INFORMATION TO USERS

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

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

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

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

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
The influence of dimension concreteness on assessors’ judgments

Parker, Debra K., M.A.
Rice University, 1991
RICE UNIVERSITY

THE INFLUENCE OF DIMENSION CONCRETENESS ON ASSESSORS’ JUDGMENTS

by

DEBRA K. PARKER

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

APPROVED, THESIS COMMITTEE

Barbara Gaugler, Assistant Professor of Psychology, Chair

William Howell, Professor of Psychology

Nancy Cooke, Assistant Professor of Psychology

Houston, Texas
February, 1991
The Influence of Concrete Dimensions on Assessors' Judgments

Debra Parker

Abstract

Assessment center dimensions have often been found to be low in convergent and discriminant validity (Hinrichs & Haanpera, 1976; Sackett & Dreher, 1982; Sackett & Hakel, 1979; Turnage & Muchinsky, 1982). Assessors' use of prototypes may interfere with assessment center ratings. Reliance on prototypes may be especially pronounced when dimensions are abstract. In this study, the influence of concrete dimensions on assessors observations, classifications, rating accuracy, and convergent and discriminant validity was investigated in an assessment center simulation. Sixty-six university students were trained as assessors. Using either concrete or abstract dimensions, they then evaluated the performance of confederates in three situational exercises. Subjects who rated concrete dimensions classified behaviors more accurately, rated dimensions more accurately according to two accuracy measures, and produced somewhat better convergent and discriminant validity than did subjects who rated abstract dimensions. Subjects who rated abstract dimensions had more accurate ratings according to one accuracy measure than did subjects who rated concrete dimensions.
DEDICATION

To the memory of my father, and to my mother, who taught me
"You can."
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Overview</td>
<td>1</td>
</tr>
<tr>
<td>The Roles of an Assessor</td>
<td>3</td>
</tr>
<tr>
<td>Reliability and Validity of Dimensions</td>
<td>5</td>
</tr>
<tr>
<td>Explanations for Lack of Convergent and Discriminant Validity</td>
<td>7</td>
</tr>
<tr>
<td>Abstractness of Dimensions: Cognitive and Social Cognition Research</td>
<td>9</td>
</tr>
<tr>
<td>Abstract Dimensions: Assessment Center Research</td>
<td>12</td>
</tr>
<tr>
<td>The Present Investigation</td>
<td>14</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>15</td>
</tr>
<tr>
<td>METHOD</td>
<td>15</td>
</tr>
<tr>
<td>Pilot Work and Creation of Dimension Sets</td>
<td>15</td>
</tr>
<tr>
<td>Design</td>
<td>19</td>
</tr>
<tr>
<td>Subjects and Procedure</td>
<td>19</td>
</tr>
<tr>
<td>Assessor Training</td>
<td>21</td>
</tr>
<tr>
<td>Minimum Standards</td>
<td>21</td>
</tr>
<tr>
<td>Videotape Exercises</td>
<td>21</td>
</tr>
<tr>
<td>Dependent Measures</td>
<td>22</td>
</tr>
<tr>
<td>RESULTS</td>
<td>27</td>
</tr>
<tr>
<td>Observation Accuracy</td>
<td>27</td>
</tr>
<tr>
<td>Classification Accuracy</td>
<td>28</td>
</tr>
<tr>
<td>Rating Accuracy</td>
<td>29</td>
</tr>
<tr>
<td>Inter-Rater Reliability</td>
<td>30</td>
</tr>
<tr>
<td>Convergent and Discriminant Validity: The Multitrait-Multimethod Matrix</td>
<td>32</td>
</tr>
<tr>
<td>Convergent and Discriminant Validity: ANOVA Analysis</td>
<td>33</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>36</td>
</tr>
<tr>
<td>Observation Accuracy</td>
<td>37</td>
</tr>
<tr>
<td>Classification Accuracy</td>
<td>37</td>
</tr>
<tr>
<td>Rating Accuracy</td>
<td>38</td>
</tr>
<tr>
<td>Inter-Rater Reliability</td>
<td>40</td>
</tr>
<tr>
<td>Convergent and Discriminant Validity</td>
<td>40</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Means and Standard Deviations of Observation Measures</td>
<td>28</td>
</tr>
<tr>
<td>2. Means and Standard Deviations of Classification Measures</td>
<td>29</td>
</tr>
<tr>
<td>3. Means and Standard Deviations of Rating Accuracy Measures</td>
<td>30</td>
</tr>
<tr>
<td>4. Analysis of Variance for Assessor Ratings, Abstract Group</td>
<td>31</td>
</tr>
<tr>
<td>5. Analysis of Variance for Assessor Ratings, Concrete Group</td>
<td>31</td>
</tr>
<tr>
<td>6. Mean Dimension and Exercise Correlations</td>
<td>33</td>
</tr>
<tr>
<td>7. Analysis of Variation of Concrete Dimension Correlations</td>
<td>34</td>
</tr>
<tr>
<td>8. Analysis of Variance for Abstract Dimension Correlations</td>
<td>34</td>
</tr>
<tr>
<td>9. Variance Components</td>
<td>35</td>
</tr>
<tr>
<td>10. Intraclass Correlations</td>
<td>36</td>
</tr>
</tbody>
</table>
Introduction

Overview

The assessment center has become a very popular method of selecting and promoting personnel. Judgments resulting from the method have frequently been found to be highly predictive of subsequent job performance and other criteria. Gaugler, Rosenthal, Thornton, & Bentsen (1987) performed a meta-analysis wherein the success of the method in various organizations and in a number of positions can readily be seen. Despite the successful utilization of the method, its underlying processes are undetermined. Several theories concerning the underlying cognitive and behavioral processes have been advanced (Klimoski & Brickner, 1987). Many of these theories center around the use of dimensions, which are used to make judgements in an assessment center.

Dimensions, which have have been defined as "clusters of behaviors that are specific, observable, and verifiable, and that can be reliably and logically classified together" (Thornton & Byham, 1982, p. 117), have traditionally been an integral part of the assessment center method. In fact, Zedeck and Cascio (1984) state that the underlying operational notion of an assessment center is that they are "dimension driven" (p. 428). Recent research on the assessment center method suggests that assessors utilize dimensions in a unexpected manner. For instance, assessors base their judgements on only
a few of the total set of dimensions that they have been instructed to use (Sackett & Hakel, 1979). Also, within-exercise dimension ratings show poor discriminant validity (Sackett & Dreher, 1982; Turnage & Muchinsky, 1982). These findings suggest that the influence of various characteristics of dimensions on the quality of assessors' judgements should be investigated.

This study explored the effects of one particular characteristic of assessment center dimensions, the continuum of abstractness/concreteness. Although all dimensions are behaviorally-based, some can be very trait-oriented (abstract) and require an assessor to make inferences when classifying behaviors and rating dimensions. For example, behaviors indicative of the dimension assertiveness may be difficult to determine. Individual assessors may have very different ideas about what constitutes assertive behavior. Consequently, reduced rating accuracy and inter-assessor agreement may result. In contrast, other dimensions may be very task-oriented (concrete) and require little inference for behavior classification and evaluation. For example, little inference may be required to determine if a behavior should be categorized into a concrete dimension such as planning. Assessors may agree more readily about what behaviors are indicative of concrete dimensions, resulting in more accurate and reliable ratings.
Moreover, research in social cognition indicates that use of abstract dimensions may be disadvantageous (Cantor & Mischel, 1977; Lord & Foti, 1986; Schank & Abelson, 1977). When assessors use trait-related dimensions, they are likely to associate behaviors they observe with a prototype. Prototypes are developed through experience, and they consist of traits and behaviors that have become associated over time. When assessors try to recall a behavior, they may have difficulty separating behaviors actually observed from behaviors not observed but associated with one of their prototypes. This would be especially problematic in the case of abstract dimensions, where the prototypes may be more well-defined. Errors in ratings could result.

Roles of an Assessor

An assessment center is a method designed to evaluate job applicants for a variety of purposes, including selection and promotion of personnel. It consists of several exercises, which are simulations of job-related situations developed to elicit behaviors determined to be important to the target job. These exercises include games, leaderless group discussions, and role plays. Candidates perform in these exercises, and are observed by assessors. The assessors, who are usually upper-level management personnel, are trained to observe and evaluate the performance of the job candidates.
Assessors perform many different duties in the assessment center method. For example, they may perform roles during exercises. While performing their roles, assessors may also be asked to carefully observe candidates’ behavior. In other exercises, assessors are simply observers. In either case, candidates are observed as they perform, and assessors are required to accurately observe and record their behavior. Although all assessors do not observe all candidates in every exercise, each assessor observes at least one, and possibly more, candidates in each exercise. The candidate to assessor ratio in a typical assessment center is 2:1 (Thornton & Byham, 1982). Therefore, an assessor will often be asked to observe and record behaviors about more than one candidate simultaneously.

Next assessors classify behaviors into dimensions. In pre-assessment center training assessors learn the definitions of the dimensions and the behaviors representative of each one to aid them in classification. Often they are asked to classify behaviors into a large number of dimensions, some of which may overlap. They are also asked to clearly communicate the behaviors that they observed to the other assessors during the assessor discussion. They must be able to utilize information given to them by other assessors because no assessor observes every candidate perform in every exercise. After all observations are relayed to the assessor group, assessors
review all information and independently make across-exercise dimension ratings. These ratings are posted and discrepancies are discussed until consensus is reached. Finally, an independent overall rating of performance is made by each assessor which is also discussed until consensus is reached.

In summary, the assessor’s job is quite complex. However, the assessment center process has been designed to aid assessors in accomplishing their numerous and difficult tasks. For example, the assessment center procedure is quite systematic; assessors make judgements in a specified order. Also, assessors are typically trained extensively on each of their tasks. Nonetheless, the requirements of assessors are quite extensive. Assessors are asked to do many things simultaneously and to make inferences when categorizing and evaluating candidates. Assessors are asked to perform in this manner in order to establish a valid and reliable personnel assessment measure.

Reliability and Validity of Dimensions

As a precursor to validity, a measurement device must be reliable. Three forms of reliability are possible in an assessment center: (a) test-retest reliability—the stability of assessors’ judgements over successive administrations of the assessment center to the same candidates; (b) inter-rater reliability—the agreement among assessors within- and
across-exercises; and (c) internal consistency—the consistency (convergence) of parallel measurements of constructs (dimensions) across exercises (Hinrichs & Haanpera, 1976). Several lines of research on the reliability of assessment center dimensions indicate that assessment centers have adequate test-retest and inter-rater reliability (see Thornton & Byham, 1982, for a review). Research has not been highly supportive of the internal consistency of assessment centers, however. Dimensions have been found to be more highly correlated within than across exercises (Hinrichs & Haanpera, 1976; Sackett & Dreher, 1982; Sackett & Hakel, 1979; Turnage & Muchinsky, 1982). This problem affects not only reliability, but also suggests that assessment centers may not be measuring the constructs they are designed to measure.

The assessment center process is based on the notion that "behavior predicts behavior" (Thornton & Byham, 1982, p. 235). Dimensions are a way of organizing and categorizing behavior. In the assessment center model, behaviors of an individual are believed to be consistent across exercises. Dimensions are a way of categorizing these consistent behaviors, and therefore are often regarded as constructs. Constructs are generally some concept, attribute, or quality which is assessed through some measure. Construct validity has been defined as "whether the ratings of personal and/or job
related traits of an individual reflect his job performance" (Kavanaugh, MacKinney, & Wolins, 1971, p. 35). Ratings of dimensions are described by some as attempts to measure constructs (Sackett & Dreher, 1982; Bycio, Hahn, & Alvarez, 1987). One way to establish the construct validity of dimensions is through demonstration of the convergent and discriminant validity of dimension ratings.

In an assessment center, the issue of convergent validity is whether behavioral constructs, or dimensions, can be reliably measured across different exercises. Discriminant validity, on the other hand, refers to the assessor's ability to discriminate among dimensions. Discriminant validity is measured in two ways. First, convergent validity, or correlations of the same dimension across exercises, should be higher than correlations among different dimensions across exercises. Second, correlations of the same dimension across exercises should be greater than correlations among different dimensions in the same exercise. If dimensions have both high convergent and discriminant validity, then they have adequate construct validity. Several researchers, however, have found dimensions to have low convergent and discriminant validity. For example, Turnage and Muchinsky (1982) found that assessment center dimension ratings gave little information beyond that given by the global rating. Sackett and Dreher (1982) and Russell (1987) found that within-exercise
correlations among dimensions were higher than correlations of the same dimension across exercises. Hinrichs and Haanpera (1976) found low reliability of dimensions across exercises. Moreover, using factor analysis, Bycio, Alvarez, and Hahn (1987) found that dimension ratings were largely situation specific, rather than being highly related across exercises. Others have found evidence of convergent validity but not of discriminant validity (Gaugler & Thornton, 1989; Silverman, Dalessio, Wood, & Johnson, 1986).

**Explanations for Lack of Convergent and Discriminant Validity**

Several explanations have been proposed to explain the lack of convergent and discriminant validity of assessment center dimensions. One explanation is that the behavior of candidates in an assessment center is situationally-determined: people perform differently according to the exercise in which they are participating (Neidig & Neidig, 1984). Neidig and Neidig propose that exercises are constructed to provide opportunities for candidates to perform in a variety of situations so that specific aspects of the same dimension can be measured in different exercises. Therefore, behaviors indicative of dimensions may vary across exercises. Exercises are developed in such a way that in different exercises, different behaviors may be relevant to the dimensions across all candidates. Consequently, low
convergent validity of within-exercise dimension ratings may not be erroneous but instead may reflect differences in exercise content.

Another explanation for the lack of convergent and discriminant validity is that individual exercises may not provide sufficient opportunity for behaviors relevant to a construct to be observed. Sackett and Dreher (1982) argue that high reliability estimates for some dimensions provide support for using these dimensions as representations of constructs, and that adding additional exercises to assessment center's could increase reliability estimates for all dimensions. This seems reasonable, although it has not been tested empirically.

Neidig and Neidig (1984) have also argued that there may be real exercise effects within individual candidates in assessment centers. Some candidates may perform better in some exercises than in others. For example, some candidates may perform well in an exercise in which they work on an individual basis, while others may do better in group situations. This hypothesis also awaits empirical testing.

Yet another argument is that candidate ratings may be unreliable due to variables such as practice effects or motivational changes within candidates. However, both Bycio (1987) and Cohen and Sands (1978) have found that exercise order does not affect post-exercise ratings. In other words, ratings of candidates in particular exercises presented in
different orders are the same. Consequently, it seems unlikely that the low internal consistency of dimension ratings is due to practice effects or motivational changes in candidates. In summary, the poor convergence of dimension ratings across exercises could indicate true performance differences within candidates, the situationally specific definition of dimensions, or the lack of dimension observability in certain exercises.

Despite their poor construct validity, dimensions, and the underlying constructs that they are purported to represent, continue to be used in the assessment center method, primarily because dimensions are psychometrically sound in other ways. For example, most dimensions have adequate inter-rater reliability (mean $r > .70$) and they predict overall job performance well (mean $r > .30$) (Gaugler, Rosenthal, Thornton, & Bentsen, 1987). However, researchers must continue to explore dimensions to determine which characteristics produce the highest quality judgments.

**Abstractness of Dimensions: Cognitive and Social Cognition Research**

Research from social cognition and cognitive psychology provides some clues concerning how dimensions can be more effectively used in the assessment center process. Dimensions are designed to help assessors make judgements by allowing them to break behavioral information into smaller
units. Social cognition research and theory, however, suggests that use of abstract, trait-related dimensions may actually decrease memory accuracy. When assessors organize information using trait-like dimensions, they are likely to associate the originally observed behaviors with other behaviors and traits that are stored in their memories. Cantor and Mischel (1977) argue that information about others may be categorized into basic personality categories called prototypes. Prototypes consist of related traits and behaviors that have become associated over time. Prototypes are useful in many social situations because they help interpret behaviors. Through their use, observations are supplemented with information from previous associations that have been stored in memory. This helps reduce incoming information and make it more manageable. In an assessment center, however, it is more important for an assessor to make accurate observations of incoming behavior than to summarize behavior or to depend upon their memories of past experiences.

People tend to particularly rely on the use of prototypes when cognitive demands are high. For instance, White & Carlston (1983) found that when subjects were faced with higher cognitive demand, they depended more on their previously formed impressions of the target to make judgements of behaviors. Subjects were less accurate in identifying behaviors that did not actually occur but which
were consistent with their ideas about the target. They reported that these types of behaviors did occur when actually they did not. Subjects were less accurate in identifying behaviors that were inconsistent with their ideas about the target, but which did occur. They were less likely to identify such behaviors as having occurred. This research seems especially relevant to the assessment center, where the cognitive demands of assessors are quite high. Assessors may rely on prototypes when using abstract dimensions, because the dimensions are so similar to actual prototypes that the assessors have formed through experience. For instance, the prototype of an assertive person might include behaviors such as "holds his/her own in arguments" and other traits such as "outgoing" or "direct." "Assertiveness" is also a frequently used dimension. Concrete dimensions may be less similar to prototypes, and therefore may be less problematic. Concrete dimensions, such as planning, might be similar to behaviors contained within prototypes, but may be less like a complete prototype. More rater error may occur, and less accurate dimension ratings within- and across-exercises may result when abstract dimensions are used.

An alternate hypothesis is that abstract dimensions may be included in fuzzy conceptual categories. Some categories, or prototypes, are less well-defined, or fuzzy, than others (Rosch, 1975). Behaviors indicative of abstract dimensions
may be more difficult for assessors to determine because these dimensional categories are more ill-defined than concrete dimensions. Concrete dimensions may be dealt with more easily because they are very task/behaviorally oriented. They may be easier to use because they are more well-defined.

Research concerning subjects' natural descriptions of others may also help explain why assessment center dimension ratings have poor construct validity. Researchers have found that when subjects are asked to describe others, they most commonly report information about appearance, attitudes, and behaviors. Schneider, Hastorf, & Ellsworth (1979) cite two unpublished studies in which it was found that traits are not the item most commonly used to describe other people. Moreover, one of these studies demonstrated that when describing others, subjects begin their descriptions with very concrete, observable features and behaviors. Only later did subjects begin to describe others using more abstract terms. When people become well acquainted with others, they are more likely to use traits to describe them. In an assessment center, however, assessors are asked to organize information about other people, many of whom they do not know, (the candidates) by dimensions many of which are trait-oriented and abstract. It would seem that dimensions based on abstract, trait-like categories are disadvantageous because assessors must process information differently than they
would naturally. Therefore, it seems that it would be more effective to have assessors organize behaviors around more concrete categories.

**Abstract Dimensions: Assessment Center Research**

One potential solution to the dimension construct validity problem is to have assessors evaluate candidates using concrete, task-oriented dimensions. Initial support for this idea has been provided in an exploratory study conducted by Gaugler, Cooke, Smith, and Swerdlin (1990). In this study, assessors were asked to sort candidates' behaviors into piles according to their similarity. Subjects then labelled the different piles of behaviors that resulted from their sorting. The majority of the labels created were concrete and action-oriented, rather than trait-oriented, indicating that in an assessment center, assessors may naturally categorize behavior into concrete, action-oriented dimensions. Examples of action-oriented, concrete dimensions include planning and organizing and decision making. Concrete dimensions should aid assessors in organizing, and therefore remembering behavioral information. Because they are not trait-oriented, these dimensions may not be as subject to unwanted inferences as abstract dimensions.

Shore, Thornton, and Shore's (1989) research also supports grouping dimensions according to their place on a
concrete/abstract continuum. In their study, dimensions were divided into two sets: (a) interpersonal style dimensions, which purportedly measured a candidate's typical interpersonal style, and (b) performance style dimensions, which measured a candidate's style when working with others. Interpersonal style dimensions included impact and personal acceptability, which are more abstract, trait-oriented dimensions, whereas performance style included oral communication and recognizing priorities, which are more concrete, task-oriented dimensions. A factor analysis supported this division of dimensions. It was found that the performance style dimensions correlated more highly with cognitive ability measures, while interpersonal style dimensions correlated more highly with conceptually similar personality measures.

Sackett and Dreher (1984) have recommended that if traditional dimensions are to be used at all, they be used to document and categorize behaviors. Concrete dimensions should fulfill this role of aiding in categorization, thus increasing accuracy. Additionally, if accuracy is increased by using concrete dimensions, then ratings of these dimensions should be more highly correlated across exercises and discrimination among dimensions within exercises should be improved. Zedeck (1986) states that dimensions may reduce the complexity of information obtained by the assessor, and that they aid processing by helping assessors to group people
and events together. In the case of concrete dimensions, this certainly could be true.

Concrete, behavioral dimensions might also reduce the inferential leap from behavioral observation to behavioral classification. It may be that it is much easier to categorize a behavior into "planning" than to categorize it into "sensitivity." Less inference may be required. Increased consistency in classifying behaviors should result in enhanced rating accuracy and convergent and discriminant validity of dimension ratings.

The Present Investigation

In the present study, undergraduates served as assessors in an assessment center simulation in which candidates for a residence hall student assistant position were evaluated. Subjects were trained as assessors and then were asked to observe videotaped performances of confederates playing the roles of candidates. They observed candidates' performance and then classified observed behaviors into either concrete, task-oriented dimensions or abstract, trait-related dimensions. After classification, subjects independently rated each dimension within-exercises. After observing each candidate in one exercise and classifying behaviors, making a total of three sets of exercise observations and classifications, assessors met in groups of three and described
their observations to the other assessors in their group. All assessors then made across-exercise ratings.

The type of dimension assessors evaluated was hypothesized to influence the quality of their judgments. Specifically, subjects who evaluated candidates using concrete dimensions were predicted to have significantly higher classification and rating accuracy than subjects who evaluated candidates using abstract dimensions. Moreover, ratings of subjects who evaluated candidates using concrete dimensions were hypothesized to have significantly higher inter-rater reliability and discriminant and convergent validity than ratings of subjects who evaluated candidates using abstract dimensions.

Method

Pilot Work and Creation of Dimension Sets

In order to allow exploration of the underlying characteristics of dimensions, several pilot tasks were administered. First, a group of community college students were asked to complete both a sorting task and a ranking task, beginning with the sorting. The pool of dimensions consisted of those dimensions found to be important to the target position of a student assistant in a job analysis conducted by Gaugler and Thornton (1989). Appendix A contains a list of
these dimensions. The community college students were asked to sort this pool of dimensions into different groupings according to any organizational scheme they felt was appropriate. Next the students labelled their groups of dimensions. A multi-dimensional scaling (MDS) analysis was performed on the students' dimension groupings in order to determine the dimensions' underlying characteristics. This analysis showed that students tended to group task-oriented dimensions such as interpersonal skills, teamwork, monitoring, and delegation together. Likewise, trait-oriented dimensions such as initiative, self-confidence, assertiveness, and creativity were often grouped together. Appendix B contains the MDS solutions.

The second task completed by the community college students consisted of a ranking task. In their study exploring the natural categories into which subjects placed assessment center behavioral descriptions, Gaugler et al. (1990) found that behaviors were often grouped into a concrete/abstract dichotomy. Therefore, the students were asked to rank the pool of dimensions according to their abstractness/concreteness. As a cross check, a different group of university students also completed this ranking task. Finally, a third group of university students rated the dimensions according to the task-oriented/trait-oriented continuum that was identified in the sorting performed by the
community college students. Each of the groups of subjects used an alternate ranking procedure to order dimensions from the dimension most abstract or trait-oriented (1) to the dimension most concrete or task-oriented (24).

Appendix C lists the mean rankings for both continua for all of the dimensions in the dimension pool. It was found that the concrete/abstract rankings and the trait-oriented/task-oriented rankings were very similar. Because the rankings provided by the three groups of subjects were so similar (r = .78), they were combined and the means of the rankings were used to divide the dimensions into two groups: (a) abstract, trait-oriented dimension and (b) concrete, task-oriented dimensions. Dimensions were divided into two sets that maximized the difference between mean rankings. To keep the most separation on the abstractness/concreteness continuum and yet allow consideration of other possible confounds (described later), the four most concrete and the four most abstract dimensions were separated from the dimension pool. The dimensions with the highest mean ranking were planning/organizing, monitoring, oral communication, and delegation (M = 15.9). The dimensions with the lowest mean ranking were sensitivity, self-confidence, energy, and assertiveness (M = 8.3).

Each of these two sets of dimensions were arranged into different combinations of three and mean rankings for the
three-dimension combinations were computed. Past research suggests that three dimensions account for most of the variance in assessment center performance (Sackett & Hake1, 1979). Four sets of three dimension groups were determined to maximize the separation of the rankings. These four sets of dimensions were further evaluated in order to control for confounds.

In previous research on the influence of dimension use on assessment center judgments, two possible confounds were identified and controlled (Gaugler & Thornton, 1989). The first possible confound was the difference in importance of dimension sets. When an assessment center is created, a job analysis is performed to identify the dimensions most important to the target position. In the present experiment, differences in importance of the dimension sets could have resulted in differences in ratings. Therefore, it was not possible to simply select the three most abstract and the three most concrete dimensions. The most abstract or the most concrete dimensions might have had differing importance and this might have confounded the results.

The second confound identified in past research was similarity of dimensions. A confound would exist if the dimensions in one experimental condition were more similar than the dimensions in the other condition. For example, if the dimensions in the concrete, task-related condition were highly
similar and dimensions in the abstract, trait-related group were quite dissimilar, assessors who evaluated concrete dimensions might classify behaviors less accurately and make less reliable and accurate dimension ratings than assessors who evaluated abstract dimensions.

In order to control for these confounds, Gaugler and Thornton (1989) had a group of undergraduate university students rate all relevant dimensions according to their importance to the target position ("1" - not important at all to "10" - extremely important) and according to their similarity ("1" - not similar at all to "10" - virtually identical). These ratings were used in the present experiment to determine importance and similarity of the different three-dimension sets described previously. Appendix D presents mean importance and similarity ratings for the four three-dimension sets. The two dimension sets selected for this study had importance and similarity ratings that were approximately equal (importance $M = 6.73$ vs. $7.17$ and similarity $M = 5.80$ vs. $5.76$, concrete group vs. abstract group, respectively). The final set of abstract dimensions consisted of assertiveness, energy, and self-confidence. The final set of concrete dimensions consisted of planning/organizing, oral communication, and monitoring. The definitions of these dimensions are presented in Appendix E.
Design

A between subjects design with two levels of the independent variable was used to test the hypotheses. Subjects were randomly assigned to either a concrete, task-oriented dimension group or an abstract, trait-oriented dimension group.

Subjects and Procedure

Eighty-one university students participated as assessors in the study in exchange for class credit. Because one part of the procedure, the integration session, required groups of three people, subjects performed as observers and raters of group process if the number of subjects in an experimental session was not a multiple of three. Research assistants randomly chose which subjects would participate as observers in cases of excess subjects. The analyses were only performed on data from those 66 subjects who participated as assessors, yielding 33 subjects per group.

After being trained as assessors, subjects watched videotapes of three confederates performing as job candidates for the target position of student assistant. The tapes were created by giving the confederates descriptions of the dimensions and minimum standards of performance for each of three exercises. Confederates were instructed concerning the general level of performance they were to perform in each
exercise and for each dimension. The confederates acted in a leaderless group discussion, a counseling session, and a roommate conflict exercise. The tapes were developed so that variation existed within and among candidates, to ensure that they would be similar to actual assessment centers. A total of seven different exercises were videotaped: the three confederates together in the leaderless group discussion and then each individually in the counseling session and roommate conflict exercise.

Subjects classified behavior into dimensions after watching the videotapes. As in a typical assessment center, all assessors did not observe all candidates in all exercises. Each assessor observed each of the three confederates in only one exercise. Subjects made observations of exercises in an order designed to ensure that they were exposed one time to all confederates and to all exercises. When they had completed this process for all three tapes, the subjects shared their observations in three member teams. The teams were set up so that when observations were shared, the assessors would have knowledge of all confederates' performance in all exercises. Finally, subjects independently made within- and across-exercise dimension ratings. Subjects did not make consensus judgments and overall assessment center ratings of candidates.
Assessor Training

Assessor training focused on helping subjects understand the dimensions. Subjects were taught how to make behavioral observations and they practiced and received feedback on their performance. They were also trained to classify behaviors into dimensions and to evaluate candidate performance on the dimensions using a common standard. Training was 1 hour in length. This time length was seen as sufficient due to the results of a meta-analysis which suggests that length of training does not moderate assessment center validity (Gaugler et al., 1987). Appendix F contains training materials and outline.

Minimum Standards

In order to provide assessors with a common standard for rating behaviors, minimum standards of performance for each dimension in each exercise were developed. These minimum standards consisted of all behaviors a candidate must have exhibited in order to deserve an acceptable rating. (The minimum standards for each exercise are presented in Appendix D).

Videotaped Exercises

As previously mentioned, confederates performed in three exercises. The counseling session simulation featured a
student who had poor hygiene and academic difficulties. The roommate conflict concerned two students who had disagreements over use of their room. The leaderless group discussion consisted of candidates talking about possible changes to their school's alcohol policy. These exercises were chosen because of their relevance to the target position and because the dimensions were observable in these exercises (Gaugler & Thornton, 1989).

**Dependent Measures**

The dependent measures consisted of observation accuracy, classification accuracy, rating accuracy, inter-rater reliability, and convergent and discriminant validity.

**Observation Accuracy.**

In order to provide a standard against which subjects' behavioral observations could be compared, transcripts of the videotapes were prepared. Subjects' recordings of behaviors were compared with the transcripts using a general meaning scoring technique. In other words, subjects' wordings of the behavioral observations could be stated differently than the exact wording in the scripts and still be counted as good observations if the meaning was the same. A good observation stated what a person said or did, was specific rather than general, was descriptive rather than evaluative, and was
confirmable by others. The numbers of good observations on
the subjects' observation forms were counted. Observations
which were repeated were only counted one time. The total
number of good observations was then computed by summing
across all exercises for each subject.

Three methods were used to evaluate observation
accuracy. First, the ratio of good observations was compared
with the total possible number of observations for each group.
The total possible number of good observations was determined
through tabulations of behavioral statements contained in the
transcripts. (Observation Accuracy I = No. Good
Observations/No. Total Possible Observations). Second, the
number of good observations was compared with the total
number of observations made by each subject for each group
(Observation Accuracy II = No. Good Observations/No. Behaviors
Recorded by Subject). Third, the simple number of good
observations for each group was compared. These three
statistics permitted examination of different components of
observation: speed of observation or how many observations
the subject was able to record, and precision or correctly
worded and observed statements.

Classification Accuracy.

Classification accuracy was measured in two different
ways. First, for each exercise the number of behaviors
correctly classified by each subject was divided by the number of behaviors correctly observed by that subject (Classification Accuracy I = No. Behaviors Correctly Classified by Subjects/No. Behaviors Correctly Observed by Subjects). This measure yielded an indication of how well subjects were able to classify, taking into account their observation ability. Second, for each dimension within each exercise, the number of behaviors correctly classified by subjects across all exercises was divided by the number of behaviors classified by experienced assessors. The experienced assessors had multiple opportunities to view the tapes and were provided scripts of the tapes. They were given the opportunity to observe the tapes as many times as they wished, and to stop and/or rewind the tapes as needed. (Classification Accuracy II = No. Behaviors Correctly Classified by Subjects/No. Behaviors Classified by Experienced Assessors). This measure yielded an indication of how well the subjects performed in comparison with the more experienced assessors.

**Rating Accuracy.**

The average of the across-exercise dimension ratings made by the experienced assessors was compared with subjects' ratings to determine the accuracy of the subjects' across-exercise dimension ratings. Four rating accuracy measures were used: elevation, differential elevation,
stereotype accuracy, and differential accuracy (Chronbach, 1955). Elevation is the average rating of an assessor over all assessees and dimensions (overall accuracy). Differential elevation is the average rating an assessor gives to each assessee across all dimensions (person accuracy). Stereotype accuracy is the mean rating an assessor gives to each dimension (dimension accuracy). Finally, differential accuracy is an assessor’s sensitivity to assessee differences in patterns of performance (person X dimension accuracy).

**Inter-Rater Reliability.**

Inter-rater reliability was compared by calculating interclass correlations of subjects' ratings (Shrout & Fleiss, 1979). Inter-rater reliability in this study was computed according to Shrout and Fleiss' Case 2 model, where a random sample of k judges (assessors) was selected from a larger population and where each judge rated each target (confederate). Therefore, a Target X Judges two-way analysis of variance was the appropriate mode of analysis. Using this analysis, the within-target sum of squares was partitioned into a between-judges sum of squares and a residual sum of squares. The analysis of variance procedure provided information necessary to calculate an estimate of the population value of the intraclass correlation coefficient among assessors, using the following formula:
BMS-EMS

\[
BMS + (k-1)EMS + k(JMS - EMS)/n,
\]

where BMS = between targets sum of squares, EMS = residual sum of squares, JMS = between judges sum of squares, k = number of judges, and n = number of targets.

**Convergent Validity and Discriminant Validity.**

Convergent and discriminant validity were assessed using two different procedures, a multitrait multimethod matrix and an analysis of variance procedure. Correlations among all dimensions were computed to form a multitrait multimethod matrix for each dimension group. Convergent validity is present when correlations among the same dimension across different exercises are high (monotrait-heteromethod correlations).

The multitrait multimethod matrix provided two different assessments of discriminant validity (Campbell & Fiske, 1959). First, monotrait-heteromethod correlations should be larger than correlations among different dimensions measured by different exercises (heterotrait-heteromethod correlations). A second criterion for discriminant validity is that monotrait-heteromethod correlations be greater than correlations among different dimensions measured by the same exercise (heterotrait-monomethod correlations).
The multitrait multimethod matrix provided a large amount of information with which the construct validity of the dimensions could be evaluated through examination of the convergent and discriminant validity of the ratings. One problem inherent in using this method, however, was that relative judgements were difficult to determine from the matrix. Comparison of effects was somewhat difficult. For this reason, an analysis of variance procedure was used (Stanley, 1971). This procedure supplied a summary of the data which allowed comparison of effects.

The sources of variation in the analysis of variance model were (a) variation due to assessees (convergent validity); (b) variation due to the assessee by dimension interaction (discriminant validity); (c) variation due to the interaction of assessees and exercises (source of exercise variation); and (d) error. Variation due to the assessees indicates the overall amount of agreement on assessees over exercises and dimensions. The assessee X exercise interaction indicates the amount of situational specificity or method variance of dimension ratings.

The ANOVA analysis resulted in two additional measures: (a) variance components (VC's), which allowed for comparison of effects within a dimension group, and (b) intraclass correlations (ICC's), which allowed for comparison of effects between dimension groups.
Results

Observation Accuracy

The first observation accuracy measure indicated speed of observation or how many observations the subject was able to record and the second measure indicated assessors' precision, meaning correctly worded and observed statements. The third measure allowed a simple comparison of the subjects' observations. Table 1 contains the means and standard deviations of the observation accuracy scores. T-tests were used to evaluate the influence of dimension type on measures of observation accuracy.

Dimension type did not significantly influence the accuracy of assessors' observations (Obs. Acc I: $t(63) = .938$, n.s.; Obs. Acc. II: $t(63) = 1.14$, n.s.; Obs. Acc. III: $t(63) = .96$, n.s.).

Classification Accuracy

The first classification accuracy measure indicated how well subjects were able to classify behaviors, taking into account their observation ability. The second classification accuracy measure indicated how well the subjects performed in comparison with the target. Results of the T-tests for independent groups performed on each accuracy measure indicated that subjects who rated concrete dimensions
Table 1
Means and Standard Deviations of Observation Measures

<table>
<thead>
<tr>
<th>Condition</th>
<th>Concrete Dimension</th>
<th>Abstract Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent measure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation Accuracy I</td>
<td>0.394</td>
<td>0.365</td>
</tr>
<tr>
<td>M</td>
<td>0.113</td>
<td>0.134</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation Accuracy II</td>
<td>0.833</td>
<td>0.803</td>
</tr>
<tr>
<td>M</td>
<td>0.090</td>
<td>0.117</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation Accuracy III</td>
<td>45.719</td>
<td>42.182</td>
</tr>
<tr>
<td>M</td>
<td>13.125</td>
<td>16.303</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

classified behaviors more accurately than subjects who rated abstract dimensions (Class. Acc. I: t(64) = 2.40, p<.02, ω²=.07; Class. Acc. II: t(64) = 3.89, p<.001, ω²=.19). Table 2 contains the means and standard deviations of both classification measures.

Rating Accuracy

A multivariate analysis of variance indicated that type of dimension significantly affected rating accuracy across the four rating accuracy measures of elevation, differential
Table 2
Means and Standard Deviations of Classification Measures

<table>
<thead>
<tr>
<th>Condition</th>
<th>Concrete Dimension</th>
<th>Abstract Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent measure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification Accuracy I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.816</td>
<td>1.533</td>
</tr>
<tr>
<td>SD</td>
<td>.427</td>
<td>.536</td>
</tr>
<tr>
<td>Classification Accuracy II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.144</td>
<td>2.417</td>
</tr>
<tr>
<td>SD</td>
<td>.801</td>
<td>.676</td>
</tr>
</tbody>
</table>

elevation, stereotype accuracy, and differential accuracy (F(4,61) = 12.98, p<.001). Univariate \( F \) tests performed on three of the four accuracy components were statistically significant, with the exception of the final accuracy measure, differential accuracy (Elevation: \( F(1,64) = 4.48, p<.058, \omega^2=.08 \); Differential elevation: \( F(1,64) = 13.81, p<.001, \omega^2=.16 \); Stereotype Accuracy: \( F(1,64) = 32.2, p<.0001, \omega^2=.32 \); Differential Accuracy, \( F(1,64) = 3.32, p>.05 \). Table 3 contains the means and standard deviations for rating accuracy measures. Ratings of subjects who evaluated concrete dimensions were significantly more accurate than those of
subjects who rated abstract dimensions on the measures of elevation ($M = 2.081$ and $2.278$, respectively) and differential elevation ($M = .418$ and $.650$ respectively). Contrary to prediction, however, subjects who evaluated abstract dimensions made ratings with significantly greater stereotype accuracy than subjects who rated concrete dimensions ($M = .878$ and $1.193$, respectively).

Table 3
Means and Standard Deviations of Rating Accuracy Measures

<table>
<thead>
<tr>
<th>Condition</th>
<th>Concrete Dimension</th>
<th>Abstract Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent measure</td>
<td>Elevation</td>
<td>Differential Elevation</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td></td>
<td>2.081</td>
<td>.390</td>
</tr>
</tbody>
</table>

Note: Lower scores indicate higher rating accuracy.
Inter-Rater Reliability

Tables 4 and 5 display the analysis of variance results on assessors' across exercise ratings, which were used to calculate intraclass correlation coefficients for both assessor groups. Examination of the interclass correlations indicates

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Targets</td>
<td>2</td>
<td>23.94</td>
</tr>
<tr>
<td>Within Target</td>
<td>96</td>
<td>1.04</td>
</tr>
<tr>
<td>Between Assessors</td>
<td>32</td>
<td>1.21</td>
</tr>
<tr>
<td>Residual</td>
<td>64</td>
<td>.95</td>
</tr>
</tbody>
</table>

Table 5
Analysis of Variance for Assessor Ratings: Concrete Group

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Targets</td>
<td>2</td>
<td>27.36</td>
</tr>
<tr>
<td>Within Target</td>
<td>96</td>
<td>1.10</td>
</tr>
<tr>
<td>Between Assessors</td>
<td>32</td>
<td>1.14</td>
</tr>
<tr>
<td>Residual</td>
<td>64</td>
<td>1.08</td>
</tr>
</tbody>
</table>

that there were no differences in inter-rater reliability between assessor groups (ICC = .40 vs .42, abstract vs. concrete group respectively).
Convergent Validity and Discriminant Validity: Multitrait Multimethod Matrix

Table 6 displays the mean exercise and dimension correlations from the multitrait multimethod matrix for both groups. Convergent validity is present when correlations among the same dimension across different exercises are high (monotrait-heteromethod correlations). Neither group met this criterion, although the concrete group (mean $r = .21$) had a larger mean correlation than the abstract group (mean $r = .14$).

Examination of the multitrait multimethod correlations reveals that subjects in both conditions produced mean monotrait-heteromethod correlations that were larger than mean heterotrait-heteromethod correlations, which is one criteria of discriminant validity (concrete group: .21 vs. .16, abstract group: .14 vs. .11). These mean correlations indicate that the ratings of the concrete dimension group had somewhat better discriminant validity than ratings of the abstract dimension group.

Also as indicated in Table 6, neither group met the criterion of discriminant validity that monotrait-heteromethod correlations be greater than heterotrait-monomethod correlations. For both groups the heterotrait-
monomethod ratings were higher than the monotrait-heteromethod ratings.

Table 6

Mean Dimension and Exercise Correlations

<table>
<thead>
<tr>
<th>Dimension (Monotrait-Heteromethod Correlations)</th>
<th>Concrete Dimension Group</th>
<th>Abstract Dimension Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning and Organizing</td>
<td>.18</td>
<td>.</td>
</tr>
<tr>
<td>Oral Communication</td>
<td>.21</td>
<td>.</td>
</tr>
<tr>
<td>Monitoring</td>
<td>.23</td>
<td>.</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>.</td>
<td>.09</td>
</tr>
<tr>
<td>Energy</td>
<td>.</td>
<td>.14</td>
</tr>
<tr>
<td>Self-Confidence</td>
<td>.</td>
<td>.19</td>
</tr>
<tr>
<td>Grand Mean</td>
<td>.21</td>
<td>.14</td>
</tr>
</tbody>
</table>

| Heterotrait-Heteromethod Correlations--Grand Mean | .16 | .11 |

<table>
<thead>
<tr>
<th>Exercises (Heterotrait-Monomethod Correlations)</th>
<th>Concrete Dimension Group</th>
<th>Abstract Dimension Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counselling Session</td>
<td>.62</td>
<td>.34</td>
</tr>
<tr>
<td>Roommate Conflict</td>
<td>.50</td>
<td>.52</td>
</tr>
<tr>
<td>Group Discussion</td>
<td>.29</td>
<td>.45</td>
</tr>
<tr>
<td>Grand Mean</td>
<td>.47</td>
<td>.44</td>
</tr>
</tbody>
</table>
Convergent Validity and Discriminant Validity: ANOVA Analysis

The ANOVA summary tables for both conditions are presented in Tables 7 and 8. As indicated in the tables, the results of the significance tests on the main effects and the interactions indicate that each source of variance was significant. For both groups of subjects, there was significant convergent validity, but this result was tempered by the fact that large method bias existed. Finally, there was significant discriminant validity for both groups.

Table 7

Analysis of Variation of Concrete Dimension Correlations

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessee (convergent validity)</td>
<td>32</td>
<td>3.10</td>
<td>6.21*</td>
</tr>
<tr>
<td>Assessee X Exercise (method bias)</td>
<td>64</td>
<td>1.45</td>
<td>2.90*</td>
</tr>
<tr>
<td>Assessee X Dimension (discriminant validity)</td>
<td>64</td>
<td>.65</td>
<td>1.30*</td>
</tr>
<tr>
<td>Error</td>
<td>128</td>
<td>.50</td>
<td></td>
</tr>
</tbody>
</table>

*p < .001
Table 8
Analysis of Variance for Abstract Dimension Correlations

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessee (convergent validity)</td>
<td>32</td>
<td>2.67</td>
<td>4.88*</td>
</tr>
<tr>
<td>Assessee X Exercise (method bias)</td>
<td>64</td>
<td>1.57</td>
<td>2.87*</td>
</tr>
<tr>
<td>Assessee X Dimension (discriminant validity)</td>
<td>64</td>
<td>.64</td>
<td>1.17*</td>
</tr>
<tr>
<td>Error</td>
<td>128</td>
<td>.55</td>
<td></td>
</tr>
</tbody>
</table>

*p < .001

Table 9 displays the variance components (VC's) for each condition. From the VC's, which allowed comparison of effects within a dimension group, it can be seen that ratings of subjects in both groups had high convergent validity and method bias and low discriminant validity. The large assessor x exercise interaction indicated that the convergent validity was quite method bound. The ICC's, which allowed comparison of effects across conditions, are presented in Table 10. For both groups, it can be seen that the largest effect was variance due to the assessee by exercise interaction (method variance), and that this effect was slightly less
Table 9  
**Variance Components**

<table>
<thead>
<tr>
<th>Source</th>
<th>Abstract Dimensions</th>
<th>Concrete Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assesssee</td>
<td>.236</td>
<td>.289</td>
</tr>
<tr>
<td>Assesssee X Exercise</td>
<td>.341</td>
<td>.316</td>
</tr>
<tr>
<td>Assesssee X Dimension</td>
<td>.031</td>
<td>.049</td>
</tr>
</tbody>
</table>

for subjects who evaluated concrete dimensions than for subjects who evaluated abstract dimensions. The next largest effect for both groups was variance due to assesssee (convergent validity), but this effect was less for subjects who evaluated abstract dimensions than for subjects who evaluated concrete dimensions. Variance due to the assesssee by dimension interaction (discriminant validity) was very low for both groups of subjects. Discriminant validity was higher for subjects who evaluated concrete dimensions, however, than for subjects who evaluated abstract dimensions. Although quite small in magnitude, the discriminant validity was almost twice as large for assessors who evaluated concrete dimensions.
### Table 10
**Intraclass Correlations**

<table>
<thead>
<tr>
<th>Source</th>
<th>Abstract Dimensions</th>
<th>Concrete Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessee</td>
<td>.204</td>
<td>.251</td>
</tr>
<tr>
<td>Assessee X Method</td>
<td>.295</td>
<td>.274</td>
</tr>
<tr>
<td>Assessee X Trait</td>
<td>.027</td>
<td>.043</td>
</tr>
</tbody>
</table>

**Discussion**

In general, it seems that dimension concreteness affected some aspects of assessor judgement, but did not affect others. Subjects were better able to categorize behaviors into dimensions and their ratings were more accurate according to two of four accuracy measures when they used concrete dimensions for evaluation. Moreover, convergent and discriminant validity were somewhat higher and method bias was lower for ratings made by subjects who used concrete dimensions than for subjects who used abstract dimensions. Unfortunately, no test was available to test the statistical significance of these convergent validity differences. Dimension concreteness did not significantly affect observation accuracy and differential accuracy of subjects' ratings. When subjects used abstract dimensions, however, their ratings were more accurate
according to one measure, stereotype accuracy, than when subjects used concrete dimensions.

Observation Accuracy

No differences were predicted for this variable. Results showed that the different dimension conditions did not affect subjects’ observations. Perhaps subjects do not attend to dimensions when they are making their observations. Possibly as subjects become more experienced, they begin to attend to dimensions as they make their observations.

Classification Accuracy

As hypothesized, subjects who worked with concrete dimensions were better able to classify behaviors than subjects who worked with abstract dimensions. Dimensions are required in assessment centers so that assessors can better store and process information (Zedeck, 1979). This data supports the idea that concrete dimensions are more helpful than abstract dimensions in aiding assessors to store and process information, thereby making their task less difficult at the classification stage of the rating process. This result is consistent with recent research in which it was found that assessors shared better classification agreement for their natural categories than
for traditional assessment center dimensions (Gaugler et al., 1990). In this research, the labels of assessors' natural categories were more concrete than are the labels of many assessment center dimensions.

**Rating Accuracy**

Subjects in the concrete dimension group made more accurate ratings according to two of the four accuracy measures, which partially confirms the hypothesis that subjects using concrete dimensions would be more accurate raters than subjects using abstract dimensions. Ratings of assessors using concrete dimensions were more accurate according to the measures of elevation and differential elevation, indicating that subjects who evaluated concrete dimensions had better overall accuracy and person accuracy than subjects who evaluated abstract dimensions.

Perhaps assessors do rely upon prototypes when they are required to use abstract dimensions. As discussed previously, this reliance on prototypes could have accounted for the less accurate ratings made by assessors who used abstract dimensions than for assessors who used concrete dimensions on two of the four accuracy measures. Assessors who evaluated abstract dimensions may have based their ratings less on behaviors which actually occurred in the assessment center and more on behaviors
and traits that were stored in their memories. Alternately, assessors who rated abstract dimensions may have been less accurate because the prototypes for abstract dimensions were less well-defined. Less well-defined prototypes could increase the cognitive load of assessors, thereby reducing their accuracy.

Contrary to prediction, however, subjects who rated abstract dimensions had significantly better stereotype accuracy than subjects who rated concrete dimensions. Stereotype accuracy is the component of accuracy associated with the average rating of a dimension across ratees. It is concerned with the rank ordering of the difficulty of the items (dimensions). Examination of the mean dimension ratings of both the experienced assessors and the subjects provides an explanation for the result of higher stereotype accuracy for subjects who rated abstract dimensions. The mean dimension rating made by the experienced assessors for each of the abstract dimensions was 2.78, with a grand mean of 2.78. The mean ratings of the subjects who rated abstract dimensions were 3.33, 2.95, and 2.82 for assertiveness, energy, and self-confidence respectively, with a grand mean of 3.03. As can be seen, the mean ratings of each dimension made by the experienced assessors lacked variation, and it was impossible to rank them. The mean ratings of each dimension made by the
subjects could therefore have been ranked in any order and still appear to be similar to the experienced subjects' dimension rankings, as long as the overall mean rating was similar to the overall mean rating made by the experienced assessors.

In contrast, the mean ratings of the concrete dimensions made by experienced assessors was 2.78, 3.00, and 2.56 for planning/organizing, oral communication, and monitoring respectively, with a grand mean of 2.78, whereas the mean ratings of these dimensions made by subjects was 3.38, 2.95, and 2.91, with a grand mean of 3.08. The subjects ranked planning/organizing and oral communication in a different order than did the experienced assessors. Consequently, superior stereotype accuracy of abstract dimension ratings seems questionable due to the lack of variability among ratings of the abstract dimensions made by the experienced assessors.

Inter-Rater Reliability

The correlations among subjects' ratings were somewhat low for both treatment groups, but were well within the range of inter-rater reliabilities found by other researchers (Hinrichs & Haanpera, 1976). In fact, the assessment center inter-rater reliabilities summarized by Hinrichs and Haanpera range from .00 to .98. Perhaps inter-
rater reliability is highly affected by the experience of the assessors, and the present manipulation was not powerful enough to overcome the lack of experience of either treatment groups of assessors.

**Convergent and Discriminant Validity**

In previous research on the construct validity of assessor ratings, it was found that within-exercise dimension ratings had poor discriminant validity (Sackett & Dreher, 1982; Silverman, Dalessio, Woods, & Johnson, 1986; Turnage & Muchinsky, 1982). Although the ratings made by subjects in the concrete dimension group had substantially better discriminant validity than the ratings of the abstract dimension group, neither group met the requirements for good discrimination among dimensions according to the magnitude of the effect sizes.

In contrast to past research where adequate convergent validity has been found (e.g. Sackett & Dreher, 1982; Silverman, Dalessio, Woods, & Johnson, 1986; Turnage & Muchinsky, 1982), both groups of assessors made ratings of only moderate convergent validity. However, when compared to convergent validity found in other analysis of variance procedures, the convergent validity in this study was adequate (Kavanaugh et al., 1971). The concrete
dimension ratings had greater convergent validity than the abstract dimensions ratings.

High method variance, moderate convergent validity, and poor discriminant validity indicate that construct validity of within-exercise dimension ratings of both groups of subjects was poor. Sackett and Dreher (1982) argued that assessors evaluate performance in exercises rather than performance on dimensions, and this study supports that contention. It was encouraging to note, however, that method variance was lessened and discriminant and convergent validity were increased when concrete dimensions were evaluated.

Conclusions

In summary, it seems that use of concrete dimensions allowed assessors to better process information and to make ratings which were psychometrically superior in some respects. Assessors who evaluated concrete dimensions better classified observed behaviors and made more accurate ratings according to two types of accuracy measures than assessors who evaluated abstract dimensions. Although results were encouraging, more research is necessary to make further conclusions about the use of concrete dimensions.
For example, further research is needed to determine exactly how concrete dimensions influence the judgment process in the assessment center method. Concrete dimensions make classification easier, but only improve some aspects of rating accuracy. Therefore, it appears that simply increasing assessors’ ability to classify behaviors is not enough to fully improve rating accuracy. The accuracy of assessors’ ratings is probably due to many factors, only one of which is the accuracy with which behaviors are classified.

The effects of use of concrete dimensions on the consensus discussion should also be studied. In the present experiment it was found that subjects who rated concrete dimensions were better able to classify behaviors and perhaps this would be helpful during consensus discussion. A larger number of correctly classified behaviors might aid assessors in discussing the candidates. This could result in consensus discussions where agreement among assessors was enhanced. Better across-exercise consensus dimension ratings of candidates, as well as better overall ratings, might result. Similarly, the fact that concrete dimensions had somewhat better rating accuracy and construct validity suggests that across-exercise consensus ratings of concrete dimensions and overall ratings may also be more accurate.
Some have advocated that dimensions be deleted from the assessment center method (Bycio, Alvarez, & Hahn, 1987; Sackett & Dreher, 1984). Others have argued that exercises should be the main focus of assessment centers, and that dimensions should only be used for organizational purposes (Zedeck & Cascio, 1984; Zedeck, 1986). The present assessment center process produces global ratings that have good criterion-related validity, however, so if such changes are made in the assessment center method, it must be demonstrated that predictive validity is not reduced. Moreover, recent research suggests that across-exercise dimension ratings, on which most administrative decisions are based, do have adequate construct validity (Shore, Thornton, & Shore, 1990). In the words of Zedeck and Cascio (1984), "if the global rating works, don't fix the process." Continued research on the method is necessary for an understanding of why the overall rating is valid.

One limitation of this research must be noted. First, the correlations among assessors' dimension ratings were low, both within and across exercises. This could indicate that subjects either disagreed on the level of performance of the confederates in the tapes or that subjects did not fully understand the level of performance necessary for adequate performance on the the dimensions. In either case, the lack of agreement would result in low reliability of the
assessment center, and low reliability would place a boundary on the validity of the entire process. Perhaps training was not long enough or thorough enough to produce consistent ratings. One study which has examined assessor training, however, found that increasing training time does not improve the psychometric quality of assessment center judgments (Gaugler et al., 1987). On the other hand, this study was concerned with the influence of concrete dimensions on the psychometric qualities of assessors' judgments. The assessors in the experiment had no assessment center experience and only a small amount of training, so it could not be expected that they would have the same inter-rater reliability or predictive validity as real assessors.

Second, several subjects, as well as the experienced raters, commented on the difficulty of observing behaviors that could be classified under one of the concrete dimensions, monitoring. Possibly results would have differed, and the effects of using concrete dimensions might have been much stronger, had the difficulty of working with this dimension been reduced.

It should also be noted that dimension characteristics other than concreteness/abstractness may exist which cause difficulty for assessors. In other words, the concreteness/abstractness continuum may only partially
capture the difference between dimensions that are easy to use and dimensions that are not. A multi-dimensional scaling procedure was used to examine the differences between dimensions, and concreteness/abstractness was one difference that was readily apparent. Other dimensions were also present, but they were quite difficult to interpret. These other dimensions need to be further explored in order to understand why some dimensions seem to be more easy to use than others.

The effect on assessor ratings of dimension concreteness might be further explored by isolating each component of the process (i.e. observation, classification, and evaluation) and examining the effects of these manipulations. For example, first subjects would be asked to classify dimensions into concrete and abstract dimensions. Next, a different sample of subjects would make within-exercise ratings for behaviors which had already been classified into concrete or abstract dimensions, and so on. This isolation of each step in the process would allow more insight into the exact effects of the concrete/abstract dimension manipulation on each step of the assessment center method.

Abstractness of dimensions might have a more pronounced effect when the cognitive demands on assessors are increased. Many have noted that the cognitive demands
on assessors are high (Bycio, et. al., 1987; Gaugler & Thornton, 1989; Silverman, et. al, 1986). In the present experiment, the demands placed upon the assessors was considerably less than what is the case in a typical assessment center. For example, subjects worked with only three dimensions, whereas typically they would work with a larger number. Assessors only observed one candidate per exercise, and in an actual assessment center they may be required to observe multiple candidates simultaneously. Future research might examine the effects of concreteness of dimensions under cases of varied cognitive demand.

The results of this research have some practical implications for assessment center research and use. Many requirements are made of assessors when they evaluate candidates in an assessment center. Because behaviors are more easily classified under concrete dimensions, use of concrete dimensions can help ease the cognitive burden placed upon assessors. This result is quite important for those who develop and administer assessment centers. Researchers must further develop and explore dimensions which are both important to the target position and concrete as well. Researchers must find alternate ways of representing the content of abstract dimensions. In an assessment center it is important for the dimensions to adequately represent the content of the job. Therefore, it is
not practical to simply eliminate all abstract dimensions without having good substitutes for them. The content of these dimensions is quite important, because content validity is the most popular method of validating assessment centers (Gaugler, Bentsen, & Pohley, 1990). Researchers must find additional ways to reduce the cognitive demands placed upon assessors. Perhaps then increased convergent and discriminant validity will be found.
References


Appendix A: Dimension Pool for Abstractness/Concreteness
Rating
Dimension Pool for Abstractness/Concreteness Rating

Oral Communication
Oral Presentation
Active Listening
Persuasiveness
Interpersonal Skills
Assertiveness
Rapport
Sensitivity
Leadership
Mediation
Teamwork
Motivation
Work Standards
Energy
Initiative
Creativity
Stress Tolerance
Interpersonal Tolerance
Self-Organization
Self-Confidence
Planning
Organizing
Delegation
Monitoring
Appendix B: Multi-Dimensional Scaling Solutions
<table>
<thead>
<tr>
<th>DIMB SYMBOL IS VALUE OF OBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>22 active listening</td>
</tr>
<tr>
<td>24 mediation</td>
</tr>
<tr>
<td>23 leadership</td>
</tr>
<tr>
<td>13 work stability</td>
</tr>
<tr>
<td>18 energy</td>
</tr>
<tr>
<td>15 initiative</td>
</tr>
</tbody>
</table>
DIMC SYMBOL IS VALUE OF OBS

oral comm
businessness oral presentation

10 self-confidence
12 assertiveness
3 initiative

4 stress tolerance
16 creativity

9 organizing
7 motivation
21 self-org.
planning 18 energy

0.2 0.5 0.8 1.1 1.4
PLOT OF DIMB*DIMC

SYMBOL IS VALUE OF

interpersonal tolerance

rappport  20  mediation
22 active listening

23 leadership

interpersonal skills  15 monitoring

2 teamwork

delegation  17  persuasiveness

oral comm  oral pres.

organiz
SAS

*DIMC  SYMBOL IS VALUE OF OBS

4 stress tolerance

18 energy

13 motivation  3 initiative

6 oral pres.  16 creativity

3 persuasiveness  12 assertiveness

21 self-org.

organizing 98 planning 10 self-confidence

DIMC
Appendix C: Mean Concreteness/Abstractness and Trait-Oriented/Task-Oriented Rankings
### Concreteness/Abstractness and Trait-Oriented/Task-Oriented Rankings

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Communication</td>
<td>18.31</td>
<td>17.44</td>
<td>20.93</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>17.54</td>
<td>15.80</td>
<td>20.06</td>
</tr>
<tr>
<td>Active Listening</td>
<td>13.69</td>
<td>12.84</td>
<td>12.40</td>
</tr>
<tr>
<td>Persuasiveness</td>
<td>12.31</td>
<td>11.52</td>
<td>13.47</td>
</tr>
<tr>
<td>Interpersonal Skills</td>
<td>10.85</td>
<td>13.67</td>
<td>10.27</td>
</tr>
<tr>
<td>Assertiveness</td>
<td>7.38</td>
<td>12.04</td>
<td>14.40</td>
</tr>
<tr>
<td>Rapport</td>
<td>9.00</td>
<td>8.52</td>
<td>7.67</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>3.31</td>
<td>4.16</td>
<td>6.20</td>
</tr>
<tr>
<td>Leadership</td>
<td>11.23</td>
<td>10.52</td>
<td>12.27</td>
</tr>
<tr>
<td>Mediation</td>
<td>17.46</td>
<td>14.40</td>
<td>14.67</td>
</tr>
<tr>
<td>Teamwork</td>
<td>17.15</td>
<td>18.08</td>
<td>15.67</td>
</tr>
<tr>
<td>Motivation</td>
<td>8.23</td>
<td>8.52</td>
<td>5.53</td>
</tr>
<tr>
<td>Work Standards</td>
<td>10.38</td>
<td>14.96</td>
<td>9.00</td>
</tr>
<tr>
<td>Energy</td>
<td>5.85</td>
<td>11.88</td>
<td>12.60</td>
</tr>
<tr>
<td>Initiative</td>
<td>11.31</td>
<td>8.64</td>
<td>13.67</td>
</tr>
<tr>
<td>Creativity</td>
<td>4.69</td>
<td>6.60</td>
<td>6.60</td>
</tr>
<tr>
<td>Stress Tolerance</td>
<td>11.23</td>
<td>12.92</td>
<td>12.20</td>
</tr>
<tr>
<td>Interper. Tolerance</td>
<td>12.69</td>
<td>12.24</td>
<td>8.07</td>
</tr>
<tr>
<td>Self-Organization</td>
<td>14.85</td>
<td>13.08</td>
<td>14.80</td>
</tr>
<tr>
<td>Self Confidence</td>
<td>4.54</td>
<td>10.00</td>
<td>5.87</td>
</tr>
<tr>
<td>Planning</td>
<td>21.23</td>
<td>15.08</td>
<td>16.00</td>
</tr>
<tr>
<td>Organizing</td>
<td>20.62</td>
<td>13.88</td>
<td>17.13</td>
</tr>
<tr>
<td>Delegation</td>
<td>18.38</td>
<td>14.84</td>
<td>18.13</td>
</tr>
<tr>
<td>Monitoring</td>
<td>19.00</td>
<td>15.16</td>
<td>14.47</td>
</tr>
</tbody>
</table>

Group 1 = undergraduates from Rice University ranking
concreteness/abstractness; Group 2=undergraduates from Houston Community College ranking concreteness/abstractness; Group 3=Graduate Students, Rice University ranking trait-oriented/task-oriented
Appendix D: Mean Similarity and Importance of Dimension Groups
<table>
<thead>
<tr>
<th>Dimension Set</th>
<th>Mean Importance</th>
<th>Mean Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning/Organizing Oral communication Monitoring</td>
<td>6.73</td>
<td>5.80</td>
</tr>
<tr>
<td>Monitoring Oral Communication Delegation</td>
<td>6.37</td>
<td>5.35</td>
</tr>
<tr>
<td>Assertiveness Energy Self-Confidence</td>
<td>7.17</td>
<td>5.76</td>
</tr>
<tr>
<td>Sensitivity Self-Confidence Assertiveness</td>
<td>7.63</td>
<td>4.96</td>
</tr>
</tbody>
</table>
Appendix E: Dimensions and Definitions
Concrete Dimensions

Planning and Organizing:
Structuring tasks/resources and establishing appropriate courses of action for self and others; coordinating the actions of others to achieve goals.

Oral Communication:
Effectively expressing oneself in individual or group situations; conveys thoughts clearly and concisely; includes gestures and other non-verbal behaviors.

Monitoring:
Ensuring that others follow rules, regulations, policies, and procedures.

Abstract Dimensions:
Assertiveness:
The ability and willingness to take action in a positive, firm, yet non-aggressive manner.

Energy:
Possessing a high activity level.

Self-Confidence:
A strong belief in one's own ideas, abilities, values, and competencies.
Appendix F: Minimum Standards
A. Group Discussion-Concrete:

Oral Communication
- appropriate voice level for the setting and the number of people
- used good grammar
- was easily understood; didn't have to repeat himself
- used appropriate non-verbal gestures
- when speaking, maintained eye contact with others
- presented ideas in a logical manner

Planning and Organizing
- made sure the discussion moved along
- initiated discussion and/or proposed a good discussion procedure
- came up with at least one suggestion about how to solve the problem
- evaluated solutions to explore possible roadblocks
- showed evidence of organizing (e.g. took notes, observed the time)
- suggested some sort of follow through on the group's recommendations

Monitoring
- seemed aware of student reaction to the present alcohol policy
- came up with ideas that were consistent with state/federal laws
- tried to find a policy that would be enforceable

B. Group Discussion-Abstract:

Self-Confidence
- made strong declarative statements, rather than tentative ones
- didn't hesitate or stammer while speaking
- appropriate voice level for the setting and the number of people
- appeared relaxed in a group situation
- was able to deal with criticism of his ideas from others
- demonstrated independence in generating ideas; came up with at least one suggestion about how to solve the problem

Assertiveness
- actively sought information from others
- supported his own ideas, but also yielded to others when appropriate
- was not domineering, but did not let others intimidate him
- evaluated solutions to explore possible roadblocks
- criticized ideas of others in a supportive manner
- initiated the discussion and/or proposed a good plan of action

Energy:
- maintained a high level of interest; didn't need prompting from others
- listened attentively
- participated; wasn't silent for long periods of time

C. Counseling Session—Concrete:

Oral Communication
- appropriate voice level for the setting and the number of people
- used good grammar
- was easily understood; didn't have to repeat himself
- used appropriate non-verbal gestures
- when speaking, maintained eye contact with others
- presented ideas in a logical manner

Planning and Organizing
- made an initial statement of the purpose at the start of the interview
- asked questions in order to discover what the problem was
- made sure the discussion moved along
- worked with the student to find possible solutions
- tried to set up some kind of follow-through plan
- came up with at least one suggestion on how to solve the problem
Monitoring
- talked with the student about dorm/floor rules and policies
  and/or good hygiene practices
- discussed how the student's actions were affecting others

D. Counseling Session-Abstract

Self-Confidence
- seemed comfortable when discussing the problem; self-assured
- used appropriate voice level
- did not retreat or take back his points/suggestions when the
  student became upset

Assertiveness
- told the student that he needed to improve his hygiene
- asked questions to identify the problem
- was persuasive but not coercive
- got commitment from the student to work toward a solution
- made constructive criticisms
- was firm, but fair
- instituted steps to make sure that the student improved

Energy:
- maintained a high level of interest; didn't need prompting from
  others
- listened attentively
- participated; wasn't silent for long periods of time
- began with a warm and friendly greeting

E. Roommate Conflict-Concrete:

Oral Communication
- appropriate voice level for the setting and the number of
  people
- used good grammar
- was easily understood; didn't have to repeat himself
- used appropriate non-verbal gestures
- when speaking, maintained eye contact with others
- presented ideas in a logical manner
Planning and Organizing
- made an initial statement of purpose
- elicited both roommates' opinions about the problem
- worked with the student to find possible solutions
- worked with the roommates on a course of action
- tried to set up a follow-through plan
- used information provided by the students when developing the plan of action

Monitoring
- talked about hall policy concerning a room change
- discussed solutions which were consistent with hall policy

F. Roommate Conflict—Abstract:

Self-Confidence
- made strong declarative statements, rather than tentative ones
- didn't hesitate or stammer while speaking
- appropriate voice level for the setting and the number of people
- appeared relaxed in the situation
- was able to deal with criticism of his ideas from others

Assertiveness
- asked questions to identify the problem
- was persuasive but not coercive
- got commitment from the students to work toward a solution
- made constructive criticisms
- was firm, but fair
- told the students that they must adjust their living habits in order to get along better together

Energy:
- maintained a high level of interest; didn't need prompting from others
- seemed interested in the discussion
- participated; wasn't silent for long periods of time
Appendix 6: Stimulus Materials
Behavioral Observation Form
Counseling Session

SS#

________________________

Date

________________________
Behavioral Classification Form
Counseling Session

SS #

Date

Planning and Organizing ( )

Oral Communication ( )

Monitoring ( )
Appendix H: Training Materials
TRAINING PROGRAM FOR ASSESSORS

OVERVIEW

PUT UP DO NOT DISTURB SIGN

INTRODUCE YOURSELF AND GIVE EACH STUDENT A PACKET

HAVE SUBJECTS SIGN CONSENT FORMS

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

SHOW OVERHEAD 1

An assessment center is not a place. It is method used to make employment decisions, such as the hiring of a job applicant, the promotion of an employee, or training and development of an employee. In an assessment center, the candidates, who are usually called 'assesseees' 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. I appreciate your help as an assessor in this research project and your willingness to
help this research project succeed so that universities will have the best system possible for hiring residence hall student assistants.

Now I will describe the job of student assistant.

SHOW OVERHEAD 2

A 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.

An SA's duties entail helping develop university and residence hall policies, communicating and enforcing residence hall policies and procedures, promoting consideration of individual needs in a group living environment, coordinating educational and social/recreational programs for students, ensuring that students develop an appropriate atmosphere on the floor, counseling students, and promoting security awareness.

Now I will discuss the assessment center in which you will be participating as an assessor.
Assessment Center Process

**SHOW OVERHEAD 3**

The phases of the assessment center in which you will participate

(a) Observation: You will watch three candidates for student assistant positions who will perform in three exercises—a counseling session with a troubled student, a conflict between two roommates, and a discussion among student assistants of the university alcohol policy. You will take detailed notes on what candidates say and do.

You won't observe each candidate in every exercise, but you will observe each candidate in one exercise. Later, you will break into groups and share your observations with other assessors.

(b) Classification: You will classify behaviors you observed and recorded into dimensions of performance, such as oral communication and sensitivity.

(c) Integration Session: You will break into groups of assessors and observers. You will become part of a three person assessor team and will share information on candidates, one candidate at a time. For each candidate, you
will rate each dimension, taking into account the candidates performance in each exercise. Finally, you will rate each dimension across all exercises, taking into account the candidates' overall performance on that dimension.

1. Observation

In the observation stage, you will observe one candidate for a student assistant position who will be discussing a university policy with two other candidates. You will take notes on everything the candidate says and does. This will include both verbal and non-verbal behavior.

Now we will focus on how to accurately observe and record behaviors.

It is very important to make good behavioral observations of others. Good behavioral observations enable you to make accurate evaluations of others and allow you to successfully communicate to the other assessors what the student assistant candidate actually said and did.
SHOW OVERHEAD 4

(a) Good observations:

- **State what a person says or does** (e.g., “He said to the student, ‘I don't have time to discuss your problem’

- **Are specific rather than general** (e.g., “She suggested a follow-up meeting in 10 days)

- **Are descriptive rather than evaluative** (e.g., “The students kept talking when he was trying to start the discussion)

- **Are confirmable by others** (e.g., “He took notes throughout the meeting”)

(b) Poor observations:

- **Make general classification statements** (e.g., “She was sensitive”)

- **Interpret actions** (e.g., “The student was getting on the SA’s nerves”)

- **Impart feelings** (e.g., “He was angry”)

- **Describe underlying personality make-up** (e.g., She is paranoid)

Are there any questions?
A. Practice in identifying good behaviors

Now you are going to practice identifying good and poor behavioral observations. Take out the sheet labelled "Behavior Example Practice Exercise" and determine whether each of the 10 behaviors listed are indicative of a good or a poor behavioral observation.

HOLD UP THE CORRECT FORM. GIVE THEM 3 MINUTES TO FILL OUT THIS FORM. DISCUSS EACH OF THE 15 EXAMPLES AND HAVE THEM INDICATE WHETHER EACH IS A GOOD OR A POOR BEHAVIORAL OBSERVATION. REITERATE WHY EACH BEHAVIOR IS A GOOD OR POOR BEHAVIORAL OBSERVATION.

Are there any questions?

B. Practice in recording behavioral observations

Now, you will practice recording behavioral observations by watching a tape of a group of people. Each candidate represents a different applicant for the SA position. You are to observe and write down both non-verbal and verbal behaviors of the person on the screen. You will be critiqued and will get feedback on your performance. Please remember that making good behavioral observations is one
of the most important things that you will do. Keep in mind that you will have to write very quickly to record the gist of what the person says and does. You need to write down details; write down verbatim quotes if possible, and note the context of things to help interpret what the person says. I will now model the first person to give you an example of how this is done.

Now, please take out your behavioral observation form. It looks like this.

**SHOW THEM A SAMPLE FORM**

You will observe ---, the man on the far left, for two minutes. Please remember to write down everything on this form in as much detail as possible. Start now.

**PLAY TAPE FOR TWO MINUTES. COLLECT OBSERVATION FORMS AT THE CONCLUSION OF THE TAPE**

II. Classification of behavior

Now, I would like to discuss the performance dimensions you will be using in this assessment center when you rate
candidates. Please take out your list of dimensions and your behavior classification form.

HOLD THEM UP.

SHOW OVERHEAD 6

READ THE DEFINITIONS OF THE DIMENSIONS AND ELABORATE ON THEM.

A. Classifying behavior into dimensions

As I mentioned earlier, the second phase of the assessment center process involves classification of behavior into the performance dimensions. Now, I will give you examples of how to do this.

GIVE EXAMPLES OF BEHAVIORS THEY SHOULD HAVE RECORDED IN THE PRACTICE SESSION. POINT TO THE DIMENSIONS ON OVERHEAD 7 AS YOU ARE DOING THIS. TRY TO GET SUBJECTS TO PARTICIPATE IN THIS PHASE

Are there any questions on how to classify behaviors into dimensions?
III. Integration session

Now we will go over the third stage in the assessment center process – the integration session.

A. Steps in the integration session

HOLD UP SHEET. READ OUT STEPS AS THEY FOLLOW ALONG.

B. Use absolute scale when rating candidates

When you are discussing a candidate's performance during the integration session you should never compare his/her performance with that of other candidates. You should use the following rating scale.

SHOW OVERHEAD 9 AND HAVE THEM REFER TO THEIR SHEET CONTAINING THE RATING SCALE

On this scale an acceptable rating (a rating of 3) does not mean average. It means that the candidate has met the minimal requirements for acceptance for the job.
C. Use minimal standards as a guide

You should use the minimal standards for performance when rating a candidate. You will be given a copy of the minimal standards later.

D. Dimension Ratings

Indicate on the classification form where dimension ratings should be written.

E. Across-Exercise Dimension Ratings

Indicate the form on which overall dimension ratings will be made.

Are there any questions at this point?

Conclusion

Now you will have a five minute break. Please go to room number ___ after your break. Please arrive at your assigned room by ___.

GIVE SUBJECTS THEIR ASSIGNED ROOMS AND THE TIME THAT THEY SHOULD ARRIVE.

COLLECT ALL MATERIALS.
Assessment Center

- used to make employment decisions (hiring, promotion, etc.)

- candidates participate in job-related simulations

Those who observe and rate behavior are called 'assessors'
WHAT IS A STUDENT ASSISTANT (SA)?

* An upper level undergraduate who is assigned to a floor section of a residence hall or college

* Is responsible for developing and maintaining an atmosphere which promotes academic, personal, and social growth in the residence hall

* Duties entail:

  – Aiding in development of policies
  – Communicating and enforcing policies and procedures
  – Promoting consideration of individual needs in a group living environment
  – Coordinating educational and social/recreational programs for students
  – Developing an appropriate academic atmosphere on the floor
  – Counseling students
  – Promoting security awareness
PHASES OF THE ASSESSMENT CENTER

1. Observation of behavior
   - Observe one candidate in a group discussion
   - Take detailed notes on what the candidate says and does

2. Classification of behavior
   - Classify behaviors into performance dimension categories

3. Integration session
   - Follow steps in the integration session
   - Rate candidates on a absolute scale; do not compare performance across candidates
   - Use minimal standards as a guide
   - Make within-exercise dimension ratings
   - After one candidate’s performance in all exercises is described, make across-exercise dimension ratings
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
Overhead 8

STEPS IN THE INTEGRATION SESSION

1. The assessor who observed a candidate shares his/her behavioral observations of that candidate on oral communication while the other assessors listen and take notes.

2. Assessors may then ask questions for clarification but discussion is limited.

3. Assessors independently rate the candidate on the dimension.

4. Steps 1–3 are repeated for each dimension.

5. Steps 1–4 are repeated for the same candidate for the other two exercises.

6. The candidate is rated on each dimension across all three exercises.

7. Steps 1–7 are repeated for the other two candidates.
BEHAVIOR EXAMPLE PRACTICE EXERCISE

READ EACH STATEMENT BELOW AND DECIDE WHETHER IT IS A BEHAVIORAL STATEMENT OR 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 THE OBSERVER MADE THE STATEMENT, WOULD YOU KNOW WHAT THE STUDENT ASSISTANT ACTUALLY DID OR SAID? 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.</td>
<td>LOOKED AT HIS WATCH</td>
<td>_______</td>
</tr>
<tr>
<td>2.</td>
<td>SEEMED NERVOUS</td>
<td>_______</td>
</tr>
<tr>
<td>3.</td>
<td>WAS THE LEADER OF THE GROUP</td>
<td>_______</td>
</tr>
<tr>
<td>4.</td>
<td>ASKED THE STUDENT HOW HE WAS FEELING</td>
<td>_______</td>
</tr>
<tr>
<td>5.</td>
<td>TOLD THE STUDENT THAT SHE LOOKED TERRIBLE</td>
<td>_______</td>
</tr>
<tr>
<td>6.</td>
<td>ASSERTIVELY STATED HER OPINION</td>
<td>_______</td>
</tr>
<tr>
<td>7.</td>
<td>SUGGESTED STUDYING SCHEDULES FOR THE ROOMMATES</td>
<td>_______</td>
</tr>
<tr>
<td>8.</td>
<td>HAD GOOD LISTENING SKILLS</td>
<td>_______</td>
</tr>
</tbody>
</table>
9. ADEQUATELY SUMMARIZED THE DISCUSSION

10. LEANED BACK IN CHAIR

11. SAID THAT THE DISCUSSION WAS PURPOSELESS

12. WAS CONSIDERATE OF OTHERS’ FEELINGS

13. CAME UP WITH CREATIVE SOLUTIONS

14. LOOKED AT OTHERS WHILE THEY SPOKE

15. GIGGLED SEVERAL TIMES

GOOD EXAMPLE   POOR EXAMPLE
Behavioral Observation Practice Form

Social Security No. ____________________________