed by applying the linear weights to the obtained values of the variables; these mean profile values are commonly referred to as the group centroids. In canonical discriminant analysis, both the centroids and the statistical weights describing the variable-variate relations are examined to determine the structure and significance of the differences between classification groups. Results of the canonical discriminant analyses for manipulation check variables are presented in Table 5.

For AMC1, discriminant canonical analysis indicated that one variate discriminated significantly between groups, $F(4, 92) = 2.84, \eta^2 = 0.110, p < 0.05$. (This is the same test, thus the same value, as the MANOVA reported above.) The
variable-variate structure was then examined for between-
group differences. Structure was assessed in terms of
variable-variate correlations and in terms of the beta
weights for predicting the centroids from the observed
variables. Correlations were judged to be meaningful if
they exceeded .30, a standard interpretative criterion for
these types of analyses (Tabachnick & Fidell, 1989, p. 216).

Analysis of the variable-variate correlations indicated
that assessors in the enhanced accountability condition
reported a higher level of feeling that the experimenters
would examine their ratings closely than did those in the
standard accountability condition ($r = -.4312$ in AMC1 and
-.4604 in AMC3), as expected. However, assessors in the
standard accountability condition tended to feel that their
ratings and observations were more justifiable than did
assessors in the enhanced accountability condition ($r = .5741$
in AMC1 and .5247 in AMC3). The beta weights further
corroborated these findings.

This pattern of findings is appropriate given the
theoretical nature of accountability. Tetlock and his
associates have shown (e.g., Tetlock & Kim, 1987) that
enhanced accountability demands lead to cognitive effort and
pre-emptive self-questioning and self-criticism. The
enhanced accountability manipulation may have increased
perceived justifiability, or the mediating effect of the
pressure to justify one's ratings to another in the enhanced
Table 5

Canonical discriminant analyses: Predicting experimental group membership from the manipulation dependent measures

1): Canonical structure of the variate discriminating between experimental group membership for AMCla:

<table>
<thead>
<tr>
<th>RAW VARIABLES</th>
<th>CORRELATION WITH DISCRIMINANT CENTROID</th>
<th>BETA-WEIGHT LOADING ON THE CENTROID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts</td>
<td>.0892</td>
<td>.0333</td>
</tr>
<tr>
<td>Closely</td>
<td>-.4312</td>
<td>-.0547</td>
</tr>
<tr>
<td>Important</td>
<td>.1266</td>
<td>.5048</td>
</tr>
<tr>
<td>Justify</td>
<td>.5741</td>
<td>.8499</td>
</tr>
</tbody>
</table>

2): Canonical structure of the variate discriminating between experimental group membership for AMC3b:

<table>
<thead>
<tr>
<th>RAW VARIABLES</th>
<th>CORRELATION WITH DISCRIMINANT CENTROID</th>
<th>BETA-WEIGHT LOADING ON THE CENTROID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thought</td>
<td>-.1427</td>
<td>-.1941</td>
</tr>
<tr>
<td>Pressure</td>
<td>-.2222</td>
<td>-.2496</td>
</tr>
<tr>
<td>Experts</td>
<td>-.1887</td>
<td>-.3203</td>
</tr>
<tr>
<td>Closely</td>
<td>-.4604</td>
<td>-.8552</td>
</tr>
<tr>
<td>Important</td>
<td>-.1090</td>
<td>.3092</td>
</tr>
<tr>
<td>Justify</td>
<td>.5247</td>
<td>.9860</td>
</tr>
</tbody>
</table>

Note. All manipulation check variables defined in text. Negative values (-) indicate that the variable is more highly related to the enhanced accountability condition. a N=97. b N=66.
condition. That is, participants in the standard accountability condition may have expressed a higher level of belief that their ratings were justifiable because they had not been pressured to think about how justifiable their ratings really were. No extra-ordinary effortful cognitive work, which could have introduced reasonable doubt about ratings, was demanded of participants in the standard condition as it was of participants in the enhanced accountability condition. Thus, the manipulation seems to have created a sense of reasoned self-doubt for assessors in the enhanced accountability condition.

Finally, a z-test was performed to determine whether assessors in the enhanced accountability condition were more likely than assessors in the standard accountability condition to report that they had been told that their data would be scrutinized. This difference was also statistically significant, $z = 4.43, p < .01$, indicating that assessors in the enhanced accountability condition were more likely to report remembering the manipulation than were assessors in the standard accountability condition ($M's = .857$ and $.323$, respectively).

**Hypothesis 1: Accountability Effects On Rating Strategies**

Hypothesis 1 stated that the imposition of enhanced accountability would significantly affect the assessor's use of nonnormative performance information in making overall judgments about candidates. Specifically, it was predicted
that assessors in the enhanced accountability condition
would use only normative dimension variance in making OAR
judgments, whereas the OAR's of assessors in the standard
accountability condition would be influenced by exercise and
personalological factors. Table 6 contains results of
hierarchical regressions for testing this hypothesis. These
regressions show the incremental prediction of the OAR for
the normative model, and three nonnormative models, within
each group. The sample size within each condition is lower
than may be desirable, but is adequate at a 5:1 (respondent-
to-variable) ratio for regression (Tabachnick & Fidell, 1989).

The results in Table 6 indicate that the increases in
model R²'s are consistently significant due to the various
nonnormative factors within the standard accountability
condition. These increments of between 5% and 10% of the
total variance in the R²'s over the dimensions-only model
are consistent with the predictions of this hypothesis. In
the standard accountability condition, assessors used
nonnormative performance information that contributed
significantly to their OAR judgments.

Table 6 also shows that the increments in R² for both
the exercise variables and personalological characteristics
within the enhanced accountable condition are not
significant. This set of findings lends further support to
Hypothesis 1: assessors in the enhanced accountable
Table 6

Hierarchical regression analysis for each condition:
Prediction of the OAR

**STANDARD CONDITION**

<table>
<thead>
<tr>
<th>Predictor Set:</th>
<th>MS</th>
<th>F (df)</th>
<th>R²</th>
<th>R² change</th>
<th>F change (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions only</td>
<td>1.88</td>
<td>29.14</td>
<td>.7259**</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4,44)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions, Exercises</td>
<td>1.16</td>
<td>20.81</td>
<td>.7804**</td>
<td>.0545</td>
<td>3.39*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7,41)</td>
<td></td>
<td></td>
<td>(3,41)</td>
</tr>
<tr>
<td>Dimensions, Personological characteristics</td>
<td>0.91</td>
<td>16.50</td>
<td>.7920**</td>
<td>.0661</td>
<td>2.48*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9,39)</td>
<td></td>
<td></td>
<td>(5,39)</td>
</tr>
<tr>
<td>Dimensions, Exercises, Personological characteristics</td>
<td>0.72</td>
<td>14.71</td>
<td>.8307**</td>
<td>.1048</td>
<td>2.79*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12,36)</td>
<td></td>
<td></td>
<td>(8,36)</td>
</tr>
</tbody>
</table>

**ENHANCED CONDITION**

<table>
<thead>
<tr>
<th>Predictor Set:</th>
<th>MS</th>
<th>F (df)</th>
<th>R²</th>
<th>R² change</th>
<th>F change (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions only</td>
<td>1.75</td>
<td>34.11</td>
<td>.7604**</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4,43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions, Exercises</td>
<td>1.05</td>
<td>22.04</td>
<td>.7941**</td>
<td>.0337</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7,40)</td>
<td></td>
<td></td>
<td>(3,40)</td>
</tr>
<tr>
<td>Dimensions, Personological characteristics</td>
<td>0.82</td>
<td>16.51</td>
<td>.7964**</td>
<td>.0360</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9,38)</td>
<td></td>
<td></td>
<td>(5,38)</td>
</tr>
<tr>
<td>Dimensions, Exercises, Personological characteristics</td>
<td>0.63</td>
<td>13.21</td>
<td>.8192**</td>
<td>.0588</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12,35)</td>
<td></td>
<td></td>
<td>(8,35)</td>
</tr>
</tbody>
</table>

Note. MS: Mean Square for the effect. Enhanced condition: N=48; Standard condition: N=49. **p<.05. *p<.01. All analyses of model fit increase made relative to the normative (dimension-only) model.
### Table 7

Hierarchical regression analysis for each condition: Identification of the incremental nonnormative component in predicting the OAR

#### STANDARD CONDITION

<table>
<thead>
<tr>
<th>Predictor Set:</th>
<th>MS</th>
<th>F (df)</th>
<th>R²</th>
<th>R² change</th>
<th>F change (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercises only</td>
<td>2.33</td>
<td>31.03</td>
<td>.6741**</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(3,45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalological characteristics only</td>
<td>0.32</td>
<td>1.55</td>
<td>.1528</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(5,43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercises, Personalological characteristics</td>
<td>0.92</td>
<td>12.41</td>
<td>.7127**</td>
<td>.0386</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>(8,40)</td>
<td></td>
<td></td>
<td></td>
<td>(5,40)</td>
</tr>
</tbody>
</table>

#### ENHANCED CONDITION

<table>
<thead>
<tr>
<th>Predictor Set:</th>
<th>MS</th>
<th>F (df)</th>
<th>R²</th>
<th>R² change</th>
<th>F change (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercises only</td>
<td>1.55</td>
<td>14.99</td>
<td>.5054**</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(3,44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalological characteristics only</td>
<td>0.05</td>
<td>0.25</td>
<td>.0292</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(5,42)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercises, Personalological characteristics</td>
<td>0.62</td>
<td>5.64</td>
<td>.5365**</td>
<td>.0311</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>(8,39)</td>
<td></td>
<td></td>
<td></td>
<td>(5,39)</td>
</tr>
</tbody>
</table>

**Note.** MS: Mean Square for the effect. Enhanced condition: N=48; Standard condition: N=49. *p<.05; **p<.01. All analyses of model fit increase made relative to the exercise-only model.
condition tended to use only dimension information in making their overall ratings.

Table 7 presents tests of the $R^2$ increment for the personalological characteristics over and above the exercise variables. The incremental partial variance (Cohen & Cohen, 1983) of the personalological characteristics over the exercise variables was analyzed for a number of reasons. First, perceptions of exercise performance have been more widely studied in the assessment center literature. Also, analysis of the increment in prediction from using personalological information have typically been conducted to show the increment over behavioral assessments (Krzystofiak et al., 1988; Day & Silverman, 1989).

Subsequent analyses investigated the incremental variance explained by personalological characteristics over the explained by both the exercises and dimensions together. (Note that Table 6 presents analyses for incremental prediction over the dimensions alone.) The incremental explanation due to personalological characteristics was not statistically significant in either the standard or enhanced accountability condition ($F (5, 36) = 2.14, p>.05$, and $F (5, 35) = 0.97, p>.05$, respectively). However, exercises accounted for significant incremental variance over both dimensions and personalological characteristics in the standard condition, $F (3, 36) = 2.87, p<.05$, but not in the enhanced condition, $F (3, 35) = 1.47, p>.05$, a finding that
corroborates the role of exercise performance in decision-making. It may be concluded that assessors in the enhanced accountability condition used only the normative dimension-based performance information in making their overall judgments, while assessors in the standard condition relied on their perceptions of exercise and personalological dispositions in making ratings.

In summary, the pattern of increments in these model R^2's is consistent with the first hypothesis. Assessors in the enhanced accountability condition used only normative performance information in making their OAR judgments. The assessors in the standard condition tended to use both the exercise performance information and personalological characteristics information to make their OAR judgments. Assessors in the enhanced condition did not use either exercise or personalological information when making their overall judgments. Also, personalological characteristics seemed unrelated to the overall ratings, the dimensions, and the exercises.

MANOVA analyses were performed to test whether or not assessors in the enhanced accountability condition arrived at different ratings than did candidates in the standard accountability condition. Unit weighted means were tested between groups for dimensions, exercises, and personalological ratings. MANOVA results indicated that the difference in the unit-weighted composite dependent variables was not
significant between groups for the dimension rating constructs, \( F(5, 91) = 1.03, \eta^2 = .053, p > .05 \). Similarly, the personalological characteristics MANOVA revealed no significant differences, \( F(5, 91) = .25, \eta^2 = .014 \). The MANOVA for between-group effects on the exercise component was marginally significant, \( F(3, 93) = 2.42, p < .07, \eta^2 = .072 \). Finally, a \( t \)-test for the significance of the group difference in the OAR was not significant, \( t(95) = .40 \). No follow-up tests (univariate ANOVA's) were performed since no MANOVA's were significant at the \( p < .05 \) level. Table 8 presents the means and standard deviations of the rating variables for each group.

Since group differences existed for several measures depending on school attended (Rice or University of Houston; see Table 1), an analysis of covariance (ANCOVA) was performed to test for the moderating effect of school affiliation of participants. The number of months of residency in a dormitory or residence hall ("Residency") was chosen as the covariate for two reasons. First, it is recommended (Thornton & Byham, 1982) that managerial assessors should serve in a supervisory capacity at least one level above the assessed position. Residency reflects "organizational tenure" in this study. Second, unlike the nominal variable indicating whether or not the participant had ever resided in a dormitory or residence hall ("Dormitory"), the range and distribution of Residency was
Table 8

Rating variable means and standard deviations

<table>
<thead>
<tr>
<th>RATING VARIABLE</th>
<th>WHOLE GROUP</th>
<th>STANDARD</th>
<th>ENHANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis &amp; Judgment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>3.15</td>
<td>3.11</td>
<td>3.20</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.500)</td>
<td>(.479)</td>
<td>(.522)</td>
</tr>
<tr>
<td>Oral Communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>2.84</td>
<td>2.79</td>
<td>2.90</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.477)</td>
<td>(.530)</td>
<td>(.413)</td>
</tr>
<tr>
<td>Planning &amp; Organizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>3.05</td>
<td>3.07</td>
<td>3.03</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.509)</td>
<td>(.527)</td>
<td>(.495)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>3.16</td>
<td>3.16</td>
<td>3.12</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.484)</td>
<td>(.491)</td>
<td>(.481)</td>
</tr>
<tr>
<td>Leaderless Group Discussion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>3.11</td>
<td>3.07</td>
<td>3.14</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.415)</td>
<td>(.453)</td>
<td>(.376)</td>
</tr>
<tr>
<td>Counseling Session</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>3.07</td>
<td>3.14</td>
<td>2.98</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.468)</td>
<td>(.500)</td>
<td>(.426)</td>
</tr>
<tr>
<td>Roommate Conflict</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>3.12</td>
<td>3.09</td>
<td>3.15</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.459)</td>
<td>(.407)</td>
<td>(.510)</td>
</tr>
<tr>
<td>Neuroticism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>46.26</td>
<td>46.24</td>
<td>46.29</td>
</tr>
<tr>
<td>(SD)</td>
<td>(5.33)</td>
<td>(5.49)</td>
<td>(5.22)</td>
</tr>
</tbody>
</table>

(Table 8 continues)
(Table 8 continued)

**Extraversion**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>42.95</td>
<td>43.57</td>
<td>42.32</td>
</tr>
<tr>
<td>(SD)</td>
<td>5.94</td>
<td>5.79</td>
<td>6.08</td>
</tr>
</tbody>
</table>

**Openness to Experience**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38.99</td>
<td>39.52</td>
<td>38.45</td>
</tr>
<tr>
<td>(SD)</td>
<td>5.33</td>
<td>5.33</td>
<td>5.33</td>
</tr>
</tbody>
</table>

**Agreeableness**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>47.46</td>
<td>47.67</td>
<td>47.25</td>
</tr>
<tr>
<td>(SD)</td>
<td>5.85</td>
<td>6.42</td>
<td>5.27</td>
</tr>
</tbody>
</table>

**Conscientiousness**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46.00</td>
<td>46.17</td>
<td>45.82</td>
</tr>
<tr>
<td>(SD)</td>
<td>5.12</td>
<td>5.69</td>
<td>4.51</td>
</tr>
</tbody>
</table>

**Overall Assessment Rating**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.08</td>
<td>3.04</td>
<td>3.11</td>
</tr>
<tr>
<td>(SD)</td>
<td>0.453</td>
<td>0.465</td>
<td>0.443</td>
</tr>
</tbody>
</table>

*Note.* All Dimension, Exercise, and Overall assessment rating means based on a 1 (Low) to 5 (High) scale. Standard condition N = 49; Enhanced condition N = 48. * Summated scores for the Personelogical characteristics range from 8 (Low) to 72 (High).
the same for both schools. Also, Dormitory correlated significantly with age (Residency did not), possibly confounding age with ever having lived in a dormitory.

Results of the ANCOVA's are presented in Table 9. Interaction vectors were computed between the covariate and the rating variables for both conditions. In all models, the covariate interaction terms are not significantly related to the OAR. This set of results indicates that within each experimental group, there is homogeneity of regression for all levels of dormitory residency. This means that OAR judgments are not directly or indirectly affected by duration of dorm residency.

**Hypothesis 2: Accountability Effects On Assessor Confidence**

Hypothesis 2 stated that assessors who experience enhanced accountability to make accurate ratings would report themselves as being less confident about their ratings than would assessors whose accountability to make accurate ratings was not enhanced. Confidence estimates were assessed for the four dimensions and the OAR, and an overall confidence rating was developed as an average of all confidence estimates made by the assessor. The means and standard deviations of these confidence ratings are presented in Table 10.

Results of the MANOVA examining the mean difference in confidence ratings between groups indicates that there were no significant differences between the two groups for the
Table 9

Analyses of covariance for the assessor rating models: Prediction of the OAR using duration of dormitory residency as a covariate

<table>
<thead>
<tr>
<th>Predictor Set:</th>
<th>STANDARD CONDITION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MS</td>
</tr>
<tr>
<td><strong>Covariate:</strong> Residency</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>Covariate and:</strong></td>
<td></td>
</tr>
<tr>
<td>Dimensions</td>
<td>1.51</td>
</tr>
<tr>
<td>and Exercises</td>
<td>1.02</td>
</tr>
<tr>
<td>Dimensions and Personalological</td>
<td>0.82</td>
</tr>
<tr>
<td>Characteristics</td>
<td></td>
</tr>
<tr>
<td>Dimensions, Exercise,</td>
<td>0.66</td>
</tr>
<tr>
<td>Personalological Characteristics</td>
<td></td>
</tr>
</tbody>
</table>

**Covariate and interaction terms for:**

<table>
<thead>
<tr>
<th></th>
<th>MS</th>
<th>F (df)</th>
<th>(R^2)</th>
<th>(R^2) change</th>
<th>F change (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>0.86</td>
<td>12.85</td>
<td>0.7478**</td>
<td>0.0190^b</td>
<td>0.73 (4,39)</td>
</tr>
<tr>
<td>Dimensions, Exercices</td>
<td>0.56</td>
<td>9.77</td>
<td>0.8162**</td>
<td>0.0303^b</td>
<td>0.78 (7,33)</td>
</tr>
<tr>
<td>Personalological</td>
<td>0.46</td>
<td>8.22</td>
<td>0.8434**</td>
<td>0.0514^b</td>
<td>1.06 (9,29)</td>
</tr>
<tr>
<td>Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Table 9 continues)
(Table 9 continued)

Dimensions, Exercises, and Personalological Characteristics 0.37 8.33 .9006** .0690b 1.33 (25,23) (12,23)

## ENHANCED CONDITION

<table>
<thead>
<tr>
<th>Predictor Set:</th>
<th>MS (df)</th>
<th>F (df)</th>
<th>R²</th>
<th>R² change</th>
<th>F change (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariate:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residency</td>
<td>0.00</td>
<td>0.00</td>
<td>.0000</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariate and:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>1.41</td>
<td>29.37</td>
<td>.7817**</td>
<td>.7817a</td>
<td>36.7**</td>
</tr>
<tr>
<td>Dimensions and</td>
<td>0.91</td>
<td>19.97</td>
<td>.8079**</td>
<td>.8079a</td>
<td>22.8**</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions and</td>
<td>0.73</td>
<td>15.28</td>
<td>.8093**</td>
<td>.8093a</td>
<td>17.0**</td>
</tr>
<tr>
<td>Personalological Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions,</td>
<td>0.57</td>
<td>12.27</td>
<td>.8286**</td>
<td>.8286a</td>
<td>13.3**</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions,</td>
<td>0.81</td>
<td>16.99</td>
<td>.8052**</td>
<td>.0235b</td>
<td>1.12</td>
</tr>
<tr>
<td>Personalological Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions,</td>
<td>0.51</td>
<td>11.49</td>
<td>.8475**</td>
<td>.0396b</td>
<td>1.15</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions,</td>
<td>0.40</td>
<td>7.44</td>
<td>.8397**</td>
<td>.0304b</td>
<td>0.56</td>
</tr>
</tbody>
</table>

(Table 9 continues)
(Table 9 continued)

Dimensions, Exercises, and Personalological Characteristics

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.32</td>
<td>6.95</td>
<td>.8922**</td>
<td>.0636b</td>
</tr>
<tr>
<td></td>
<td>(25,21)</td>
<td></td>
<td>(12,21)</td>
<td></td>
</tr>
</tbody>
</table>

Note. **MS**: Mean Square of the effect. Standard N = 49; Enhanced N = 47. *a* Increment is assessed relative to the covariate-only model. *b* Increment is assessed relative to the covariate plus variable model.
Table 10

Means and standard deviations of confidence ratings for each dimension and the OAR (by condition)

<table>
<thead>
<tr>
<th>RATING VARIABLE</th>
<th>WHOLE GROUP</th>
<th>STANDARD</th>
<th>ENHANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis &amp; Judgment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>71.81</td>
<td>72.93</td>
<td>70.67</td>
</tr>
<tr>
<td>(SD)</td>
<td>(14.47)</td>
<td>(15.22)</td>
<td>(13.73)</td>
</tr>
<tr>
<td>Oral Communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>74.61</td>
<td>75.38</td>
<td>73.83</td>
</tr>
<tr>
<td>(SD)</td>
<td>(12.79)</td>
<td>(12.94)</td>
<td>(12.72)</td>
</tr>
<tr>
<td>Planning &amp; Organizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>73.31</td>
<td>74.37</td>
<td>72.22</td>
</tr>
<tr>
<td>(SD)</td>
<td>(13.08)</td>
<td>(13.64)</td>
<td>(12.54)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>72.96</td>
<td>73.71</td>
<td>72.19</td>
</tr>
<tr>
<td>(SD)</td>
<td>(13.21)</td>
<td>(13.36)</td>
<td>(13.16)</td>
</tr>
<tr>
<td>Overall Assessment Rating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>73.67</td>
<td>74.67</td>
<td>72.66</td>
</tr>
<tr>
<td>(SD)</td>
<td>(12.74)</td>
<td>(13.85)</td>
<td>(11.56)</td>
</tr>
<tr>
<td>Average Confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>73.27</td>
<td>74.21</td>
<td>72.31</td>
</tr>
<tr>
<td>(SD)</td>
<td>(11.83)</td>
<td>(11.96)</td>
<td>(11.74)</td>
</tr>
</tbody>
</table>

Note. Standard condition N=49; Enhanced condition N=48. All confidence means based on a 0% (Absolutely no confidence that the dimension rating is correct) to 100% (Absolutely sure that the dimension rating is correct) scale.
confidence ratings, $F(5, 91) = 0.15$, $\eta^2 = .008$, $p > .05$.
Additionally, there was no significant mean difference between groups for the overall confidence level, directional $t(95) = -.79$. Thus, the accountability manipulation does not appear to have decreased significantly the absolute level of confidence reported by the assessors in the enhanced accountability condition as compared to the assessors in the standard accountability condition.

The lack of a significant difference was unexpected. Further review of accountability research indicated a possible explanation for this finding. Previous examinations of rater confidence (Tetlock & Kim, 1987) have assessed the appropriateness of confidence for accountable raters. These researchers used a "Brier score" (Lichtenstein & Fischhoff, 1977), a metric that adjusts dichotomous decisions for expressed confidence. For dichotomous decisions, the Brier score (adapted with corrections from Tetlock & Kim, 1979, p. 704) is computed as:

$$\frac{1}{N} \left[ \sum (r_i - c_i)^2 \right] =$$

$$\alpha(1 - \alpha) + \frac{1}{N} \left[ \sum \operatorname{count}(r_i - c_i)^2 \right] - \frac{1}{N} \left[ \sum \operatorname{count}(c_i - \alpha)^2 \right] \quad (1)$$

where: $N$ is the total number of confidence assessments made by a given assessor (here, four dimension ratings and one overall rating for each candidate);

$r$ is the confidence assigned to a particular rating;

c is either 1 (if the rating matched that of experts)
or 0 (if not);

$c_t$ is the proportion of total ratings that were correct;

$n_t$ is the total number of confidence judgments that fell into each confidence category (i.e., the total number of "100% confidence" ratings, etc.);

$r_t$ is the numerical value of each confidence category (0, .1, .2, etc.);

$c_t$ is the proportion of confidence judgments in each category that were in fact associated with correct ratings.

A post hoc test was performed for the differences between conditions in Brier scores. It is unclear in the literature whether Brier scores may be used for polychotomous decisions (as in this study); however, using Brier scores, the difference in the levels of appropriate rating confidence between assessors in the enhanced and standard accountability conditions is statistically significant, directional $t$ (95) = 1.66, $p<.05$. In other words, post hoc analyses indicated that the manipulation seemed to have had the anticipated effect on the appropriateness, but not on the absolute level, of assessor confidence.

**Hypothesis 3: Accountability Effects On Rating Accuracy**

Hypothesis 3 stated that assessors in the enhanced accountability condition would make more accurate ratings than would assessors in the standard accountability condition. The relations between expert
and student ratings were assessed through the use of Cronbach's (1955) four components of rating accuracy: elevation; differential elevation; stereotype accuracy; and differential accuracy. Elevation is the tendency of the assessor to use the same part of all dimension scales for all candidates. Differential elevation is the assessor's tendency to rank-order the across-dimensions average rating of all candidates correctly. Stereotype accuracy is the assessor's tendency to rank-order the across-candidates average rating for each dimension correctly across all candidates, or the perception of the prevalence of certain dimensions in the sample. Differential accuracy is the assessor's tendency to rank-order candidates correctly within each dimension (Schneider et al., 1979). Finally, a composite accuracy component score was developed from the sum of the accuracy components. This measure reflects overall accuracy (Cronbach, 1955, Formula 3a, p. 192).

The data aggregation strategy for this hypothesis differed from tests for other hypotheses. For other analyses the ratings for each candidate are averaged to create one rating (for instance, of sensitivity) per assessor. However, in the accuracy component analyses ratings were not averaged. Each assessor's component scores reflect dimension rating accuracy across all three candidates. In all, five accuracy scores were computed for each participant.
Table 11

Rating accuracy component scores

<table>
<thead>
<tr>
<th>RATING VARIABLE</th>
<th>WHOLE GROUP</th>
<th>STANDARD</th>
<th>ENHANCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.210</td>
<td>0.231</td>
<td>0.187</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.156)</td>
<td>(0.167)</td>
<td>(0.142)</td>
</tr>
<tr>
<td>Differential Elevation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.372</td>
<td>0.392</td>
<td>0.352</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.156)</td>
<td>(0.157)</td>
<td>(0.154)</td>
</tr>
<tr>
<td>Stereotype Accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.307</td>
<td>0.321</td>
<td>0.293</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.155)</td>
<td>(0.166)</td>
<td>(0.144)</td>
</tr>
<tr>
<td>Differential Accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.725</td>
<td>0.739</td>
<td>0.711</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.163)</td>
<td>(0.149)</td>
<td>(0.177)</td>
</tr>
<tr>
<td>Overall Accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>1.614</td>
<td>1.683</td>
<td>1.543</td>
</tr>
<tr>
<td>(SD)</td>
<td>(0.379)</td>
<td>(0.373)</td>
<td>(0.376)</td>
</tr>
</tbody>
</table>

Note. SD=Standard Deviation. All Accuracy component means based on a 0.0 (perfectly accurate) to 1.0 (perfectly inaccurate) scale. Overall accuracy is the sum over all components for each assessor. Standard condition N = 49; Enhanced condition N = 48.
Table 11 shows the mean accuracy component scores of each group. A t-test for the overall accuracy component was significant as predicted, directional \( t (95) = 1.85, p<.05, \) \( \omega^2 = .02. \) A MANOVA of the four accuracy components between groups was not significant, \( F (4,92) = 0.95, p>.05, \) \( \eta^2 = .04. \) (The MANOVA matrix for the combined overall and four component comparison was singular.) These results indicate that the enhanced accountability manipulation increased overall dimension rating accuracy, but the effect of the accountability enhancement on particular facets of accuracy was statistically negligible though largely in the predicted direction.

**Hypothesis 4: Accountability Effects on Observation And Classification Accuracy**

Several indices were developed to assess predictions in Hypothesis 4. The first index described assessors's behavioral observations, while the second index described assessors's behavioral classifications of observations into dimensions. Appendix E contains the observations and classifications made by experts.

**Observation Accuracy.** Hypothesis 4 predicted that assessors in the enhanced accountability condition would record more accurate observations than would assessors in the standard accountability condition. The ratio of the number of correct behavioral observations (defined below) to the number of target behavioral observations was developed.
for each assessor; this ratio was then weighted by the total number of observations recorded by each assessor (CORRECT/TARGET<sub>pr</sub>) (Gaugler & Thornton, 1989; R. Martell, March 20, 1989, personal communication).

The assessment center experts who provided target scores also provided target behavioral observations. A target behavioral observation is a behavior recorded by the expert raters for each candidate in each exercise that is behaviorally specific and states what the candidate said or did (Gaugler & Thornton, 1989; McIntyre & Bentson, 1984). In this study, an observation was retained as a target behavioral observation if at least two of the three expert raters recorded the behavior on their observation forms.

In order to distinguish between observations made by experts and subject assessors in this study, an observation was called a target behavioral observation if agreed upon by the experts and a correct behavioral observation if made by an assessor. Subject assessors received a score of one toward their correct behavioral observation score for each target behavioral observation they recorded within each exercise. Finally, each assessor received a score for the total number of observations they recorded.

This study adopted the rationale of Gaugler & Thornton (1989), who noted that the correct-to-target ratio could be inflated spuriously due to the tendency of some assessors to make copious observations. Gaugler & Thornton (1989) called
this tendency "prolificacy". The effect of prolificacy was assessed in this study by dividing each assessor's correct-to-target ratio for each exercise by the number of behaviors recorded within that exercise (suggested by R. Martell, March 20, 1989, personal communication), resulting in an index called CORRECT/TARGET_{pro}. This index indicates the observation efficiency for each subject assessor. Specifically, assessors who recorded more correct observations, but fewer observations overall, than other assessors were more efficient than other assessors. Thus, more efficient assessors will have higher CORRECT/TARGET_{pro} ratios than will less efficient assessors. The index of observational efficiency, CORRECT/TARGET_{pro}, served as the dependent measure for the first test of Hypothesis 4.

Analyses indicated that assessors in the enhanced accountability condition had a significantly higher ratio of good-to-target observations after weighting for prolificacy (CORRECT/TARGET_{pro}) than did assessors in the standard accountability condition, M's = .053 and .044 respectively, directional t (95) = 5.28, p<.01, omega^2 = .22. This result indicates that assessors in the enhanced accountability condition were more efficient when making observations than were assessors in the standard accountability condition.

**Classification Accuracy.** It was also predicted that assessors in the enhanced accountability condition would classify their observations into dimensions more accurately
Table 12

Behavioral observation and classification efficiency mean scores and standard deviations

<table>
<thead>
<tr>
<th>RATING VARIABLE</th>
<th>WHOLE GROUP</th>
<th>STANDARD</th>
<th>ENHANCED</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORRECT/TARGET&lt;sub&gt;pro&lt;/sub&gt;:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>.048</td>
<td>.044</td>
<td>.053</td>
<td>5.28**</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.009)</td>
<td>(.008)</td>
<td>(.001)</td>
<td></td>
</tr>
<tr>
<td>CLS/TARGET&lt;sub&gt;pro&lt;/sub&gt;:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>.086</td>
<td>.082</td>
<td>.091</td>
<td>1.87*</td>
</tr>
<tr>
<td>(SD)</td>
<td>(.026)</td>
<td>(.027)</td>
<td>(.024)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Higher mean values indicate greater efficiency. All t-tests are directional; df for t-tests = 95. *=p<.05. **=p<.01. Enhanced condition: N=48; Standard condition: N=49.
than would assessors in the standard accountability condition. Using the same rationale for classifications as for observations (see above), scores for target behavioral classifications and correct behavioral classifications were developed. Expert assessors again provided the target behavioral classifications for this analysis. Subject assessors received one point toward their correct classification score for each of these target behavioral classifications they themselves correctly classified.

Classification accuracy efficiency was assessed as the averaged ratios of correct-to-target behavioral classifications within each dimension, averaged across dimensions and candidates. This value was then weighted by the number of classified observations within each dimension to indicate the effects of prolificacy on classifications (CLS/TARGET\textsubscript{pro}). The rationale for the construction of the CLS/TARGET\textsubscript{pro} ratio is similar to the rationale for CORRECT/TARGET\textsubscript{pro} above, i.e., to assess efficiency of classification. The difference is that CORRECT/TARGET\textsubscript{pro} focuses on efficient behavioral observations, while CLS/TARGET\textsubscript{pro} focuses on efficient behavioral classifications. These measures are conceptually linked in that classification accuracy is dependent on observation accuracy, as assessors cannot readily classify behaviors they did not observe.

As with CORRECT/TARGET\textsubscript{pro}, CLS/TARGET\textsubscript{pro} provides an
index of classification efficiency. Analysis of the classification efficiency measure indicated that assessors in the enhanced accountability condition exhibited significantly greater classification efficiency (higher CLS/TARGET<sub>pro</sub> scores) than did assessors in the standard accountability condition, directional $t = 1.87$, $p<.05$, $\omega^2 = .03$.

Taken together, results from tests of Hypothesis 4 indicate that enhanced accountability to justify one's observations and classifications led assessors to be more efficient than assessors in the standard condition as they recorded and classified observations. These results show that the process of assessment can be altered through enhancing accountability, even if the rating outcomes (see Table 8 above) were not significantly different between the two conditions.

**Generalizability of the Results**

**Proportion Of Significant Results.** One issue arising from a review of the results of this study may be the probability that the number of "significant" results reported above may have occurred by chance. Out of twenty-six tests, sixteen were significant at or beyond the $p<.05$ level. (These numbers do not include the findings from the covariate analyses.) Brozek & Tiede (1952) developed a measure to address this issue. Their Critical Ratio statistic is distributed as a standard score with mean = 0
and standard deviation = 1, and for this study the Critical Ratio value was 13.23, p<.001. Thus, one may be confident that the proportion of significant findings in this study did not arise by chance alone. (ANCOVA tests were not included, but adding them did not change the main conclusion to be drawn that the number of significant results did not arise by chance alone.)

**Comparability Of Effect Sizes.** Another issue may be the level of obtained effect sizes for this study's accountability manipulation. At the outset of this study, it was hypothesized that the effect of an accountability manipulation in the decision making and person perception literatures was at least "small" (Cohen, 1988), but the actual typical effect size was unknown. As much as the current study may contribute to knowledge about the effects of accountability, it is desirable to know the effect sizes both in the current study and in the accountability literature.

Table 13 contains the average effect sizes derived through a literature review of eleven studies examining the effects of accountability. This list is based on a review of commonly referenced accountability articles that have been published within the past ten years. The studies contained in this list have tested the effect of accountability in a variety of representative experimental situations (see Appendix F).
All main effect or two-way interaction effect sizes including accountability were included for analysis (no higher-order interactions were considered). All effect sizes were based on reported $\eta^2$'s, or $\eta^2$ as estimated from $F$-values or analyses of partial variances (Rosenthal & Rosnow, 1984, p. 357; Cohen, 1988, p. 412). The effect sizes reflected the influence of accountability on the following: (1) evaluative ratings made by subjects; (2) confidence levels reported by subjects; (3) accuracy component scores based on evaluative ratings; (4) measures of behavioral observation or classification accuracy; and (5) manipulation checks. For each study, the number of effect sizes reported ranged from 1 to 46 ($M=9.4$, $SD=14$). Four of the studies were published in Organizational Behavior and Human Performance or Journal of Applied Psychology, while the remaining seven studies were published in Journal of Personality and Social Psychology or Social Psychology Quarterly. Thirteen total experiments were reported. The sample sizes in these experiments ranged from 19 to 325 subjects ($M=92.5$, $SD=86.1$).

Each $\eta^2$ effect size from the literature review was transformed to standard score units ($z$; Rosenthal & Rosnow, 1984). After transformation, the mean and standard deviation of $z$ was computed within each study for the effect of accountability on each of the five areas listed above.

There are two interpretive cautions that should be
emphasized. First, for the "Evaluative Ratings" (dimension, exercise, and personalological ratings) effect size analysis, only ratings made by subjects are included; interpretive scores based on subject thought protocols (e.g., "integrative complexity") were not included because the current study did not use these types of measures. Also, there was a marked tendency in the literature to report $F$ values only for statistical tests that reached conventional significance levels ($p < .05$). Thus the literature effect sizes reported here may be inflated.

The effect sizes in the current study reflect averages of the effects for each hypothesis in $z$ units. For the obtained observation and classification accuracy effect sizes, only the MANOVA $F$ was used and not the $t$-tests. Similarly, only the MANOVA $F$'s for the manipulation check variables were used.

Results from the literature review indicated that the main effect of accountability is associated with what is defined as at least a "medium" effect size for each type of hypothesis tested in this study (according to Cohen, 1988, a medium effect size is indicated by $r = .30$ or $d = .5$). Table 13 presents 95% confidence intervals around the mean effect sizes to indicate whether the effect size results obtained in this study were likely to be similar to those obtained in the literature. Note that no confidence bands were estimated for accuracy components, analysis of partial
Table 13

Tests of differences for obtained effect sizes between this study and average effect sizes found in the accountability literature

<table>
<thead>
<tr>
<th>ANALYSIS</th>
<th>LITERATURE EFFECT</th>
<th>CIa</th>
<th>OBTAINED EFFECT</th>
<th>NUMBER OF ESTIMATESb</th>
</tr>
</thead>
</table>

1: FOR EVALUATIVE RATINGSC:

| MEAN      | .3086             | (.247, .370) | .1671 | 61 |
| SD        | .2446             |               |       |    |
| N         | 113               |               |       |    |

Analysis of Partial Variance

| MEAN      | .3661             | ------ | .1039 | 1  |
| SD        | 0.0               |         |       |    |
| N         | 120               |         |       |    |

2: FOR CONFIDENCE RATINGS:

| MEAN      | .3132             | (.229, .397) | .0718 | 7  |
| SD        | .1134             |               |       |    |
| N         | 89                |               |       |    |

3: FOR ACCURACY COMPONENTS:

| MEAN      | .3879             | ------ | .1940 | 2  |
| SD        | .0175             |         |       |    |
| N         | 60                |         |       |    |

5: FOR OBSERVATION/CLASSIFICATION ACCURACY:

| MEAN      | .3699             | ------ | .3541 | 3c |
| SD        | .131              |         |       |    |
| N         | 60                |         |       |    |

6: FOR MANIPULATION CHECK RATINGS:

| MEAN      | .3486             | (.137, .560) | .3430 | 13 |
| SD        | .3892             |               |       |    |
| N         | 177               |               |       |    |

(Table 13 continues)
(Table 13 continued)

Note. All tabled values reported in standard score units. N for all obtained effects in this study is 97. a 95% confidence interval of the standard error of the mean comparing obtained and literature effect sizes. b Indicates the number of effect size estimates comprising the literature effect. c Analysis of partial variance results reported in Gordon et al., 1988. d Too few studies to form stable estimates for intervals. e Effect size is for the number of observations recorded, not accuracy of recording.
variance, or for observation/classification efficiency. These measures have not been employed widely enough in the literature and are presented here only for the sake of completeness.

Results in Table 13 indicate that for the evaluative ratings and confidence ratings, the effects found in this study fell outside of the respective confidence intervals. These effects are not within the typical range of effects for accountability. For the manipulation check ratings, the obtained values are within the range expected in the literature.

It was noted in the description of the confidence measure that literature results concern appropriate confidence as opposed to absolute confidence. The effect size for appropriate confidence (.34) is within the range found for results in the literature.

**Power Analyses.** Power was estimated for each analysis reported in this study. Power was not estimated for literature effects.

Power estimates, reported in Table 14, were derived in three ways. First, exact power estimates were obtained through the MANOVA procedure and POWER option in the SPSSx data analysis system (SPSS, 1988). Next, in the case where several MANOVA's had been performed, the tabled value is the average of the exact MANOVA power levels. For instance, the reported power for evaluative ratings is .35. This value
Table 14

Estimates of power for the obtained accountability effect

<table>
<thead>
<tr>
<th>ANALYSIS</th>
<th>OBTAINED POWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: FOR EVALUATIVE RATINGS:</td>
<td></td>
</tr>
<tr>
<td>MANOVA</td>
<td>.35&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Analysis of Partial Variance</td>
<td>.29&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>2: FOR CONFIDENCE RATINGS:</td>
<td></td>
</tr>
<tr>
<td>MANOVA</td>
<td>.08&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>3: FOR ACCURACY COMPONENTS:</td>
<td></td>
</tr>
<tr>
<td>MANOVA</td>
<td>.29&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>t-Test</td>
<td>.45&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>4: FOR OBSERVATION/CLASSIFICATION ACCURACY:</td>
<td></td>
</tr>
<tr>
<td>t-Tests</td>
<td>.78&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>5: FOR MANIPULATION CHECK RATINGS:</td>
<td></td>
</tr>
<tr>
<td>MANOVA</td>
<td>.56&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note.  <sup>a</sup> Average of exact power values.  <sup>b</sup> Average of estimates.  <sup>c</sup> Exact power value.
reflects the average of MANOVA power values for the
dimension ratings (power=.35), the exercise ratings
(power=.59), and the personalological characteristics
(power=.11). The third method used for estimating power was
to average estimates based on analysis of partial variance
in regression.

Power values presented in Table 14 vary widely.
Several factors should be considered in evaluating these
results. First, there is considerable range in these power
estimates. For instance, for the manipulation check
ratings, the power of the three MANOVA's were .81 (MANOVA
for AMC1), .12 (AMC2), and .75 (AMC3). Similar ranges occur
for other tests. Also, note that the reported power for
absolute confidence (.08) is lower than the power for
appropriate confidence (.52).

These power estimates are attenuated due to three
factors (Cohen, 1988). First, there are differences in
construct reliability. The MANOVA for exercise ratings had
a power of .59, while the MANOVA power for the somewhat
less-reliable personalological characteristics was .11. Higher
reliability was associated with higher power. Second,
reliability was attenuated from averaging across stimuli.
Discrepancies in reliabilities were found for candidate-
level vs. aggregate-level ratings. Finally, the sample size
for these analyses were smaller than desirable, especially
in light of the uneven reliabilities. A larger sample size
would have increased effect size power considerably.

DISCUSSION

The imposition of enhanced accountability for assessors to make accurate observations and ratings resulted in the following, as compared to assessors in the standard accountability condition: (1) greater efficiency in making behavioral observations and classifications; (2) reliance on normative (dimension-only) performance information when making overall assessment ratings; (3) higher overall dimension rating accuracy; and (4) more appropriate expressed confidence in ratings. Enhanced accountability did not affect either particular facets of rating accuracy, the level of confidence in ratings, nor the ratings made in the assessment center.

Accountability. Accountability had a broad range of effect on assessor behavior in this study. Enhanced accountability altered assessors' use of assessment center performance information as compared to that of assessors in the standard accountability condition. Overall, assessors in the enhanced accountability condition made dimension ratings that were more accurate than those made by assessors in the standard accountability condition (though facets of accuracy were unaffected). Also, assessors within the enhanced accountability condition were more efficient in observing and classifying their observations than were those in the standard accountability condition, and had more
appropriate confidence in their ratings.

The effect of enhanced accountability on the absolute level of confidence assessors held in their ratings was negligible. As mentioned above, this may be a methodological artifact. Analysis of Brier scores in the current study led to the post hoc conclusion that the imposition of enhanced accountability resulted in significantly more appropriate confidence in ratings. Thus, the conclusion that the imposition of accountability had "no effect" on confidence may not be warranted. Instead, enhanced accountability seems to have had an effect on appropriate confidence.

The effects of enhanced accountability in this study have implications for studies of social judgment. Both assessment center developers (Thornton & Byham, 1982) and social judgment researchers (Fiske & Taylor, 1984) have emphasized that interpersonal feedback is the mechanism through which personal judgment idiosyncrasies are reduced. One contribution of this study is the demonstration that interpersonal communication alone does not necessarily lead to better judgments (setting aside for now the potential negative effects of group decision-making).

Enhanced accountability changes the structure of the judgment task through the mediating effect of the pressure to justify one's judgments. A comparison of judgments made in interpersonal situations where there seems to be a pre-
existing level of accountability, such as in the standard accountability condition of this experiment, to judgments made under conditions of enhanced accountability, indicates that the judgment process can be altered appreciably by this accountability manipulation.

Another implication of this experiment for future applications of accountability is in refining the distinction between justification and explanation of judgments. In this study, the standard accountability condition instructions presented assessors with the possibility that they would need to explain their classifications to other assessors in their integration groups. Regardless, assessors in the standard condition had significantly lower observation and classification efficiency than assessors in the enhanced condition, who were told that they may need to justify their observations and classifications. It could be that other applied studies may have failed to find a significant accountability effect (Gordon et al., 1987) because they did not emphasize the need to justify judgments, not just explain them.

Assessment Center Method. This study replicated the findings from field research (e.g., Bycio et al., 1987) that assessors's judgments are influenced by nonnormative data. Specifically, the source of this influence seems to have arisen from situational cues. Assessors in the standard accountability condition tended to use both normative
(dimension) and nonnormative (exercise, personallogical) performance information in their decisions about the candidates' overall performance.

Assessors in the standard accountability condition tended to use personalological information in making their overall ratings of candidate performance. However, this effect was not as strong as the effect of personallogical evaluation in related research. Other literature investigating the effects of personalological component IPC's (Krzystofiak et al., 1987; Raza & Carpenter, 1988; Day & Silverman, 1989) has been more successful in demonstrating the relatedness of personalological evaluations.

Even though personalological evaluations on behavioral ratings have been significant in other applied research contexts, there are at least three alternative explanations for the relative lack of personalological contribution in this study. First, the personalological constructs themselves were not explicitly job-related. Instead of broad personalological constructs, future assessment of more specific and fidel personallogical work dispositions may increase the predictiveness of this type of rating construct (Anastasi, 1988; Guion, 1987). For example, assessors may be able to match personalological "types" of candidates to "types" of jobs (as in Jackson's interview research) if the personalological constructs are redefined for specific assessment center situations based on incumbent or expert
ratings (see Rynes & Gerhart, 1990).

Second, assessment centers are designed so that evaluation of personality factors is minimized (Thornton & Byham, 1982). Thus, it may be that the behavioral focus of the assessment center decreased the chance of finding a significant effect of personalological characteristic evaluations.

Finally, the candidate stimuli in these tapes may not have exhibited enough range of variability in personalological characteristics. In a functional assessment center, candidates may be irritable and under duress, or even excited and active. With a greater range of candidate personalities may come greater salience of personality for making assessment decisions.

The effect of exercise evaluations was more in line with previous research findings, but these results show that exercise evaluation is more extensive than thought earlier. In their review of assessor decision-making, Adams & Thornton (1988) noted that previous findings of exercise influence resulted when assessors made within-exercise dimension ratings. This practice may have increased the chance of finding an exercise performance influence on ratings, as Zedeck (1986) predicted, by increasing the salience of the behavioral medium -- the exercise. In other words, within-exercise dimension ratings may unintentionally lead to high method variance through emphasis on the same-
surface features of the behaviors. Similarly, Silverman et al. (1986) presented evidence that this finding may be a methodological artifact induced by requiring assessors to think of candidate performance within, as opposed to across, the assessment center exercises.

Alternatively, the process of making within-exercise ratings may have unintentionally resulted in the introduction of systematic rating bias ("areal" bias; Wherry & Bartlett, 1982). Areal bias refers to rating bias introduced from contacts with the ratee in distinct contexts ("areas", or potentially exercises).

Areal bias may be an important component of assessor ratings. Assessors must decide how observable each dimension is in each exercise for the purpose of combining information. To the extent that an assessment center is designed so that the only relevant true rating score variance is dimension-based, all other sources of performance information are either unspecified or irrelevant. However, when making within-exercise ratings, assessors appear to give nonzero statistical weight to an areal component such as "counseling session performance variance". It may be that within-exercise dimension ratings are intercorrelated to the extent that they are areal (reflect different facets of performance within the same exercise).

The current study discovered similar evidence of areal
bias in ratings made by assessors in the standard accountability condition. However, unlike previous studies, exercise ratings were made at the close of the assessment center and dimension ratings were made across exercises. Thus the 'surface feature salience' argument cannot explain the significant influence of exercise performance on ratings made by assessors in the standard accountability condition. These findings suggest that contextual, or areal, information are a significant component of assessor decisions in the assessment center even when exercise ratings are made after all dimension ratings had been made.

Is this finding a consequence of assessment center design, such that dimensional performance is inextricably bound into the schematic functions of exercises, as in Zedeck's (1986) model? This study provides evidence that this may not be the case. The unique variance attributable to exercises is a significant 5.5% of the total OAR in the standard accountability condition. Furthermore, the imposition of enhanced accountability seems to have assessor reliance on exercise and personalogical performance variance. In Zedeck's model, there is no mechanism to explain this result. This does not seem to support the hypothesized organizing function of an exercise schema: shouldn't exercises account for more variance in assessors' overall judgments if exercises organize performance perception and guide performance recall? The question may
be a theoretical one; specifically, instead of whether exercise performance information contributes to dimension ratings, the question may be how much variance should exercise performances contribute to ratings.

The influence of exercise performance on the judgments of assessors in the standard accountability condition may be explainable through the "person-within-situation" phenomenon (Mischel, 1984). The diagnosticity of a dimension for predicting the OAR is related to the dimension's consistency across exercises (which is the rationale for making across-exercise dimension ratings). Variability due to exercises may reflect the prototypicality of certain behaviors emitted within exercises, which is why multiple exercises are used. Behaviors that are more prototypical across each temporal interval (i.e., exercise) should be indicative of consistent performance. Differences in terms of the prototypicality of behaviors emitted within each exercise reflect inconsistency in performance.

The assessor needs to determine whether behavioral inconsistency is diagnostic or whether it is due to the (in)ability of the exercise to bring out the desired dimension-related behaviors. If the inconsistency in behavior prototypicality is due to unusual behavior within a particular exercise, the increment in performance variance due to exercises may reflect salient and diagnostic exercise variability. Enhanced accountability seems to have reduced
the salience of exercise variability. This attenuation is "good" to the extent that it reinforces the content validity of decision processes in the assessment center.

Wherry's psychometric theory of ratings (Wherry & Bartlett, 1982) may help explain the effect of accountability on the influence of nonnormative information on assessors' judgments. In his model, performance ratings consist of true variance, systematic bias variance, and random error variance in both perception and recall. This model is adversarial because the sources of variance (true, bias, and error) have weights that when squared must sum to unity. Under most circumstances, this model states, the effects of bias on obtained ratings are substantial since there are many sources of bias. In particular, memory is affected by perceptual defects and schema-induced recall biases, both of which may be related to exercises. (Wherry's model seems to make allowance for appropriate IPC's.)

Wherry's Theorem 12 and subsequent corollaries (Wherry & Bartlett, 1982) address the effects of enhanced accountability. His model postulated that the weights of "true" perception and recall factors increase in opposition to the weights of areal and overall bias. This appears to have happened in this study, where the importance weights of exercise-specific and personalological-specific performance were negligible in the enhanced accountability condition.

Another consideration of these data arises from the
perspective of IPC's. Imposing accountability seemed to change the nature of the IPC's for assessors in the enhanced condition as compared to those in the standard condition. Based on this study and those reported in the literature, "exercise" and "personological evaluation" may be two natural categories that assessors tend to use along with dimension information when making overall judgments.

The psychometric, person-within-situation, and IPC explanations presented above are tenable, and may be consistent with each other. This study was not designed to test one model against another, though, and so their applicability must be judged in hindsight. The results of this study should lead to further research which directly tests predictions made by each model. These tests would be important for the future development of assessment center content.

Determining which situations are most indicative (and contraindicative) for different assessment center applications could be achieved through the social judgment vantage point (Mischel, 1984; Bycio et al., 1987). Alternately, situational assessment center IPC's, or IPC's based on "natural" performance categories, may be more realistic than dimension based IPC's. An experiment could be designed to assess the effectiveness of IPC's (measured by accuracy components or confirmatory factor analysis) within these indicative situations to examine the method
variance vs. performance category explanations. Finally, it may be that simple design techniques such as enhanced accountability or statistical control of exercise variance could increase assessor reliance on desirably diagnostic dimensions. In this case, the psychometric approach may prove most advantageous.

As noted above, statistical power in this study was low even though the levels of effect sizes were generally appropriate given published studies of accountability. Of course, the low sample size contributed greatly to this finding. However, some effects had high power while others didn't; this may have resulted from construct unreliability. Ratings for each construct in this study were averaged across stimuli in order to facilitate generalizations to other stimuli. It appears that generalizability was bought with the price of lowered reliability, and therefore power, due to the data analysis strategy. It is possible that considered separately, the effects were much more powerful for one candidate than for others, but this is a theoretically and empirically uninteresting result. Given this data analytical option, it is preferable to have found an appreciable proportion of significant findings with moderate power than only a few with higher power. This indicates that the results reported in this study may be a conservative indicator of the true effects of enhanced accountability in assessment centers.
Even if a more appropriate assessment center IPC should be exercise-based (Bycio et al., 1987), the main issues remain those of increasing content- and criterion-related validity and reducing inaccuracy of measurement in the assessment center. As compared to results for assessors in the standard accountability condition, the imposition of enhanced accountability increased the efficiency of observations and classifications made by assessors, the appropriateness of their confidence in their ratings, and the overall accuracy of their ratings. All of these issues are related to the construct validity of the assessment center.

A striking result in this study was that the imposition of enhanced accountability was clearly more effective in altering assessors' observations and classifications than in altering their evaluations of candidate performances. There were no rating variable score differences between the two conditions, but assessors were clearly better at making observations and classifications when put in a situation of higher accountability. In other words, the judgment process was affected by enhanced accountability, but the decision outcomes remained the same compared to those made under standard accountability conditions.

There are several implications of these findings for future research. In this study, only two levels of accountability were compared. Other types of accountability
(e.g., to the candidate; to minorities) may influence rating outcomes; in any case, a wider sampling of accountability enhancement manipulations should be examined in applied settings to determine their relative effects. For instance, meta-analytic evidence has shown that assessment centers conducted for experimental purposes have superior predictive validity; this may be the result of some level of enhanced accountability (Gaugler et al., 1987). Additionally, accountability may be enhanced by altering the process assessors use to make final judgments. Specifically, personal accountability may reduce tendencies toward group polarization or domination by one assessor within an integration group (Thornton & Byham, 1982). Future research should examine the influence of different decision-making strategies on assessors' perceptions of accountability and the quality of their judgments.

Surprisingly, enhanced accountability affected the rating process (CORRECT/TARGET \textsubscript{pro} and CLS/TARGET \textsubscript{pro}) but not rating outcomes. This was unexpected, since it is widely believed that better observation processes note-taking, accurate behavioral classification, etc. lead to better performance ratings (Guion, 1986).

Further research should investigate the linkages between the processes of observation and classification, and the nature of subsequent information use in deriving ratings. There are a number of reasons why the manipulation
in this study did not alter rating outcomes. The accountability manipulation may not have been powerful enough over the entire experimental session to alter rating results. Alternately, it may be that crucial elements have not been properly identified. Other factors may have influenced rating process. For instance, it could be that the target ratings should have been made by experts who were more similar to the subjects. Also, it may be that assessors would have been more accurate rating exercises rather than dimensions.

One final reason why the observation and classification process may not have affected rating outcomes is that the link between process and ratings is more subtle than suspected. This study took as a premise that observation efficiency would result in rating accuracy, a premise that has been widely accepted in performance appraisal research (Guion, 1986). This premise may hold true for dimensions that are globally defined or trait-related, but it may not for dimensions (such as the ones in this study) that are behaviorally specific.

Postscript

The title of this study, "Is there judgment bias in the assessment center method?", remains a central question for assessment center research. Given the considerable body of evidence that bias exists in other forms of personnel evaluation (interviews, performance appraisal), we should
expect that the answer to the title's question is "yes". However, whether or not this bias is a cause for alarm is a matter for future enhanced accountability research.

CONCLUSIONS

The following conclusions may be drawn:

1: Assessors whose accountability to make accurate ratings was not enhanced used nonnormative sources of performance data in making assessment center decisions;

2: Enhanced assessor accountability influenced assessors to use only normative, dimension-based information when making OAR's;

3: Assessors in the enhanced accountability did express increased appropriateness in the confidence they held in their ratings as compared to assessors in the standard accountability condition, but the overall confidence expressed by assessors did not differ significantly between the two conditions;

4: Assessors in the enhanced accountability condition had higher overall accuracy in their dimension ratings, though not facets of accuracy, as compared to assessors in the standard accountability condition;

5: Assessors in the enhanced accountability condition had higher levels of behavioral observation and classification efficiency than did assessors in the standard accountability condition.
REFERENCES


Campbell, J.P. (1986). Labs, fields, and straw issues. In E.A. Locke (Ed.), Generalizing from laboratory to field


Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. Organizational Behavior and Human Performance, 3, 552-564.


APPENDIX A

Target Dimension Scores Made By Expert Raters.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Brad</th>
<th>Chuck</th>
<th>Greg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis &amp; Judgment</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Oral Communication</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Planning &amp; Organizing</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Overall Assessment Rating</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
APPENDIX B

Training Session Forms And Script.

TRAINING OUTLINE


"I'd like to thank you all for coming here today! Your help is going to go up a long way toward making this research project a success!"

"The purpose of this experiment is to evaluate a new technique called an assessment center for hiring resident hall student assistants."

"An assessment center is not a place. It is a method that is used to help make employment decisions, such as the hiring of a job applicant, the promotion of an employee, or the development of an employee."

"In an assessment center, job applicants or employees are put through a series of standardized, job-related simulations that are specially designed to bring out the important behaviors necessary for good performance on the job in question."

"In an assessment center, the candidate, who are usually called 'assessee', are observed by trained people, called 'assessors', under controlled conditions for a specific amount of time. Today, you will be trained as assessors so that you can evaluate student assistant candidates later on."

"Finally, now that you're here today, I'm assuming that
you'll try your best to help this research project succeed."

---Residence Hall Student Assistant

"A residence hall student assistant (SA) is an upper-level undergraduate who is assigned to a floor section of a residence hall (or college). SAs are responsible for working closely with other SAs to develop and maintain an atmosphere which promotes academic, personal, and social growth in the residence hall."

"An SAs duties include: assisting in counseling students; ensuring that students develop an appropriate academic atmosphere in the residence hall; promoting consideration of individual needs in a group living environment; communicating and enforcing residence hall policies and procedures; coordinating educational and social/recreational programs for students; and promoting security awareness. Is it pretty clear what an SA is supposed to do?"

---The assessment center itself

"There will be several phases of this experiment. I'd like to give a brief overview of each to familiarize you with what's going to happen. You don't have to memorize these phases, and you can ask the administrator during the experiment any questions you have at the time. But please do pay attention so that you'll know what you have an idea of what you'll be doing in this experiment."

"You'll observe three candidates for the SA position,
each of whom will participate in three exercises: a discussion of University policy among the candidates; a counseling session with a troubled student; and a conflict between two roommates. You won't observe each candidate in every exercise, but you will observe each candidate in one exercise."

"Your main job will be to make detailed notes on what the candidates say and do in each exercise. You'll classify the behaviors you observed and recorded into well-defined dimensions of performance, such as oral communication and sensitivity."

"After observing the candidate, you'll join others to form a group of assessors. You'll become part of an assessor team and share information on candidates, one candidate at a time. For each candidate you will independently rate each dimension, taking into account the candidate's performance in each of the exercises; than you will independently evaluate the candidate's overall performance in the assessment center."

---Observing and recording behavior

"The training you will receive today will focus on how to observe and record behaviors accurately. It is very important to make good behavioral observations of others. Good behavioral observations enable you to make accurate evaluation of others and allow you to communicate successfully to other assessors what the student assistant
candidate actually said and did."

"You have a handout that reads 'Tips for observing and recording behaviors'—if you could follow along with me for a moment, let's go over this form."

**Good observations:**

--State what a person says or does (for instance, "he said to the student, 'I don't have time to discuss your problem");

--Are specific rather than general ("she suggested a follow-up meeting in 10 days");

--Are descriptive rather than evaluative ("the student kept talking when he was trying to start the discussion"); and

--Are confirmable by others ("he took notes throughout the meeting").

**Poor observations,** on the other hand:

--Make general classification statements (such as "she was sensitive");

--Interpret actions ("the student was getting on the SA's nerves");

--Impart feelings ("he was angry"); and

--Describe underlying personality make-up ("he is paranoid").

"Is it fairly clear what make a good or poor observation?"

---Observation practice #1: BET
"Now you're going to have practice identifying good and poor observations. On the sheet labeled 'Behavior Example Test', please read each observation and indicate on your form whether each is good or poor observation. Once you're all finished, we're going to go over the correct answers."

---Observation practice #2: Nancy

"OK, now you're going to get some practice making actual behavioral observations by watching a tape of a group of people who must decide how to allocate money to various employees. Each of these candidates represents a different employee and is arguing for that employee getting a raise."

"Your goal will be to watch the tape and write down both verbal and nonverbal behaviors. Please use the blank piece of paper I gave you, and put your name on the form first; I'll be collecting these at the end of training and I'll go over your observations and give you feedback on how well you did."

"Remember that making good behavioral observations is one of the most important tasks you'll have in this experiment. Keep in mind that you will have to write what each person says and does quickly; you won't have time to write complete sentences or to use correct grammar. You do want to write verbatim quotations if possible, the context of what's happening to help interpret what the person did, and any nonverbal behaviors."

"You'll be making observations on someone named Nancy."
I'll start off the tape making observations to give you an idea for what we're looking for in observations. Use the Behavior Example Test sheets to make your observations." (Model observations for George; let them watch Nancy for 2 minutes).

"How was that? It gets easier as you get more practice!"

---Dimensions

"Now let's talk about classifying behaviors into dimensions. After watching each tape you will classify the behaviors you observed into dimensions of performance. These dimensions have been identified as important for performance of the SA's job. Please follow along with me as I read the definitions for each dimension on the Minimal Standards of Performance sheets I handed out."

**Oral Communication:** The ability to express oneself effectively in individual or group situation; includes gestures and other nonverbal behaviors; conveys thoughts clearly and concisely; does not go off on tangents.

**Sensitivity:** The ability to appraise accurately the needs, feelings, skills, and competencies of others in interpersonal situation and act accordingly.

**Planning and Organizing:** The ability to establish a course of action for self and/or others to accomplish a specific goal; makes proper assignments of personnel and allocates resources appropriately.
**Analysis and Judgement:** The ability to identify problems and possible causes; gathers relevant information and relates data from different sources to solve problems.

"Do you have any questions about these definitions?"

"As I mentioned earlier, after you've made your observations you'll classify behaviors into dimensions as follows. Consult the minimal standards of performance for each dimension and compare them to your observations. The minimal standards are behavioral examples that a candidate could be expected to perform for that dimension in that exercise. You'll see that one of the minimal standards for Oral Communication is "spoke clearly, concisely, and fluently"; you'll read over your observations and decide which observations come closest to that minimal standard, and write it down for that dimension. For instance, Nancy looked at the other candidates during her speech, repeated herself, looked down at her notes-- all of these are examples of oral communication as exemplified by the minimal standards, and so you'd write down that Nancy performed each of these behaviors on your sheets."

---Ratings

"Once you've discussed each candidate's performance, you'll need to make an evaluation of each dimension for each candidate, one candidate and one dimension at a time. You'll share observations for each dimension and each candidate you observed with other assessors. You'll compare
your pooled observations to minimal standards of performance for each dimension. If the candidate exhibited most of the minimal standards of performance for that dimension, you'd give him a "3"--acceptable performance, not low, not high. If he didn't meet most of the minimal standards, you'd give him a "2"--fairly low performance on the dimension. And if he hardly met any of the minimal standards or did a particularly poor job, you'd give the candidate a "1"--low performance on the dimension."

"On the other hand, if the candidate met the minimal standards and did a fairly good job, you might give him a "4"--fairly high performance on the dimension. Finally, if the candidate meets most of the minimal standards of performance for that dimension and does a very good job in your opinion, you might give a "5"--high performance on the dimension."

"Please keep in mind that a 3 doesn't mean AVERAGE performance; a 3 means that the candidate met most of the minimal standards for that dimension, performing acceptably well. Also, only compare the candidates against the minimal standards, not in comparison to each other."

"I realize that's a lot of ground to cover today; do you have any questions I could answer right now? Remember that you can ask questions at any time during the experiment."

"The next session is at _______. Please be on time. I'll see you all then! Thanks!" Collect materials.
CHARACTERISTICS OF GOOD AND POOR BEHAVIORAL OBSERVATIONS

Good observations:
State what a person says or does
Are specific rather than general
Are descriptive rather than evaluative
Are confirmable by others

Poor observations:
Make general classification statements
Interpret actions
Impart feelings
Describe underlying personality make-up
BEHAVIOR EXAMPLE TEST

Read each statement below and decide whether it is a behavioral statement, or whether it is too general, vague, evaluative, subjective, etc. The statement may be an example of something "effective" or "ineffective". The question you are to decide is whether it is a behavior or someone's subjective report.

Imagine that you were listening to someone report observations from an exercise. If that observer made this statement, would you know what the student assistant actually said or did? If you think it is a behavioral statement, put an "X" in the "good example" column. If you consider it too vague or general, put an "X" in the "poor example" column.

<table>
<thead>
<tr>
<th></th>
<th>Good Example</th>
<th>Poor Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. GOT UP AND MOVED HIS CHAIR CLOSER TO THE STUDENT.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. WROTE DOWN THE STUDENT'S SUGGESTIONS.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. HER ARGUMENT BROKE DOWN.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. WAS VERY SENSITIVE TO THE STUDENT'S REQUEST.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. SUGGESTED GOING TO A MOVIE.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. DIRECTIONS FOR PERFORMING THE TASK WERE UNCLEAR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. TOLD THE SA HE WASN'T BEING CLEAR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. BEGAN THE PRESENTATION BY IDENTIFYING THE PROBLEM.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. WAS TOO SYMPATHETIC.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. GAVE THE STUDENTS THREE OPPORTUNITIES TO SUGGEST REASONS FOR THE PROBLEM.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

Experimental Session Forms.

BEHAVIOR OBSERVATION FORM

Group Discussion Exercise

Assessor #:___________  Candidate:_________________
BEHAVIOR OBSERVATION FORM

Counseling Session Exercise

Assessor #:__________  Candidate:_________________
BEHAVIOR OBSERVATION FORM

Roommate Conflict Exercise

Assessor #:__________  Candidate:______________
LIST OF DIMENSIONS

Oral Communication: The ability to express oneself effectively in individual or group situations; includes gestures and other nonverbal behaviors; conveys thoughts clearly and concisely; does not go off on tangents.

Sensitivity: The ability to appraise accurately the needs, feelings, skills, and competencies of others in interpersonal situations and act accordingly.

Planning and Organizing: The ability to establish a course of action for self and/or others to accomplish a specific goal; makes proper assignments of personnel and allocates resources appropriately.

Analysis and Judgment: The ability to identify problems and possible causes; gathers relevant information and relates data from different sources to solve problems.
MINIMAL STANDARDS OF PERFORMANCE:

ROOMMATE CONFLICT

Oral Communication:

The ability to express oneself effectively in individual or group situations; includes gestures and other nonverbal behaviors; conveys thoughts clearly and concisely; does not go off on tangents.

--Spoke clearly, concisely, and fluently

--Did not have to repeat self to be understood

--Used acceptable grammar

--Spoke in a conversational tone and volume (i.e., voice didn't crack; no fast-paced or high-pitched voice; no giggling or stuttering)

--Maintained train of thought (i.e., not easily diverted more than once by other topics or intruding thoughts)

Sensitivity:

The ability to appraise accurately the needs, feelings, skills, and competencies of others in interpersonal situations and act accordingly.

--Listened attentively (i.e., maintained forward posture; nodded; maintained eye contact with the speaker)

--Did not interrupt the roommates

--Expressed an understanding of each roommate's feelings and needs

--Didn't take sides

--Made at least one positive and reinforcing statement about each roommate

--Was persuasive, but not coercive (i.e., did not "give orders" to the roommates, but got them to improve their relationship)
ROOMMATE CONFLICT (continued)

Analysis and Judgment:

The ability to identify problems and possible causes; gathers relevant information and relates data from different sources to solve problems.

--Used information provided by both students to identify the problem

--Followed up on clues provided by the roommates, and did not merely follow the interview plan

--Encouraged both John and Steve to express their ideas regarding the source of their conflict

--Gathered information about possible causes of their conflict before coming up with a plan for the roommates to get along better (e.g., how come you are not getting along? When do each of you like to study?)

--Sought common ground in each roommate's position as a start toward resolving the conflict

Planning and Organizing:

The ability to establish a course of action for self and/or others to accomplish a specific goal; makes proper assignments of personnel and allocates resources appropriately.

--Made an initial statement of purpose at the start of the interview

--Ensured that the discussion moved along (i.e., did not get bogged down in one area of concern, but devoted a few minutes to each area)

--Devised a plan for the roommates to follow to improve their relationship and asked for commitment from them to follow the plan

--Set up a follow-up meeting, specifying its time and date, to review the roommates' progress with their problems
MINIMAL STANDARDS OF PERFORMANCE:

LEADERLESS GROUP DISCUSSION

Oral Communication:

The ability to express oneself effectively in individual or group situations; includes gestures and other nonverbal behaviors; conveys thoughts clearly and concisely; does not go off on tangents.

--Spoke clearly, concisely, and fluently
--Did not have to repeat self to be understood
--Used acceptable grammar

--Spoke in a conversational tone and volume (i.e., voice didn't crack; no fast-paced or high-pitched voice; no giggling or stuttering)

--Maintained train of thought (i.e., not easily diverted more than once by other topics or intruding thoughts)

Sensitivity:

The ability to appraise accurately the needs, feelings, skills, and competencies of others in interpersonal situations and act accordingly.

--Questioned the ideas of others in a supportive manner (e.g., did not insult the other group members or their ideas)

--Showed evidence of awareness of the possible reactions of other students and the administration to suggestions for enforcing the alcohol policy

--Listened attentively (i.e., maintained forward posture; nodded; maintained eye contact with the speaker)

--Did not interrupt the other group members
LEADERLESS GROUP DISCUSSION (continued)

Analysis and Judgment:

The ability to identify problems and possible causes; gathers relevant information and relates data from different sources to solve problems.

--Encouraged other group members to express their ideas

--Gathered information about the possible causes of the alcohol problem before coming up with a plan to solve the problem

--Examined the strengths and weaknesses of proposed solutions

--Incorporated ideas of other group members when suggesting solutions to the problem

Planning and Organizing:

The ability to establish a course of action for self and/or others to accomplish a specific goal; makes proper assignments of personnel and allocates resources appropriately.

--Made an initial statement of purpose at the start of the discussion

--Ensured that the discussion moved along (i.e., did not get bogged down in one area, but devoted a few minutes to each area; paid attention to how much time was left for discussion)

--Made statements about what should be done with the suggestions the group came up with (e.g., print them in the student newspaper)

--Showed evidence of handling the problem in a methodical, organized way (e.g., took notes during the discussion; organized the group's consensus into a clear, concise statement)
MINIMAL STANDARDS OF PERFORMANCE:

COUNSELING SESSION

Oral Communication:

The ability to express oneself effectively in individual or group situations; includes gestures and other nonverbal behaviors; conveys thoughts clearly and concisely; does not go off on tangents.

--Spoke clearly, concisely, and fluently

--Did not have to repeat self to be understood

--Used acceptable grammar

--Spoke in a conversational tone and volume (i.e., voice didn't crack; no fast-paced or high-pitched voice; no giggling or stuttering)

--Maintained train of thought (i.e., not easily diverted more than once by other topics or intruding thoughts)

Sensitivity:

The ability to appraise accurately the needs, feelings, skills, and competencies of others in interpersonal situations and act accordingly.

--Listened attentively (i.e., maintained forward posture; nodded; maintained eye contact with the speaker)

--Did not interrupt Bob

--Expressed an understanding of Bob's feelings and needs

--Made at least one positive and reinforcing statement about each Bob

--Was persuasive, but not coercive (i.e., did not "give orders" to Bob, but got Bob to improve his hygiene and to get him help from others to deal with his family problems)
Counseling Session (continued)

Analysis and Judgment:

The ability to identify problems and possible causes; gathers relevant information and relates data from different sources to solve problems.

--Followed up on clues provided by Bob, and did not merely follow the interview plan

--Encouraged Bob to express his ideas regarding the source of his depression

--Asked questions about Bob's current hygiene practices (e.g., how often he wears the same clothes; how often he showers)

--Asked questions about Bob's parents' separation (e.g., how do you feel about it?; was it a surprise to you?)

--Gathered information about possible causes of Bob's problems before coming up with a plan for Bob

Planning and Organizing:

The ability to establish a course of action for self and/or others to accomplish a specific goal; makes proper assignments of personnel and allocates resources appropriately.

--Made an initial statement of purpose at the start of the interview

--Ensured that the discussion moved along (i.e., did not get bogged down in one area of concern, but devoted a few minutes to each area)

--Devised a plan for Bob to follow to improve his hygiene and asked for commitment from him to follow the plan

--Devised a plan for Bob to follow to help him deal with his parents' separation and asked for commitment from him to follow the plan

--Set up a follow-up meeting, specifying its time and date, to review Bob's progress with his problems
DIMENSION RATING FORM

Oral Communication

Candidate: ________________  RATING: _____

ASSESSOR #: ____________  CONFIDENCE: _______

Group Discussion:

Counseling Session:

Roommate Conflict:
DIMENSION RATING FORM

Planning and Organizing

Candidate: ________________  RATING: _____
ASSESSOR #: ____________  CONFIDENCE: ______

Group Discussion:

Counseling Session:

Roommate Conflict:
DIMENSION RATING FORM

Analysis and Judgment

Candidate: ____________  RATING: _____
ASSESSOR #: ___________  CONFIDENCE: _____

Group Discussion:

Counseling Session:

Roommate Conflict:
DIMENSION RATING FORM

Sensitivity

Candidate: _______________   RATING: _____

ASSESSOR #: __________   CONFIDENCE: _______

Group Discussion:

Counseling Session:

Roommate Conflict:
OVERALL RATING FORM

BRAD: _____  confidence in this rating: _____

CHUCK: _____  confidence in this rating: _____

GREG: _____  confidence in this rating: _____
ADJECTIVE RATING FORM

CANDIDATE: GREG

ASSESSOR #: ________

DIRECTIONS: On the form below are listed adjective pairs. Your goal is to rate each adjective pair as it applies to the candidate listed above.

EXAMPLE: Suppose you were rating your best friend on the first adjective pair, "at ease-nervous". Read each adjective in the pair and decide which one best describes your friend. Then circle the number that is closest to the adjective that best describes your best friend. Suppose your friend is somewhat more at ease than nervous; you might circle "3", since "3" is closer to "at ease" than to "nervous". Now please be careful and rate the candidate in terms of each adjective pair. Thank you.

<table>
<thead>
<tr>
<th>Adjective Pair</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT EASE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALLOUS</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>CALM</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>CARELESS</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>CONVENTIONAL</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>CRITICAL</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>DISORGANIZED</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>DOWN TO EARTH</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>IRRITABLE</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>LATE</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>LAX</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>LAZY</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>LIKABLE</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>LONELY</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>LONER</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>NARROW INTERESTS</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>NEGLIGENT</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

NERVOUS                     |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
SYMPATHETIC                 |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
WORRYING                    |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
CAREFUL                     |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
ORIGINAL                   |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
LENIENT                     |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
ORGANIZED                  |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
IMAGINATIVE                 |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
GOOD-NATURED               |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
PUNCTUAL                   |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
SCRUPULOUS                 |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
HARDWORKING                |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
UNLIKABLE                  |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
NOT LONELY                 |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
JOINER                     |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
BROAD INTERESTS           |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
CONSCIENTIOUS           |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
<table>
<thead>
<tr>
<th>Trait</th>
<th>CANDIDATE: GREG</th>
<th>ENVIOUS/ JEALOUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOT ENVIOUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td></td>
</tr>
<tr>
<td>PASSIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>ACTIVE</td>
</tr>
<tr>
<td>PATIENT</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>IMPATIENT</td>
</tr>
<tr>
<td>PREFER ROUTINE</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>PREFER VARIETY</td>
</tr>
<tr>
<td>QUIET</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>TALKATIVE</td>
</tr>
<tr>
<td>RELAXED</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>HIGH STRUNG</td>
</tr>
<tr>
<td>RESERVED</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>AFFECTIONATE</td>
</tr>
<tr>
<td>RETIRING</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>SOCIABLE</td>
</tr>
<tr>
<td>SECURE</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>INSECURE</td>
</tr>
<tr>
<td>SELFISH</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>SELFLESS</td>
</tr>
<tr>
<td>SELF-SATISFIED</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>SELF-PITYING</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>COMPLEX</td>
</tr>
<tr>
<td>SLOPPY</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>NEAT</td>
</tr>
<tr>
<td>SOBER</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>FUN LOVING</td>
</tr>
<tr>
<td>STINGY</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>GENEROUS</td>
</tr>
<tr>
<td>SUSPICIOUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>TRUSTING</td>
</tr>
<tr>
<td>RUTHLESS</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>SOFT-HEARTED</td>
</tr>
<tr>
<td>UNADVENTUROUS1</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>DARING</td>
</tr>
<tr>
<td>UNCREATIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>CREATIVE</td>
</tr>
<tr>
<td>UNCURIOSOUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>CURIOUS</td>
</tr>
<tr>
<td>UNEMOTIONAL</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>EMOTIONAL</td>
</tr>
<tr>
<td>UNFEELING</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>PASSIONATE</td>
</tr>
<tr>
<td>VENGEFUL</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>FORGIVING</td>
</tr>
<tr>
<td>WEAK-WILLED</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>SELF-DISCIPLINED</td>
</tr>
</tbody>
</table>
ADJECTIVE RATING FORM

CANDIDATE: CHUCK  ASSESSOR #: ________

DIRECTIONS: On the form below are listed adjective pairs. Your goal is to rate each adjective pair as it applies to the candidate listed above.
EXAMPLE: Suppose you were rating your best friend on the first adjective pair, "at ease-nervous". Read each adjective in the pair and decide which one best describes your friend. Then circle the number that is closest to the adjective that best describes your best friend. Suppose your friend is somewhat more at ease than nervous; you might circle "3", since "3" is closer to "at ease" than to "nervous". Now please be careful and rate the candidate in terms of each adjective pair. Thank you.

AT EASE 1 2 3 4 5 6 7 8 9 NERVOUS
CALLOW 1 2 3 4 5 6 7 8 9 SYMPATHETIC
CALM 1 2 3 4 5 6 7 8 9 WORRYING
CARELESS 1 2 3 4 5 6 7 8 9 CAREFUL
CONVENTIONAL 1 2 3 4 5 6 7 8 9 ORIGINAL
CRITICAL 1 2 3 4 5 6 7 8 9 LENIENT
DISORGANIZED 1 2 3 4 5 6 7 8 9 ORGANIZED
DOWN TO EARTH 1 2 3 4 5 6 7 8 9 IMAGINATIVE
IRRITABLE 1 2 3 4 5 6 7 8 9 GOOD-NATURED
LATE 1 2 3 4 5 6 7 8 9 PUNCTUAL
LAX 1 2 3 4 5 6 7 8 9 SCRUPULOUS
LAZY 1 2 3 4 5 6 7 8 9 HARDWORKING
LIKABLE 1 2 3 4 5 6 7 8 9 UNLIKABLE
LONELY 1 2 3 4 5 6 7 8 9 NOT LONELY
LONER 1 2 3 4 5 6 7 8 9 JOINER
NARROW INTERESTS 1 2 3 4 5 6 7 8 9 WIDE INTERESTS
NEGLIGENT 1 2 3 4 5 6 7 8 9 CONSCIENTIOUS
<table>
<thead>
<tr>
<th>Candidate: Chuck</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOT ENVIous</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>PASSIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>PATIENT</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>PREFER ROUTINE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>QUIET</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>RELAXED</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>RESERVED</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>RETIRING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SECURE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SELFISH</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SELF-SATISFIED</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SLOPPY</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SOBER</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>STINGY</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SUSPICIOUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>RUTHLESS</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>UNADVENTUROUS1</td>
<td>2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>UNCREATIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>UNCURIous</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>UNEMOTIONAL</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>UNFEELING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>VENGEFUL</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>WEAK-WILLED</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>ENVIOUS/JEALOUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>ACTIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>IMPATIENT</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>PREFER VARIETY</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>TALKATIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>HIGH STRUNG</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>AFFECTIONATE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SOCIABLE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>INSECURE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SELFLESS</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SELF-PITYING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>COMPLEX</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>NEAT</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>FUN LOVING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>GENEROUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>TRUSTING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SOFT-HEARTED</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>DARING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>CREATIVE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>CURIOUS</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>EMOTIONAL</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>PASSIONATE</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>FORGIVING</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>SELF-DISCIPLINED</td>
<td>1 2 3 4 5 6 7 8 9</td>
</tr>
</tbody>
</table>
ADJECTIVE RATING FORM

CANDIDATE: BRAD  ASSESSOR #: _______

DIRECTIONS: On the form below are listed adjective pairs. Your goal is to rate each adjective pair as it applies to the candidate listed above.

EXAMPLE: Suppose you were rating your best friend on the first adjective pair, "at ease-nervous". Read each adjective in the pair and decide which one best describes your friend. Then circle the number that is closest to the adjective that best describes your best friend. Suppose your friend is somewhat more at ease than nervous; you might circle "3", since "3" is closer to "at ease" than to "nervous". Now please be careful and rate the candidate in terms of each adjective pair. Thank you.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT EASE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALLOUS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CARELESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONVENTIONAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRITICAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISORGANIZED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOWN TO EARTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRRITABLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAZY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIKABLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LONELY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LONER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NARROW INTERESTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEGLIGENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NERVOUS  SYMPATHETIC  WORRYING  CAREFUL  ORIGINAL  LENIENT  ORGANIZED  IMAGINATIVE  GOOD-NATURED  PUNCTUAL  SCRUPULOUS  HARDWORKING  UNLIKABLE  NOT LONELY  JOINER  WIDE  INTERESTS  CONSCIENTIOUS
<table>
<thead>
<tr>
<th>Word</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOT ENVIOUS</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>PASSIVE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>PATIENT</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>PREFER ROUTINE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>QUIET</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>RELAXED</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>RESERVED</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>RETIRING</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SECURE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SELFISH</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SELF-SATISFIED</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SLOPPY</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SOBER</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>STINGY</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>SUSPICIOUS</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>RUTHLESS</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>UNADVENTUROUS1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>UNCREATIVE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>UNCURIous</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>UNEMOTIONAL</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>UNFEELING</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>VENGEFUL</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>WEAK-WILLED</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

**CANDIDATE: BRAD**

ENVIous/JEALOUS
ACTIVE
IMPATIENT
PREFER VARIETY
TALKATIVE
HIGH STRUNG
AFFECTIONATE
SOCIABLE
INSECURE
SELFLESS
SELF-PITYING
COMPLEX
NEAT
FUN LOVING
GENEROUS
TRUSTING
SOFT-HEARTED
DARING
CREATIVE
CURIOUS
EMOTIONAL
PASSIONATE
FORGIVING
SELF-DISCIPLINED
EXERCISE RATING FORM

ASSESSOR #: __________

On this form, please rate each candidate’s performance from 1 (low performance on the exercise) to 5 (high performance on the exercise). Remember to evaluate each candidate’s exercise performance in terms of the 5-point scale, not in comparison to each other. Please be careful and respond for each candidate. Thank you.

<table>
<thead>
<tr>
<th></th>
<th>GROUP DISCUSSION</th>
<th>COUNSELING SESSION</th>
<th>ROOMMATE CONFLICT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAD:</td>
<td>______</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>CHUCK:</td>
<td>______</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>GREG:</td>
<td>______</td>
<td>______</td>
<td>______</td>
</tr>
</tbody>
</table>
RATING SCALE

1 --- LOW PERFORMANCE ON THE DIMENSION
2 --- FAIRLY LOW PERFORMANCE ON THE DIMENSION
3 --- ACCEPTABLE PERFORMANCE, NOT LOW, NOT HIGH
4 --- FAIRLY HIGH PERFORMANCE ON THE DIMENSION
5 --- HIGH PERFORMANCE ON THE DIMENSION

CONFIDENCE SCALE

0 ------ ABSOLUTELY NOT CONFIDENT THIS RATING IS CORRECT
10
20
30
40
50 ------ 50% CONFIDENT THIS RATING IS CORRECT
60
70
80
90
100 ------ ABSOLUTELY CONFIDENT THIS RATING IS CORRECT
POST-EXPERIMENTAL QUESTIONNAIRE

ASSESSOR #: ________ YEAR IN SCHOOL: 1 2 3 4 G

SEX: MALE FEMALE AGE: ______

Have you ever lived in a residence hall or residential college?______  (If so, for how long? ________________)

What do you think is the purpose of this experiment?

__________________________________________________________

Were you told during this session that the experimenters would examine your classifications and ratings to see how justifiable they were as compared to expert classifications and ratings?

YES NO

How much pressure did you think there was for you to make accurate and thorough classifications and ratings?

NO SLIGHT MODERATE VERY MUCH EXTREME
PRESSURE PRESSURE PRESSURE PRESSURE PRESSURE

How much thought did you put into making your classifications and ratings?

NOT MUCH SLIGHT MODERATE AMOUNT VERY MUCH EXTREME
THOUGHT OF THOUGHT THOUGHT

How important did you think it was for you to make accurate classifications and ratings during this experiment?

NOT AT ALL SLIGHTLY MODERATELY VERY EXTREMELY
IMPORTANT IMPORTANT IMPORTANT

How justifiable would your classifications and ratings be if you had to explain to the experimenters why and how you made the classifications and ratings?

NOT AT ALL SLIGHTLY MODERATELY VERY EXTREMELY
JUSTIFIABLE JUSTIFIABLE JUSTIFIABLE

How well do your classifications and ratings match classifications and ratings made by experts?

NOT AT ALL SLIGHTLY MODERATELY VERY WELL EXTREMELY WELL
MATCHED MATCHED MATCHED MATCHED

How closely do you think the experimenters will study your classifications and ratings?

NOT AT ALL SLIGHTLY MODERATELY VERY EXTREMELY
CAREFULLY CAREFULLY
CONSENT FORM:

RICE UNIVERSITY ASSESSMENT CENTER PROCESS

INTERVIEW FORM

I hereby allow assessment center researchers at Rice University, Wright State University, and Colorado State University to have access to the ratings and descriptions I have provided today as a participant in the "Assessment Center" experiment. I understand that my data will be used to evaluate cognitive decision making processes of assessment center assessors. No data, including my name, will be used for any other purpose. I further understand that questions I may have about my data will be answered by Dr. Barbara Gaugler or Ted Hayes, Department of Psychology, Rice University, x2216.

Signed: ________________________________

Date: __________________________
APPENDIX D

Experimental Session Scripts.

ASSESSMENT CENTER ADMINISTRATION INSTRUCTIONS

(BRACKETED PORTIONS OF THE SCRIPT ARE USED ONLY IN THE ACCOUNTABLE CONDITION)

I. Overview of the Assessment Center

BEFORE YOU START, BE SURE YOU HAVE A COPY OF ALL FORMS, THE TIMETABLE, VIDEOTAPE CHART, AND LIST OF SUBJECTS AND THEIR ROOM ASSIGNMENTS

(a) Close door and put up do not disturb sign

(b) Introduce yourself to the subjects; make sure they are in the correct room. Give each subject an envelope of materials. Tell them NOT TO PUT their names on any of the materials; instead, they will have an assessor number.

(c) Tell the students that:

"We are here today to evaluate decision making processes in an assessment center. This particular assessment center is designed for hiring resident hall student assistants."

(d) "If you'll remember from last time, a residence hall student assistant, known as an SA, is an upper level undergraduate who is assigned to a floor section of a residence hall or college. SAs are responsible for working closely with other SAs to develop and maintain an atmosphere which promotes academic, personal and social growth in the residence hall."
"Now that you're here today, I'm assuming that you'll put forth your best effort. It is important that you work hard for this research project to succeed so that assessment center researchers across the country can have the best data possible to evaluate decision making processes in the assessment center."

"Now that you're here today, I'm assuming that you'll put forth your best effort. It is important that you work hard for this research project to succeed so that universities across the country can have the best system possible for hiring resident hall student assistants).

"The main purpose of this assessment center experiment tonight is to help understand the assessment center judgment process, and your judgment processes in particular. To do that, we're going to evaluate the justifiability of your classifications and ratings. To do this, we'll compare your behavior classifications and responses to those of experts who have rated the same candidates in these tapes. If your individual results seem to be unjustified, you'll be contacted by phone in two days to justify your ratings and behavioral classifications. The researcher will ask you detailed questions about your responses to help understand why you made your classifications and gave the ratings you did. Just to make it official, please sign the additional form acknowledging that you're aware that someone may contact you by phone. Please do this now."
"Although universities have a number of really good SAs, some perform their jobs so poorly that they should never have been hired. I hope that the assessment center that you are going through today will help universities hire the best SAs possible. But, before universities can start using the assessment center to hire SAs, we need your help in evaluating it to make sure that it will work. We appreciate your help by participating in the assessment center today".

(e) "You will be evaluating 3 candidates in one of 3 different situations:

(1) a group discussion of all 3 candidates about a policy change; (2) a roommate conflict; and (3) a counseling session with a student who has a personal problem"

(f) Go over the highlights of the assessment center timetable

"I can't impress upon you enough the importance of working quickly, but thoroughly and accurately and keeping on schedule. Please note that the time allotted for each phase of the assessment center is tight."

II. Recording Observations

(a) "Last time we practiced making behavioral observations. The trainer pointed out several reasons why it is important to make good behavioral observations

Good behavioral observations:

(1) enable you to make accurate evaluations of others;"
and (2) allow you to successfully communicate to the other assessors what the candidate actually said and did;

Although its difficult not to get engrossed while watching the tapes and forget to take notes, without notes, you can't accurately remember what happened - things get distorted.

(b) Go over what constitutes a good behavioral observation

----------------------------------------
TIPS FOR OBSERVING AND RECORDING OBSERVATIONS
----------------------------------------

(c) Pass out their critiqued practice observation forms.

"Most of the feedback focused on encouraging you to record more detail and pay closer attention to nonverbal behavior."

(d) "Do you have any questions about how to make good behavioral observations or about the feedback you got on your practice observation forms?"

(e) "Please note that if a person says or does the same thing several times during an exercise, you don't have to write it down every time. You can just write that the SA said or did a particular thing several times. For example, you can write 'He nodded his head several times while listening to the students.'"
(f) "Again, remember to write down details; write down verbatim quotes if possible, and note the context of things the SA said so you will later be able to interpret what the person said."

III. Dimension Training

(a) "Now let's talk about the dimensions on which you will be evaluating the assessees. After each exercise, you will classify behaviors under dimensions."

-----------------------------------------
DIMENSION LIST
-----------------------------------------

(b) "Please carefully read each dimension and its definition. Do you have any questions? It's very important that you look for both verbal and nonverbal behaviors that represent each dimension."

GO OVER THE FIRST EXERCISE, THE GROUP DISCUSSION

The student assistants have gotten together to come up with recommendations on how to best enforce the alcohol policy in the residence halls.

TELL THEM WHAT FORM TO USE AND WHAT PERSON TO WATCH.
TELL THEM TO BE SURE TO WRITE THE CANDIDATE'S NAME AND TODAY'S DATE ON EACH FORM THEY FILL OUT. ASK THEM NOT TO DISCUSS THE TAPES WITH EACH OTHER. TELL THEM THAT IT IS IMPORTANT THAT THEIR OBSERVATIONS AND RATINGS ARE NOT INFLUENCED BY OTHERS.

[RE Mind them that they need to make justifiable
OBSERVATIONS AND CLASSIFICATIONS.]
START THE FIRST TAPE.
AS SOON AS THE GROUP DISCUSSION IS OVER, STOP THE TAPE.

IV. Classification Training

(a) "Now let's talk about how to classify behaviors under dimensions. After watching each exercise and recording the behaviors observed, you will classify those behaviors under the appropriate dimension using the Dimension Rating Forms. There is one form for each dimension."

-------------------------
DIM RATING FORMS
-------------------------

"You will use 4 Dimension Rating Forms for each candidate, one for each of the four dimensions. Each dimension is on a different color sheet of paper. Since you have just watched the group discussion exercise, classify the behaviors which represent Oral Communication on the pink Dimension Rating Form under the group discussion. Then classify the behaviors which represent Sensitivity on the blue Dimension Rating Form under the group discussion. You will then classify behaviors under Analysis and Judgment and Planning and Organizing. During the integration session in which you share your observations with the other assessors, you will take notes on the candidates' performance on dimensions under the other two exercises."

HOLD UP FORMS AND POINT TO SPACE UNDERNEATH THE GROUP
DISCUSSION

"You may classify a particular behavior under more than one dimension if you feel that behavior is indicative of more than one dimension."

"Also, some behaviors may not be classifiable under the dimensions. If you feel that a behavior doesn't fit any of the dimensions, then don't classify it and don't report it to the other assessors.

"You should give a (+) to the behaviors that are positive for the dimension and a (-) for behaviors that are negative for the dimension"

-------------------

MINIMAL STANDARDS: LGD

-------------------

(b) "To help classify behaviors under the proper dimension, refer to the minimal standards of performance for each exercise. Examples of behaviors that represent each dimension are given."

"If you notice a behavior on the minimal standards that the candidate exhibited, but you forgot to record, go ahead and classify that behavior on the appropriate dimension rating form."

"Do you have any questions? You will have 15 minutes to classify behaviors. Remember, you need to work quickly but accurately."

"Go ahead and classify the behaviors you just observed
on the videotape, using the minimal standards for the group discussion as a guide. [Remember that you need to make your classifications as justifiable as possible.]

GIVE THEM A 2 MINUTE WARNING. WHEN ALL ASSESSEES ARE FINISHED OR AFTER 15 MINUTES AT THE MOST, GO OVER THE COUNSELING SESSION EXERCISE.

Bob, a student at the university, has been neglecting his hygiene and coursework because he is upset that his parents are getting divorced. The SA is to discuss the problem with Bob, help him deal with his parent's situation and get him to improve his personal hygiene.

TELL THEM WHICH SA THEY WILL WATCH

HAVE THEM OBSERVE AND CLASSIFY BEHAVIORS FOR THE SECOND AND THIRD EXERCISE. REMIND THEM THAT THEY CAN USE THE MINIMAL STANDARDS TO HELP CLASSIFY BEHAVIORS UNDER DIMENSIONS. MAKE SURE THEY USE A NEW SET OF DIMENSION RATING FORMS. GIVE THEM A 2 MINUTE WARNING WHEN CLASSIFYING BEHAVIORS. [REMIND THEM THAT THEY NEED TO MAKE JUSTIFIABLE CLASSIFICATIONS.]

BEFORE STARTING THE THIRD TAPE, GO OVER THE ROOMMATE CONFLICT EXERCISE.

Steve and John are roommates, not by choice, and have not been able to adjust their lifestyles to get along. They fight constantly. There are no available rooms so they must continue to live together for the rest of the semester. The SA is to listen to both sides of the story and help them
arrive at a compromise so that they can start getting along better.

TELL THEM WHICH SA THEY WILL WATCH. [REMEMIND THEM TO MAKE JUSTIFIABLE CLASSIFICATIONS.]

TURN OFF THE TV AND VIDEO PLAYER WHEN THE LAST EXERCISE IS OVER.

AFTER THEY HAVE CLASSIFIED BEHAVIORS FOR ALL 3 EXERCISES, GIVE THEM THEIR ASSESSOR GROUP ROOM ASSIGNMENTS. TELL THEM WHO THE ADMINISTRATOR IS IN THAT ROOM. TELL THEM THAT THEY HAVE 5 MINUTES TO GET TO THEIR ROOM AND RELAX. TELL THEM WHAT TIME TO BE AT THEIR ROOM. TELL THEM THAT THE NEXT PHASE OF THE ASSESSMENT CENTER, THE INTEGRATION SESSION, WILL START PROMPTLY AT _____!

TELL THEM NOT TO GO INTO THE ROOM UNTIL THE DOOR IS OPEN

----------------------5 MINUTE BREAK----------------------

MAKE SURE THAT YOU HAVE THE CORRECT PEOPLE. START THE INTEGRATION SESSION ON TIME. IF EVERYONE GETS TO YOUR ROOM AHEAD OF TIME AND WANTS TO START EARLY, THAT'S OK.

[Before you start the integration session, get a time estimate from each assessor of when would be the most convenient time for the researcher to call and have a brief discussion of their results if they seem unjustifiable.]

V. Rater Training

(A) Go over the rating scale (put the scale on the blackboard ahead of time):
"1" - low performance on the dimension
"2" - fairly low performance
"3" - acceptable performance, not low, not high
"4" - fairly high performance
"5" - high performance on the dimension

"Please note that a "3" denotes acceptable performance, not average performance. If you give someone a "3", you're saying that the candidate has performed at an acceptable level in that exercise"

(b) Have the subjects separate their minimal standards from the rest of their forms. Go over the minimal standards of performance for each exercise:

"Use the minimal standards of performance to help you evaluate each candidate's performance on the dimensions. The behaviors listed under each dimension are indicative of performance on that dimension. If a candidate exhibits most of the behaviors listed under a particular dimension across all exercises, he is "acceptable" on that dimension and receives a rating of at least a "3". However, if a candidate does not meet these requirements he has performed "below standard" on that dimension and therefore should usually be rated a "1" or a "2" on that dimension."

"In some instances, however, a candidate does not meet the minimal standards, but may still earn a "3" rating. If a candidate exhibits half or more behaviors indicative of acceptable performance on the dimension exceptionally well,
assessors may use their own judgment to decide whether that compensates for not exhibiting a "minimal" behavior."

"If a candidate exhibits additional behaviors indicative of good performance on the dimension, he is "above standard" on that dimension and should receive a rating of a "4" or "5". "

VI. Overview of Integration Session

(c) Tell the assessors:

"Now you will share observations and rate the candidates on the dimensions, one candidate at a time. You will first share behavioral observations and rate (the first candidate's) performance on oral communication, taking into account his performance in all three exercises. Also independently, you will give an estimate of how confident you are (on a 0-100% scale) that your rating is correct; you'll see the confidence scale on the dimension rating sheet. A rating of 0 for confidence indicates absolutely no confidence in your rating, while a 100 means you're absolutely certain. You will repeat this process for each dimension. Then you will independently rate his overall performance on the Overall Assessment Form and indicate your confidence in that rating."

["Remember that your ratings will be evaluated to determine the justifiability of the ratings to help determine your decision processes in the assessment center."]
OVERALL RATING FORM

"Finally, you will share behaviors and rate (the second candidate), and then (the final candidate)."

VII. Integration session ground rules:

"Here are the ground rules for the integration session. You are to listen to the assessor's report and take notes on what he/she says. You are not to discuss what the assessor reports, even if you disagree with what the assessor said. You are only allowed to ask questions to clarify what the candidate did or said."

"If the reporting assessor reports a behavior which you feel is not indicative of the dimension under discussion, you are free to classify the behavior under the dimension or dimensions that you think are appropriate."

"You will have 3 - 4 minutes to read your observations and then everyone has 2 minutes to make their ratings. You do not have to write down everything the reporting assessor says. However you do need to take enough notes so that you can make your ratings. [Remember, it is very important to make justifiable ratings.]"

(b) Make sure that the members of the assessor team face one another.
Steps:

(1) Share observations on (the first candidate's) performance on oral communication in all exercises

(2) Independently rate (the first candidate's) oral communication, and then make a confidence rating

(3) Repeat steps 1 - 2 for each dimension

(4) Rate his overall performance in the assessment center and make a confidence rating.

Repeat this process for each candidate in turn.

After the assessors have shared their observations of the first candidate on oral communication but before they make their ratings, remind them to use the minimal standards to make their ratings.

VII. Adjective and exercise ratings

(A) After all candidates have been rated and have received their overall ratings, hand out the adjective inventory and say:

"Now we'd like you to assess each candidate on each one of these adjective pairs. They are in alphabetical order. Just indicate by circling the number for each pair which adjective in the pair best describes the candidate. You don't have to spend too much time thinking about each choice, but please be accurate. Start with the first candidate and rate him on all adjective pairs. Then start a new form and evaluate the second candidate, and once you're through with his form, evaluate the third candidate on the
final adjective set. Remember, you're using a new set of forms for each candidate. [Keep in mind that your responses are going to be evaluated in comparison with expert ratings."

(B) After they are through with the adjective ratings, pass out the exercise ratings and say:

"Now please rate each candidate's performance in each exercise. Make a rating on a 1 to 5 scale, with 1 being low performance in the exercise and 5 being high performance in the exercise. There are no minimal standards, so just make your best estimate. Please evaluate the exercise performance of the first candidate start with the first candidate and then move on to the other two. I know this has been a long experiment, but it's almost over!"

(C) "Great. There's a final form to fill out. It asks you for some of your reactions to this experiment. It will only take a minute or two to complete."

VIII. Closing

Thank the assessors for participating in the experiment. TELL THEM THAT YOU REALLY APPRECIATE THEIR INPUT. TELL THEM THAT IT SEEMS LIKE THEY ALL DID A GOOD JOB.

Collect all materials.

Tell them that they will receive a memo, debriefing them about the experiment after all subjects have been run (near the end of the semester).
Sign their "credit cards". The study is called "Assessment Centers" (#72) and they should receive 4 credits. Again, thank them for their participation.

Take the "do not disturb" sign off the door. Return materials to my office that night.

-----------------------------REMINDERS-----------------------------

BE ENTHUSIATIC!!!!!!!!!!!!!!!!!!!

REMEMBER: This is a long experiment. Your assessors will become tired and may gripe. Don't take it personally. Just remind them in a friendly manner that you're all in it together, and that if they can cooperate a bit longer they will get through it faster. If there are cookies and soda available, you may want to distribute these a bit early to appease them. However, don't forget that you're in control of the session, and they agreed to participate.

BE CAREFUL NOT TO AD LIB THE INSTRUCTIONS BUT DON'T READ CONSTANTLY

Impress upon the subjects the need to do well and to feel free to ask questions. Ask for questions several times during the assessment center.

Be friendly but task-oriented.

Make sure your packets have the correct forms. Bring extra materials with you.

Make sure the subject's ID # is on each form.

Make sure that the subjects are using the correct forms
at all times.

Do not allow the subjects to take any materials from the room.

If an assessor wants to know if a behavior they recorded is a good behavioral observation, refer them to their copy of "Tips for observing and recording behaviors".

Neither you nor any of the other assessors are to help subjects classify behaviors into dimensions. However, you may clarify and help assessors understand the dimension definitions and the minimal standards. Make sure that the assessors change tasks on time!!!! TIME EACH SEGMENT. The schedule is very tight. If you fall behind, you will probably not be able to catch up later on.

Do not sign their "credit cards" until after the assessment center is over.

If an assessor reports a behavior that didn't occur, don't say anything about it

Dress professionally (e.g., no jeans, shorts or sweats)

I have double checked the rooms. If someone hassles you, tell them that you have reserved the room through the appropriate university channels and they are to contact me if they have any problems. Do not give up the room. Write down the person's name and telephone number and tell him/her that I will be in touch. Give them my name and number if the instructor asks for it (Ted Hayes, Psychology Dept., ext. 2216)
If the tape runs into the next exercise, quickly shut off the monitor. When you start the next exercise, rewind the tape to the beginning of the exercise you want to show.

Hold up each form as they are to use it. Continually check to make sure that subjects are using the correct forms.

You can read while they watch the tapes and classify behaviors but don't make noise.

If only 1 or 2 people show up to watch your tapes, go to the other rooms and take one of their overflow subjects (fourth person on their list). Put that person in your vacant slot. If only 1 or 2 people come to your room for the integration session, you can't run it. Have them become observers in the other rooms. Take them to their new room and explain the situation to the administrator.
APPENDIX E

Observations and Classifications Made by Experts.

BEHAVIORAL OBSERVATIONS

BRAD: GROUP DISCUSSION

--(Looks at others as they talk. Twirls pencil, nods head, says "Umm")
--He's always wanted to make recommendations to the administration
--Look at alcohol policies at other Universities
--Setting drinking guidelines would be redundant because the law says you have to be over 21 to drink already. The SAs need to work with the students to control the problems better, not simply enforce the law.
--Banning alcohol is not the best way to enforce the policy. Students will still buy it and smuggle it into their rooms; some die as a result.
--The present alcohol policy is fair (letting those over 21 drink in the halls)
--Students are going to encounter the drinking law off-campus and they've got to learn to deal with it. Why should college be different?
--Let's keep the current system, because in a controlled space it's easier to keep watch on those who drink
--Education is a major part of what we do (agrees with Chuck--"right")
--SAs need to work with each other to establish consistency
in policies between floors.

**BRAD: COUNSELING SESSION**

--Introduces himself to Bob, asks how school is going
--He's heard Bob's parents were separating, and that Bob
seems depressed
--Other people's parents have separated before. It's an
unfortunate fact of life. Bob will live through it by
concentrating on the 'now'
--Asks whether he's talked with any friends about the
separation
--'Umm'
--He know's it's tough. He can't really understand since
his parents haven't separated.
--Bob did really well in school last semester
--Bob's parents wouldn't want their split to affect him,
because they know he cares for them but they care for him
too. They want him to take care of the little things like
school and health and go from there
--Other people have noticed Bob's deteriorating condition
and have brought it to Brad's attention. Maybe they can
help Bob out
--Have you talked with your parents about the separation?
How long has it been going on?
--Talking to someone who's been through a similar situation
would help
--Talking to his brother would help; he'd make time for it
—He'll live through it by taking it one day at a time
(health and school)
—What's on your list of priorities? Bob needs to
prioritize, get motivation, and start like his brother
—Did you see the game last weekend?
—Brad knows it's tough but he's got to keep himself up and
stay on top of things. When others start to notice the
problem it's bad
—Keep the problem on the back burner to see if it will
resolve itself
—You've got your own life to lead at school that is
different from that of your parents
—If I can help, let me know
—I'd like to see you get a little more motivation

BRAD: ROOMMATE CONFLICT

—Introduces himself to the roommates (shakes hands)
—I've heard you've been having problems getting along
—I've looked all over and can't find a spare room because
of overcrowding
—I've had roommates before, and no one's perfect ("just be
glad you don't have a kitchen")
—Steve, living together shouldn't affect your studying
—Everyone has different study habits; you'll need to work
something out
—Talk ahead of time and develop a plan for studying and
partying
--Suggests alternatives to the roommates for the weekend
--'Umm'
--Leans forward
--You're making progress and need to learn to live together
and learn from each other

CHUCK: GROUP DISCUSSION

--Tells others that they are there to review the
UNiversity's alcohol policy and make recommendations about
policy and enforcement (reviews current policy)
--Asks each member to give input (reviews notes, listens
to/looks at others, 'OK')
--Summarizes each ('Don't want the students to think we're
against them'; 'It's hard to play both roles of friends and
enforcers'), makes notes
--Is the current system fair to those under 21? Won't they
feel left out?
--Yes, the campus situation has to be more like the real
world in terms of drinking laws
--How does everyone feel about the current system (such as
designated drinking areas)?
--How do the others feel about putting on alcohol awareness
presentations to make students aware of alcohol dangers?
--Agrees with Brad that we need a standard system and
policies from SA to SA
--Summarizes the discussion's main points; keep track of
time
--I like each of your ideas. Thanks for your input

CHUCK: COUNSELING SESSION

--Introduces himself, asks how Bob is doing, tells him he seems bummed out. I'm here to see if I could help; tell me why you're depressed ('Wow', leans forward, etc.)

--You seem pretty unmotivated

--Others are concerned about you; he needs to get back into normal patterns to get his mind off things. Your parents wouldn't want to see you this way

--It hasn't happened to me and I can't understand exactly, but I'm concerned anyway

--Getting back into habits will help (changing clothes, going to school, etc.)

--The separation is 'a slap in the face'--shocking

--Do you have any close friends you could talk to about this? Do you belong to a study group?

--Tries to get Bob to agree to take a shower nightly

--Tries to get Bob to agree to study again

--Why don't we get some people together for notes and studying

--Does your brother live on campus? Can you talk with him about it to rid each other of some tension? Maybe they can talk to parents together

--I'll get physics notes from a lady down the hall

--I don't understand exactly but I'm here to talk when you need to
--Arranges follow-up. Thanks Bob

CHUCK: ROOMMATE CONFLICT

--Introduces himself. I've heard you've been having some problems; asks what's going on

--Listens to both roommates

--The university is full, there's no chance of getting a room now, but I'll keep trying. Can't you get along in the meantime?

--It's no use to scream at each other; you need to compromise

--What's the problem

--Summarizes each roommate's concerns

--You'll have to give each other equal time in the room, so you need to work out schedules for the week, weekends, and exams

--Suggests partying/studying location alternatives

--Asks whether they can get together Monday, watch the game together, and come up with a schedule at half-time

--Who's your favorite team?

--Give the schedule 2 weeks; sets up a follow-up meeting to discuss progress. Thanks each roommate

GREG: GROUP DISCUSSION

--What are other universities' policies on alcohol that work

--Listens to Chuck, plays with pencil, leans forward, rubs hands, stutters

--The solutions are causing another problem: SAs are
supposed to be students' friends and help them, but also enforce rules. They need to draw the line somewhere by making students aware of rules.

---Can call police or issue a series of warnings. Knowing who was over 21 would help with enforcement

---It's fair to those over 21 to let them drink as long as they know the alcohol policy

---A bar's a bar, on or off campus, you still need to be over 21 to drink. Need to support our kids

---The law may not be fair but we didn't make the rules; we were only asked to come up with creative ways to enforce rules. Students will drink anyway, but we can't watch over them 'hand and foot'

---Starts to talk, loses train of thought ('what was I going to say?')

---Maybe programs can't hurt. I doubt they will help since most are aware of alcohol effects before college, since students are semi-intelligent

GREG: COUNSELING SESSION

---Introduces himself to Bob. Asks how he's feeling, why he hasn't been to class. Do you want to talk about it? (leans forward, rubs hands)

---'We all get that way sometimes'. I may not understand, but try me

---I can imagine it's pretty hard on you. Does it look permanent?
What do you plan to do about it? (leave school?)

--Do you want to talk about it? Even though I can't understand I can sympathize

--Do you have any brothers or sisters you can talk to about the split who would understand the pressure?

--Your brother wouldn't be too busy to talk, he's reacting to pressure too

--Have you been eating right and changing clothes?

--Getting yourself 'back in line' would help you start to deal with the problem

--Others on the floor care about you and are concerned. Talking to them would make you feel better and that would help him feel better about himself

--Break the problem down into parts and working on one part at a time

--Have you tried to talk to your parents?

--Tries to get him to agree to talk with his brother

--Come back to talk again in a few days

GREG: ROOMMATE CONFLICT

--Introduces himself to the roommates; I understand you've been having problems

--The university is overcrowded and you can't get another room (looks at both, sits back, rubs hands) even though there are tons of rooms

--If you can work it out without my help that's fine. Otherwise you have to work out a compromise and I'm here to
help. What problems do you have?
--I can empathize with both of you, you each have a legitimate gripe. I want to work a compromise
--The university is made for studying but not every night
--Why don't you check with each other at least one day in advance before studying/partying
--The 'key' is common courtesy to each other
--They need to check with each other ('Do you make reservations at home?')
--I'm not trying to make you (John) out to be anything
--Do you have any common interests? ('Yelling can be fun too')
--Makes suggestions for alternatives to each roommate
--I expect that each of you will respect the other's rights;
I'll return if respect doesn't work, or come by if you need to
--I'll get back with you in a couple of days; compromise won't be difficult.
BEHAVIORAL CLASSIFICATIONS

CHUCK: PLANNING AND ORGANIZING

GROUP DISCUSSION:
--Stated the purpose of the discussion
--Summarized what each member had said before going to next topic
--Asked different members for their opinions
--Ensured the discussion moved along
--Made suggestions about what to do with the proposed solutions
--Took notes on the discussion, kept track of time
--Summarized the discussion at the end

COUNSELING SESSION:
--Made a statement of why he was there at the beginning
--Brought up new topics to keep the discussion going
(studying, talking with brother/others, etc.)
--Asked Bob for commitment to a plan to improve hygiene and start studying again
--Tried to develop a plan to deal with the separation
--Arranged a follow-up meeting in 2 days to review Bob's progress

ROOMMATE CONFLICT:
--Asked what was going on
--Brought up new topics to keep discussion moving and to identify problems
--Asked for commitment to make a room schedule and to try
and get along better (watch football together).
--Arranged follow-up meeting in 2 weeks

CHUCK: ORAL COMMUNICATION

GROUP DISCUSSION:
--Spoke clearly, concisely, and fluently
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Took notes, looked at papers
--Said "OK"

COUNSELING SESSION:
--Spoke clearly, concisely, and fluently
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Used hand gestures to emphasize points
--Said "OK"

ROOMMATE CONFLICT:
--Spoke clearly, concisely, and fluently
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Said "OK"
CHUCK: SENSITIVITY

GROUP DISCUSSION:
--Didn't interrupt others
--Asked if some would feel left out by the alcohol policy
--Questioned others in a supportive manner
--Listened attentively

COUNSELING SESSION:
--Listened attentively (leaned forward, said "Oh wow")
--Expressed empathy with Bob ("It's a slap in the face")
--Asked Bob if he would consider support from others
--Didn't interrupt Bob
--Made positive statements about Bob ("We're really concerned about your well-being")
--Persuaded Bob to give his suggestions a chance; wasn't coercive

ROOMMATE CONFLICT:
--Listened attentively to each roommate
--Didn't interrupt the roommates
--Didn't take sides between the roommates
--Expressed empathy with both roommates
--Persuaded the roommates to try and change the situation (wasn't coercive)

CHUCK: ANALYSIS AND JUDGMENT

GROUP DISCUSSION:
--Asked for input from others
--Reviewed current policy, agreed they needed a standard
policy
--Examined possible causes of the alcohol problem in the
dorms
--Examined +/- of proposed solutions ("Would it be unfair?")
COUNSELING SESSION:
--Took clues from Bob ("Talking with your brother would be
good")
--Asked about the sources of Bob's problems
--Didn't ask about Bob's hygiene, but got him to talk about it
--Didn't ask about his parents' separation
--Asked about plans for change (suggested study groups,
talking with friends)
--Followed interview plan
ROOMMATE CONFLICT:
--Asked about the problem
--Took clues from each roommate to gather more information
--Encouraged both to express ideas and explore solutions
--Sought common ground between roommates to resolve the
problem

BRAD: ORAL COMMUNICATION

GROUP DISCUSSION:
--Spoke clearly, concisely; slight hesitation
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Played with pencil, looked at papers
--Said "Umm", "Hmm"

COUNSELING SESSION:
--Spoke clearly, concisely, and fluently
--Voice trailed off when making statements
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Used hand gestures to emphasize points
--Said "Umm", "Hmm"
--Laughed after asking Bob if he had seen the game

ROOMMATE CONFLICT:
--Spoke clearly, concisely, and fluently
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Said "Umm", "Hmm"

BRAD: PLANNING AND ORGANIZING

GROUP DISCUSSION:
--No initial statement of purpose
--Didn't take notes, monitor time, etc.
--No ideas about what to do with the group's suggestions
--Didn't summarize the group's suggestions
COUNSELING SESSION:
--Made a statement of purpose (he heard Bob's parents had separated)
--Introduced new topics to move discussion along
--Suggested talking with his brother to help Bob deal with the separation
--Told Bob to prioritize
--Didn't talk about hygiene
-- Didn't set follow-up meeting time
ROOMMATE CONFLICT:
--Made a statement of purpose (he heard they were not getting along)
--Talked about a number of topics to move discussion along
--Told them to develop a plan to split time in the room
--Didn't set a follow-up meeting

BRAD: SENSITIVITY

GROUP DISCUSSION:
--Agreed with others (didn't question them)
--Didn't interrupt others
--Said they needed a standard policy to reduce problems between SAs and students
--Wondered about the reaction of students to the policy
COUNSELING SESSION:
--Asked Bob if he was depressed
--Listened attentively (eye contact, leaned forward)
--Said he couldn't relate to Bob's feelings ("my parents..."
haven't separated")

--Asked if Bob had seen the football game
--Made a positive statement about Bob ("you did well last semester")
--Tried to persuade Bob to get through his current problems, but wasn't coercive ("take it one day at a time")

ROOMMATE CONFLICT:
--Listened attentively
--Didn't take sides
--Didn't interrupt roommates
--Didn't express an understanding of wither roommate, but indicated he understood how the conflict could arise ("no one's perfect")
--Told them they could learn from each other
--Persuaded them to talk and compromise (e.g., set up a room schedule)

BRAD: ANALYSIS AND JUDGMENT

GROUP DISCUSSION:
--Didn't encourage others to express ideas (responded to Chuck's questions)
--Didn't gather information about possible causes of the alcohol problem
--Wanted to look at policies at other Universities to see what works
--Indicated that the cause of the problem was not their concern ("not our job to enforce the law")
--Examined +/- of proposed solutions ("Alcohol bans don't work")

--Didn't incorporate ideas of others into solutions

COUNSELING SESSION:

--Took clues from Bob about priorities and talking to brother

--Encouraged Bob to express his ideas about the source of his depression ("I heard your parents have separated")

--Didn't ask about Bob's hygiene

--Asked about his parents' separation ("Have you talked w/ them?")

--Didn't ask questions to develop a plan for Bob to follow

ROOMMATE CONFLICT:

--Used information from both roommates to solve the problem

--Followed up clues about the problems (Cathy, partying)

--Asked if living together would affect their study habits

--Didn't encourage them to express their ideas about the conflict

--Didn't gather further information about the problems

--Didn't seek a common ground

GREG: PLANNING AND ORGANIZING

GROUP DISCUSSION:

--Didn't make a statement of purpose

--Said they were only asked to suggest creative policies, not make law

--Didn't make suggestions about what to do with ideas
COUNSELING SESSION:
--Didn't make a statement of purpose
--Asked questions to determine the problem
--Ensured the discussion moved along
--Asked for commitment from Bob to improve hygiene and talk with his brother
--Didn't set follow-up meeting, but suggested Bob contact him in a couple of days

ROOMMATE CONFLICT:
--Stated that he heard they were having problems getting along
--Told them they had to live together and needed to compromise
--Ensured that the discussion moved along
--Persuaded them to show more courtesy for each other
--Told them he'd get back with them in a couple of days

GREG: ANALYSIS AND JUDGMENT

GROUP DISCUSSION:
--Asked about the current alcohol policy
--Said the problem was being a friend but also enforcing the law
--Asked Brad what other Universities do and whether it worked; didn't gather other information
--Suggested a series of warnings, and finding out who was over 21
--Examined +/- of proposed solutions
COUNSELING SESSION:
--Followed Bob's clues and asked about the separation
--Asked what Bob planned to do (stay in school, talk with brother)
--Asked whether Bob was taking care of himself (changing clothes, going to classes)
--Asked if anything else was wrong
ROOMMATE CONFLICT:
--Used information from each to ID the problem (asked each roommate)
--Looked for common ground between the roommates

GREG: SENSITIVITY

GROUP DISCUSSION:
--Said the alcohol awareness program might work but students have already been exposed to alcohol before college
--Expressed concern that the SA had to be both friend and monitor
--Said the law may be unfair but that wasn't their concern
--Listened attentively
--Didn't interrupt others

COUNSELING SESSION:
--Listened attentively
--Didn't interrupt Bob
--Expressed empathy with Bob's situation ("I can imagine...")
--Made positive statement about Bob (others were concerned
about him)

**ROOMMATE CONFLICT:**

--Listened attentively

--Didn't interrupt the roommates

--Didn't take sides

--Expressed an understanding of each roommate's position

--Didn't make a positive statement about them

--Suggested they respect each other

**GREG: ORAL COMMUNICATION**

**GROUP DISCUSSION:**

--Spoke clearly, concisely; occasional hesitation

--Didn't have to repeat self to be understood

--Used acceptable grammar

--Spoke in a conversational tone and volume

--Lost train of thought ("What was I going to say?")

--Jokes

--Said "Umm", "Hmm"

--Played with (rubs) hands

--Looked down

**COUNSELING SESSION:**

--Spoke clearly, concisely, but slowly

--Didn't have to repeat self to be understood

--Used acceptable grammar

--Spoke in a conversational tone and volume

--Maintained train of thought

--Used hand gestures to emphasize points
--Said "Umm", "Hmm"
--Leaned forward

ROOMMATE CONFLICT:
--Spoke clearly, concisely, and fluently
--Didn't have to repeat self to be understood
--Used acceptable grammar
--Spoke in a conversational tone and volume
--Maintained train of thought
--Said "Umm", "Hmm"
--Moved around in chair
--Looked down, leaned forward
APPENDIX F

Studies included in the accountability literature review for effect size and power analyses.


