



3 1293 00891 9304

This is to certify that the
thesis entitled
UNDERSTANDING ASSESSMENT CENTER DIMENSION
AND EXERCISE CONSTRUCTS: A CENTER DESIGN
AND EXERCISE DESIGN APPROACH
presented by
Jeffrey Robert Schneider

has been accepted towards fulfillment
of the requirements for
M.A. degree in Psychology

Major professor

Date 10/17/90



PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
204		

MSU is An Affirmative Action/Equal Opportunity Institution

c:\circ\dtedue.pm3-p.1

UNDERSTANDING ASSESSMENT CENTER
DIMENSION AND EXERCISE CONSTRUCTS:
A CENTER DESIGN AND EXERCISE DESIGN APPROACH

Jeffrey Robert Schneider

A THESIS

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

MASTER OF ARTS

Department of Psychology

1990

ABSTRACT

UNDERSTANDING ASSESSMENT CENTER

DIMENSION AND EXERCISE CONSTRUCTS:

A CENTER DESIGN AND EXERCISE DESIGN APPROACH

By

Jeffrey Robert Schneider

Two related avenues of research aimed at understanding assessment center dimension and exercise constructs were pursued in this study. First, the center was designed according to previous researchers' recommendations for designing exercises, dimensions, scoring guidelines and assessor training; for reducing the cognitive load on assessors; and for controlling for rater effects. Confirmatory factor analyses of ratings of 89 high school student assesseees suggest these design interventions did not result in better evidence of convergent and discriminate validity than in previous studies. Second, in an attempt to understand the dominant influence of exercises on assessment center ratings, exercise form and content were proposed as distinct, exercise-based, sources of method variance. The results suggest that exercise form accounts for 16% of method variance whereas the effect of exercise content, defined as cooperative and competitive task designs, is negligible. The implications of these findings for future research and practice are discussed.

Copyright by
JEFFREY ROBERT SCHNEIDER
1990

ACKNOWLEDGEMENTS

I am proud of this thesis, and I wish to express my appreciation to those who contributed to its quality. My deepest appreciation goes to my committee members, Dan Ilgen and Kevin Ford, and especially to my advisor, Neal Schmitt. Neal helped me obtain the sample, sharpened my hypotheses and conclusions, and supported me on many days when I was certain that the whole project would collapse. A big thank you also goes to Paul Stemmer and Bill Brown who helped this project succeed and who taught me the ropes of dealing within the school system. I also have great appreciation for the time and effort of many school administrators, teachers, students, role-players, and assessors, all of whom shall go nameless for confidentiality reasons.

On a more general level, Dave Heine and Rob Silzer deserve much credit for stimulating my interests in assessment centers. I have come a long way since Dave and Rob taught me how to create my first project presentation and discussion exercise. My fellow graduate students at MSU have been an extraordinary source of support and friendship and have made most every day in the office fun and enjoyable.

Above all, thanks, Mom and Dad, for the financial and emotional support that has allowed me to pursue passionately my career and life goals. You have made all of this possible!

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	viii
INTRODUCTION.....	1
Assessment Center Validity.....	4
Assessment Centers and Criterion-Related Validity.....	4
Assessment Centers and Content Validity.....	5
Construct Validity.....	7
Assessment Centers and Construct Validity.....	11
Summary and Conclusions.....	15
Exercise Effects--Explanations and Recommendations.....	17
Dimension/Exercise Match.....	17
Assessor Training.....	17
Dimension Definitions and Scoring System.....	19
Cognitive Overload.....	21
Number of Dimensions.....	21
Independence of Dimensions.....	22
Checklist Scoring Systems.....	23
Rater Effects.....	24
Conclusions--Design Interventions.....	25
Method Variance as a Function of Exercise Design.....	26
Behavior/Exercise Confounds.....	26
Exercise Form and Content.....	27

	Page
Exercise Forms--The Leaderless Group Discussion and	
Role Play Exercise.....	30
Exercise Content--Cooperation and Competition.....	31
Exercise Form and Content--Conclusions.....	32
Hypotheses.....	33
METHOD.....	36
Description and Sample.....	36
Dimensions.....	37
Exercises.....	39
Grant Allocation Exercise (competitive-group discussion)...	41
Team Manufacturing Exercise (cooperative-group discussion).	41
Customer Service Exercise (cooperative-role play).....	42
Team Selection Exercise (competitive-role play).....	43
Pre-test of Exercises.....	44
Role Players and Training.....	45
Assessors and Assessor Training.....	46
Rating Procedure and Scoring Guidelines.....	48
Design Issues.....	50
Statistical Analyses.....	50
RESULTS.....	55
Rating Distributions and Inter-rater Reliability.....	55
Multitrait-Multimethod Matrix.....	58

	Page
Convergent and Discriminant Validity.....	60
Exercise Form and Content.....	69
Estimates of Variance Attributable to Facets in the Design....	75
DISCUSSION.....	77
APPENDIX A: Correspondence to Participants.....	84
APPENDIX B: Dimensions and Definitions.....	86
APPENDIX C: Exercises.....	89
APPENDIX D: Role Player Training Materials.....	114
APPENDIX E: Assessor Training Materials.....	119
APPENDIX F: Scoring Guidelines.....	121
LIST OF REFERENCES.....	134

LIST OF TABLES

Table	Page
1 Summary of Multitrait-Multimethod Mean Correlations from Previous Construct Validity Studies.....	12
2 Hierarchically Nested Models for Hypothesis Testing.....	34
3 Dimensions.....	38
4 Questions Regarding Previous Contact with Students.....	47
5 Model IVD: Three Trait Factors (Oblique); Two Content Factors (Oblique).....	51
6 Dimension Means, Standard Deviations, and Alphas by Exercise..	56
7 Means and Standard Deviations for Previous Contact and for Dimension Ratings by Rater.....	57
8 Multitrait-Multimethod Correlations.....	59
9 Convergent and Discriminant Validity: Column A Indices of Fit.....	61
10 LISREL Estimates for Model IIA: One Trait Factor.....	62
11 LISREL Estimates for Model IIIA: Three Trait Factors (Orthogonal).....	63
12 Convergent and Discriminant Validity: Column B Indices of Fit.....	65
13 LISREL Estimates for Model IIB: One Trait Factor; Four Exercise Factors (Orthogonal).....	66

Table	Page
14 LISREL Estimates for Model IIIB: Three Trait Factors (Orthogonal); Four Exercise Factors (Orthogonal).....	67
15 LISREL Estimates for Model IVB: Three Trait Factors (Oblique); Four Exercise Factors (Orthogonal).....	68
16 Exercise Form and Content: Row I Indices of Fit.....	70
17 LISREL Estimates for Model IB: Zero Trait Factors; Four Exercises Factors (Orthogonal).....	71
18 LISREL Estimates for Model IC: Zero Trait Factors; Four Exercise Factors (Oblique).....	72
19 LISREL Estimates for Model ID: Zero Trait Factors; Two Form Factors (Oblique); Two Content Factors (Oblique)...	73
20 Variance Attributable to Facets in the Design.....	76

INTRODUCTION

The use of assessment centers has grown rapidly since their predictive validity for managerial jobs was first demonstrated (Bray & Grant, 1966). Assessment centers are now used for selection, placement, early identification of management potential, and employee development in large and small, manufacturing, government, educational, military, and service organizations (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Klimoski & Brickner, 1987). Given their widespread use, assessment center results are used in important decisions regarding the careers of thousands of people annually (Thornton & Byham, 1982).

Assessment centers represent a work sample or simulation approach to employment testing. *Assesseees*, the participants being evaluated in the assessment center, take part in a series of exercises in which they are required to perform behavioral tasks that represent actual or simulated job behaviors. *Assessors*, specially trained raters, observe assesseees during the exercises and assign numerical ratings to the assesseees' performance on a number of job-relevant behavioral *dimensions*. Dimensions are defined by a series of representative behaviors in order that the dimensions can be observed and rated. In theory, assessment center dimension ratings represent stable individual characteristics or traits. The dimension ratings, then, serve as quantitative indices for personnel decision making.

Interest in assessment center research has grown along with the growth in their application. In addition to continued validity related research, a number of studies have examined process related hypotheses to explain how assessment centers work (Russell, 1985; Sackett & Hakel, 1979; Sackett & Wilson, 1982). Over the past decade, researchers have examined assessment center ratings for evidence of construct validity (Archambeau, 1979; Bycio, Alvares, and Hahn, 1987; Neidig, Martin, & Yates, 1979; Robertson, Gratton, & Sharpley, 1987; Sackett & Dreher, 1982; Silverman, Dalessio, Woods, & Johnson, 1986; Turnage & Muchinsky, 1982). The findings of these studies have been remarkably consistent and robust; ratings across dimensions within an exercise correlate higher than ratings across exercises for particular dimensions.

These findings are troublesome for assessment centers. Assessment centers are thought to measure "enduring characteristics" of assessees that influence their behavior in various settings (Thornton & Byham, 1982, p. 7). Yet, there is little evidence that assessment centers produce dimension scores that serve as valid representations of constructs (Sackett & Dreher, 1982) or that assessment centers identify consistent behavioral characteristics of assessees (Guion, 1987). In short, the empirical findings run contrary to the conceptualization of assessment centers.

These findings have also shaken the confidence in assessment centers. For instance, Guion (1987) wrote relative to the inability of assessment centers to find evidence of behavioral consistency, "I do, however, see a growing skepticism among those less fully involved in assessment centers" (p. 204). Reporting the comments of one of their reviewers, Klimoski and Brickner (1987) wrote that research establishing

assessment centers as measures of valid constructs would eliminate "the nagging feeling that when assessment centers actually work, one has just been through some sort of 'voodoo rite'" (p. 256). These comments reinforce the importance of continuing to pursue the question of what assessment centers measure. Further construct validity research may be able to determine whether the skepticism and nagging feelings about assessment centers are well founded.

This study is an extension of previous assessment center construct validity studies; its purpose is twofold. First, previous construct validity studies offer a number of recommendations for designing centers, all of which aim at improving construct validity. Design interventions based on these recommendations guided the design of the center in this study and may very well be sufficient for finding better evidence of construct validity. Second, this study pursues specific hypotheses regarding the effects of the design of exercise tasks on assessee performance. While many assessment center studies have shown that method (exercise) variance dominates the factor structure of assessment centers, no previous studies have examined how exercise design contributes to the large method variance factor. Specifically, I propose that exercise form and content represent two dimensions along which exercise designs can be distinguished. I explore whether form and content represent a reasonable system for thinking about exercise design and to what extent method variance can be explained by exercise form and content differences.

The literature review section unfolds as follows. First, I describe criterion, content, and construct-related defenses of assessment center validity and highlight the importance of evidence of construct validity

to the conceptualization and use of assessment centers. Second, I will review the assessment center literature, focusing specifically on the design recommendations that researchers have offered for improving construct validity. Third, I present the conceptual foundation for a system of classifying exercise designs according to form and content.

Assessment Center Validity

Assessment centers and criterion-related validity

Assessment centers have been widely used on the basis of criterion-related validity. The criterion-related validity evidence is summarized in meta-analyses of assessment center validation studies. In a meta-analysis involving 107 validity coefficients by Gaugler, Rosenthal, Thornton, and Bentson (1987), the corrected mean coefficient across all types of criteria was .37. In terms of the specific criteria used, validity coefficients were significant for predicting job performance, managerial potential, performance in training, and career advancement. The highest corrected coefficient, .53, was found when potential was the criterion. In a meta-analysis study involving different types of predictors by Schmitt, Gooding, Noe, and Kirsch (1984), assessment center validity coefficients were higher, overall, than many traditional predictors such as measures of general mental ability, personality, special aptitude, and biodata. Based on criterion-related validity, then, assessment centers seem to be a statistically significant predictor of a wide variety of criteria of managerial success and seem to demonstrate higher validity coefficients than other types of predictor measures when averaged across many validation studies.

Assessment centers and content validity

The results of criterion-related studies have been consistently favorable regarding the predictive validity of assessment centers, yet local considerations such as small sample sizes do not always make a criterion-related validity study feasible for all applications (Schmidt & Hunter, 1980). As a result, a content-oriented defense is often adopted as the most prudent strategy for assessment center validation (Sackett, 1987).

Diverse opinions can be found in the published literature about the appropriateness of a content-related validity defense for assessment centers. On one hand, assessment centers are commonly viewed as work samples and a case is made for content validity as long as the job content domain is appropriately reflected in the assessment center. For example, Norton (1977) recommends that assessment centers can be used on the basis of content validity to select among candidates for a position which is substantially managerial in content, even in the absence of an empirical validation study. He suggests that the content validity of an assessment center is based on the resemblance of the assessment center as a whole to the managerial job as a whole.

Other researchers have applied more stringent parameters in determining whether a content validation strategy is appropriate for assessment centers. First, a distinction between predictor measures as signs or samples is commonly made. Following Wernimont and Campbell (1968), Sackett and Dreher (Dreher & Sackett, 1981; Sackett & Dreher, 1982) suggest that a content validation strategy is appropriate when the predictor is used as a work sample to measure the current level of

performance of the candidate. If a predictor is used as a sign to predict performance on work tasks not previously performed by candidates or performance after training, a content validation strategy alone may not be sufficient. Following Sackett and Dreher's logic, content validation may not be sufficient for assessment center applications where the primary use of the data is the prediction of long term performance, commonly called potential or managerial potential.

Second, the nature of the dimensions to be measured also seems to dictate whether a content or construct strategy is more appropriate. A number of researchers and professional publications suggest that a content oriented strategy is appropriate when dimensions can be measured through direct observation of behaviors (Dreher & Sackett, 1981; Sackett & Dreher, 1982; Society for Industrial and Organizational Psychology's Principles for the Validation and Use of Personnel Selection Procedures, 1987; Tenopyr, 1977). Byham (1980) argues further that a content-oriented strategy is also appropriate when the measured dimensions are clusters of observable work behaviors, even though dimension labels may be similar to personality or aptitude constructs. His position also seems to be backed by the authors of SIOP Principles who suggest that a content-oriented strategy may be appropriate when ability dimensions represent clusters of knowledges or skills relative to task performance. A content-oriented strategy is not recommended when higher level mental or psychological constructs such as intelligence, common sense, judgment, leadership, and spatial ability are measured (Uniform Guidelines on Selection Procedures, 1978).

Despite the various positions, it is reasonable to conclude that a content validation strategy for assessment centers is not always

appropriate; the use of the results and the nature of the dimensions measured are important considerations. As the ratings are used to make predictions about managers' future behavior or as the dimensions measured represent higher level mental or psychological processes, a construct validation strategy may be warranted.

Construct validity

This section departs briefly from the discussion of assessment center validity to review the conceptualization of construct validity. This departure is warranted since construct validity is a vital part of this study, and its proof involves complex conditions. A more specific discussion of assessment centers and construct validity follows this section.

In an early conceptualization of construct validity, Cronbach and Meehl (1955) suggested that a construct is defined implicitly by a network of associations among constructs and observable variables. Construct validation is pursued through examining whether predicted relations among constructs are witnessed in observable variables. SIOP Principles offers a definition representing the current usage of the term construct:

"Knowing whether a construct is measured validly requires, if not a theory, at least some fairly well articulated ideas about what is being measured, what a measure of the construct should reasonably expect to be related to, and, perhaps more importantly, what it should not be related to" (p. 25).

Campbell and Fiske (1959) offer the multitrait-multimethod (MTMM) matrix as a framework for examining the relationships among observable variables for evidence of construct validity. The multitrait-multimethod matrix is a correlation matrix among traits measured by different methods. One condition for construct validity is convergent

validity. Evidence of convergent validity is determined by the extent to which ratings of the same trait correlate (converge) across independent methods of measurement. A second condition for construct validity is discriminant validity. Measures of one construct should be uncorrelated with measures of constructs that are conceptually different.

More specifically, Campbell and Fiske list four conditions for construct validity. The first condition is evidence of convergent validity, and the second, third, and fourth conditions are evidence of discriminant validity. First, entries in the validity diagonals (monotrait-heteromethod correlations) should be significantly different from zero. Second, the validity value for a variable (monotrait-heteromethod correlations) should be higher than the correlations obtained between that variable and any other variable having neither trait nor method in common (heterotrait-heteromethod correlations). Third, a measurement of a trait should correlate higher with an independent method of measuring the same trait (monotrait-heteromethod) than with correlations of different traits measured by the same method (heterotrait-monomethod). Fourth, trait relationships should be consistent throughout the matrix; the same pattern of trait inter-correlations should be shown in monomethod and heteromethod blocks.

Campbell and Fiske also emphasize the importance of the independence of methods. Following Cronbach and Meehl (1955), they suggest that the strongest evidence of construct validity comes from numerous and diverse methods of measuring the same construct.

The MIMM matrix has a number of deficiencies for construct validity research. One problem is that judgments of the degree of convergent and

discriminant validity are based on comparison of an often large number of correlation coefficients and subjective assessments of their size relative to other correlations in the matrix. Widaman (1985) cites two additional problems. First, precise estimates of the amounts of trait-related or method-related variance for specific measures are not obtainable, even though these estimates would be quite useful. Second, judgments of convergent and discriminate validity using the MIMM procedure are based on observed correlations among measured variables, with no consideration of how the reliability differences among measures distort these correlations. Other procedures for analyzing MIMM data which overcome these problems have emerged.

First, Analysis of Variance (ANOVA) has frequently been used following the procedure proposed by Guilford (1954) and refined by Stanley (1961), Boruch, Larkin, Wolins, and Mackinney (1970), Boruch and Wolins (1970), and Kavanagh, Mackinney, and Wolins (1971). In this model, the sums of squares are calculated based on average correlations from the MIMM matrix, and mean squares are expressed as a function of the average variances and covariances among the variables in the matrix. Significant tests can also be performed. A significant effect for person is hypothesized to indicate the degree of convergence or consistency in judgments about individuals over exercises and traits; a significant effect for person X trait is hypothesized to be indicative of discriminant validity; and a significant effect for person X situation is hypothesized to be indicative of the degree of method variance.

Since the ANOVA procedure allows for overall significance tests, the procedure overcomes one of the problems in the Campbell and Fiske

procedure. The ANOVA procedure estimates the effects for person, trait and method across all persons, traits and methods and like MIM, does not allow for estimates of variance for specific traits or methods. For example, the person effect (convergent validity) is tested across traits and methods. Thus, a significant effect for person, the proposed test of convergent validity in the ANOVA procedure, tests the extent to which the measure rank orders candidates on one general trait. This procedure does not test the more stringent case of convergent validity based on the correlations of single dimension ratings (for a measure with multiple dimensions) across multiple methods. Lastly, ANOVA also uses observed correlations without attention to reliability.

Factor analysis can also be used to summarize and test construct validity hypotheses. Using traditional factor analysis in an exploratory mode, factors should represent a clustering of dimension scores rather than exercise scores (Smith 1976). More recently, confirmatory factor analysis has been proposed (Schmitt & Stults, 1986, Widaman, 1985). Widaman (1985) has outlined a procedure for testing a series of hierarchically nested models of various trait and method relationships to test for evidence of convergent and discriminant validity. The CFA procedure allows for the comparison of a series of models on the basis of their fit and parsimony and allows for a statistical judgment of which model of trait and method relationships best fits the data, overall. The CFA procedure addresses the problems with the MIM procedure. Statistical judgments of overall fit are possible; variances associated with specific traits and methods can be separately estimated; and reliabilities for observed variables are used in calculating the relationships between latent constructs.

Assessment centers and construct validity

Assessment center researchers have employed all of the methods above in testing for construct validity. Simply stated, evidence for construct validity in assessment centers exists when ratings of one dimension agree more closely with ratings of the same dimension in other exercises than with ratings of other dimensions in the same exercise (Sackett and Dreher, 1982). The exact opposite has been found in all of the published studies; heterotrait-monomethod correlations have been larger than monotrait-heteromethod correlations. Table 1 reports these values allowing for comparison of the findings across these studies. I will summarize the findings of specific studies in detail in the paragraphs that follow.

Sackett and Dreher (1982) used factor analysis to test the factor structure of assessment center data from three organizations: a multinational firm selecting upper level managers (N = 86); a state civil service commission selecting upper level managers (N = 311); a retailer selecting store managers (N = 162). Using the factor analysis procedure, they hypothesized that factors corresponding to dimensions and not exercises should be found if assessment centers are working as they are conceptualized. The opposite result was found in all three cases; factor loadings corresponded with exercises and not with dimensions. As for convergent validity, the level of agreement among various ratings of a dimension was near zero for two of the three data sets (see Table 1). More convergence was demonstrated in the third data set, yet an interpretation that this data demonstrated better convergent validity was tempered by evidence of pervasive halo error in all the ratings (the mean heterotrait-heteromethod correlation was .45). The

Table 1

Summary of Multitrait-Multimethod Mean Correlations from Previous
Construct Validity Studies

Study	N	HTMM	MTMM	HTHM
Archambeau (1979)	29	.47	.33	NA
Neidig et al. (1979)	260	NA	NA	NA
Sackett & Dreher (1982)	86	.638	.074	.06
	311	.395	.109	.07
	162	.648	.508	.45
Turnage & Muchinsky	1028	.51-.90	.18-.70	NA
(1982)	1028	.52-.90	.20-.69	NA
Silverman et al. (1986)	45(w/d)	.65	.54	.44
	45(w/e)	.68	.37	.31
Robertson et al. (1987)	41	.64	.28	NA
	48	.66	.26	NA
	84	.60	.23	NA
	49	.49	.11	NA
Bycio et al. (1987)	1170	.75	.36	NA

Note. The following mean correlation values are reported: heterotrait monomethod (HTMM); monotrait heteromethod (MTMM); heterotrait heteromethod (HTHM). Values not available (NA); Ranges are reported where means were unavailable. (w/d) and (w/e) distinguish within dimension and within exercise consensus discussions.

relatively high level of agreement among the dimension ratings within exercises was evidence of a lack of discriminant validity in all three data sets. In their discussion of the results, Sackett and Dreher raised strong concerns about the lack of evidence that assessment center techniques generate scores representing complex constructs such as leadership and decision making skills and offered perhaps the strongest statement of caution against the use of assessment centers on the basis of a content validity defense alone. In a later work, based on the same findings, Sackett and Dreher (1984) advocated a reconceptualization of assessment centers where assessees would be rated on exercises rather than on dimensions.

Turnage and Muchinsky (1982) used the ANOVA procedure to examine convergent and discriminant validity for two matched samples from the assessment center from a large manufacturing firm (N total = 2056). All three effects, person, person X trait, person X situation, were significant in both samples, yet person and person X situation effects were considerably more robust. Based on the significant effect for person, they concluded assessors generally agree to a great extent on the ordering of individuals on a global basis (i.e. as if ratings were based on one general trait factor). The findings suggest a lack of discriminant validity across traits, however. Some caution is warranted in interpreting these results since a large halo factor was present across all ratings, much like what was found in the third data set in the Sackett and Dreher (1982) study. Pervasive halo in ratings tends to inflate estimates of convergent validity when using the ANOVA procedure.

Somewhat better evidence for dimension factors was found by Silverman, et al. (1986) using a retail department manager assessment

center (N total = 90). They hypothesized that different relationships between dimensions and exercises may be observed when ratings are made independently by assessors after each exercise (within exercise method, N = 45) and when ratings are made on each dimension after considering the data from all the exercises (within dimension method, N = 45). They used the MTMM matrix, ANOVA, and factor analysis to test for convergent and discriminant validity. Based on the MTMM and the ANOVA procedures, they found evidence for convergent validity with both samples, yet evidence for convergent validity was significantly better (t test at $p < .01$) for the within dimension method. The conditions for discriminant validity were not supported in either sample, yet the within dimension method showed slightly better evidence of discriminant validity. Factor analysis for the within exercise condition suggested simple factor loadings represented by exercises, confirming the Sackett and Dreher (1982) findings. The results for the within dimensions method revealed statistically significant loadings on more than one factor, offering more support for assessment centers as a measure of individual difference dimensions than previous findings. Nevertheless, in the within dimension method, the exercise loadings were considerably higher than the ability loadings suggesting a relatively strong exercise effect and a lack of discriminant validity. They concluded that the design of centers, and the design of the rating process in particular, with the aim of improving construct validity may be a productive research direction.

Robertson, et al. (1987) used data sets from four centers assessing candidates for bank branch managers (N = 41), undergraduate recruits for positions in an oil company (N = 48) and in a foodstuff manufacturing

and distributing company (N = 84), and branch managers in a retail chain (N = 49). Using multitrait-multimethod comparisons and factor analysis, their findings were consistent with the findings of previous studies. Like Sackett and Dreher (1982), they advocated a reconceptualization of assessment centers as a series of work sample measures.

Bycio, et al. (1987) used confirmatory factor analysis on data from a manufacturing supervisor assessment center (N = 1170) to test a series of ability and exercise factor models. Three models were tested: (a.) eight ability factors, five exercise factors This model was based on the center designer's intent that all abilities be measured by all exercises. The model was eventually collapsed into five ability factors and five exercise factors because the eight and five model produced estimated ability correlations greater than 1.0. (b.) one ability factor, five exercise factors This model was based on previous empirical research suggesting that assessors made ratings on the basis of one general factor and on the basis of previous research finding exercise factors. (c.) zero ability factors, five exercise factors This model tested whether ratings reflected exercise performance only. While there were small differences in the indices of fit used to judge the models, the authors could not unambiguously choose which model best fit the data. Their overall conclusions were consistent with previous studies, however. The results suggested that ratings were largely situation specific and that assessors did not distinguish among the eight ability dimensions (a lack of discriminant validity).

Summary and conclusions

It is reasonable to consider what level of concern is appropriate in light of the lack of construct validity findings. A range of opinions

about the lack of construct validity in assessment centers have been expressed in the published literature. Some researchers are not surprised by the findings (Neidig & Neidig, 1984; Zedeck & Cascio, 1986) whereas others express more strong cautionary concerns (Dreher & Sackett, 1981; Sackett & Dreher, 1982; 1984; Gorham, 1978). Assessment center users driven by only a legal or practical perspective will likely continue to use assessment centers based on a content-oriented validation defense. Despite the seven published studies failing to demonstrate construct validity, I know of no statistics suggesting that the use of assessment centers is declining. There is a segment of professionals, then, that are not concerned with addressing this issue or respond to the issue by suggesting that construct validity is not essential to the application of assessment centers.

Theoretically, a test should be able to meet the conditions outlined for criterion-related, content, and construct validity. Thus, it is at least theoretically inconsistent that a test which is scored on the basis of construct-like dimensions ratings would be job related but that the ratings of those dimensions would have no conceptual meaning or relevance. The practitioner should also be swayed by this position since it is also practically important that assessment centers work according to their conceptualization.

Regardless of the position, all parties seem to agree upon the need for more research aimed at understanding why assessment centers are not working according to design. As Sackett and Dreher (1984) have suggested, it is important to merge the conceptualization of assessment centers with the empirical findings. Further research can improve the understanding of how assessment centers work, and thus, the

conceptualization of assessment centers may be able to be matched more closely with the empirical findings.

Exercise Effects--Explanations and Recommendations

Each of the construct validity studies mentioned above attempt to explain their findings, and a number of these explanations can be translated into prescriptions for designing a center. The authors seem to be suggesting that a correctly designed and executed center should produce ratings with a factor structure that agrees with the conceptualization of assessment centers. These explanations represent a good place to start in developing hypotheses about why assessment centers do not work as predicted. In this section, I review all of the design recommendations offered and conclude with a list of design interventions that may improve the probability of finding convergent and discriminant validity.

Dimension/exercise match

It is important that exercises provide ample opportunity to observe the dimensions. If assessors are required to rate dimensions that are unrelated to the behaviors observed in an exercise or to make ratings based on too few behaviors, it is likely that assessors will rely on their overall impression of a candidate. This, in turn, may contribute to the halo and lack of discriminant validity in assessment center ratings. All dimensions are not commonly rated in all exercises in the administration of a particular center, since it is recognized that certain exercises are designed to elicit certain types of behavior more than others. This is not always the case, however. For example, exercises were completely crossed with dimensions in the data set used

by Turnage and Muchinsky (1982). Turnage and Muchinsky note that different dimensions may have different levels of relevance depending on the exercise and that assessors would likely have less rational basis for rating less relevant dimensions. This occurrence likely creates an exercise effect. Thus, it is imperative that exercises and dimensions are matched with a high likelihood of observing relevant behaviors with relatively equal frequency across different exercise conditions.

Assessor training

Assessor training ensures that assessors do their job correctly. If assessors are poorly trained, the center is not likely to work according to design. At minimum, assessor training should provide assessors thorough knowledge and understanding of the assessment techniques and dimensions used, observation and recording skills, evaluation and rating procedures, and organizational practices and uses of the data (Task Force on Assessment Center Standards, 1980).

Turnage and Muchinsky (1982) and Klimoski and Brickner (1987) suggest that assessor training may improve the construct validity of centers, particularly discriminant validity, if assessors are trained to discriminate among dimensions and avoid halo error. While this topic has not been explored in assessment center research, research in performance appraisal has frequently explored the efficacy of training designed to reduce rater errors. Favorably, a number of studies suggest that rater training can reduce halo (Bernadin, 1978; Bernadin & Walter, 1977; Borman, 1975; Ivancevich, 1979; Latham, Wexley, & Pursell, 1975). On balance, some evidence suggests that rater training with an emphasis on avoiding rater errors may result in sacrifices in rating accuracy on behalf of reduced halo and leniency error (Bernadin & Pence, 1980).

Dimension definitions and scoring system

It is also important that assessors clearly understand the dimensions. While this can be achieved to some extent through assessor training, the dimension definitions and the use of anchors and scoring guidelines can clarify the assessor task. First, it is important that dimensions are observable; ratings made on less observable dimensions require inferential leaps on the part of assessors (Sackett & Dreher, 1982) and introduce greater error in measurement. In fact, research in personality suggests that evidence for behavioral consistency is more likely to be found when dimensions are highly observable (Kenrick and Stringfield, 1980). Second, as Turnage and Muchinsky (1982) suggest, exercise scoring guidelines and procedures such as behavioral description observation forms or BARS may improve the potential for finding discriminant validity. Research in performance appraisal has explored the effects of structured formats into the rating process. First, research has explored whether the use of rating scales improves convergent and discriminant validity. Two studies examining construct validity in performance ratings made with behavior expectation scales (Smith & Kendall, 1963) serve as an example for this study. Zedeck and Baker (1972) used head nurse and supervisor ratings to examine the construct validity of the performance ratings of 98 nurses. The correlations in the validity diagonals differed significantly from zero suggesting moderate evidence of construct validity, yet none of Campbell and Fiske's (1959) criteria for discriminant validity were met. Similarly, Dickenson and Tice (1973) examined construct validity of fire fighter performance ratings and also found moderate convergent validity but little evidence of discriminant validity. Campbell, Dunnette,

Arvey, and Hellervik (1973) examined behavior anchored scale performance ratings of 537 department managers for evidence of construct validity. Campbell and Fiske's criterion for convergent validity was met; the first condition for discriminant validity was met in 136 of 144 comparisons; yet, only 16 of 72 comparisons met the second condition for discriminant validity.

Second, a number of studies have compared behaviorally anchored rating scales with other types of formats (e.g. scales without behavioral anchors) and compared them against a number of criteria (halo, leniency, etc.). Yet, differences in effect due to type of format have been trivial (e.g. Borman & Dunnette, 1975; and Burnaska and Hollmann, 1974).

The implications of the performance appraisal findings regarding scoring guidelines is not conclusive, making it likely that the use of behaviorally anchored scales alone is not apt to ensure that evidence of both convergent and discriminant validity will be found. Landy and Farr's (1980) summary of research on rating scales does provide some guidance about what features of rating scales may be helpful to assessment centers, however. First, they suggest that some form of graphic rating scale represents an improvement over more arbitrary forms of judgments. Second, they note some advantage in using behavioral anchors rather than simple numerical or adjectival anchors, particularly in the absence of good dimension definitions. Summarizing the first two points, scoring guidelines which clarify the observation and rating responsibilities of assessees can contribute to a more effective procedure, yet the specific format of a rating system (i.e. BARS vs. BOS) does not seem to matter as long as the format is clear and

understandable. Third, they highlight the importance of using rigorous item selection and anchoring procedures in developing rating scales.

In conclusion, the use of a rigorously developed and clearly defined and anchored scoring system may not completely resolve the construct validity issues, yet it seems to be the best solution for accurate and reliable measurement available at this time. Assessment center dimensions should be well defined with examples of observable, relevant, behaviors and scoring guidelines should provide good and poor behavioral examples upon which assessors can base exercise ratings.

Cognitive overload

Number of dimensions. Assessment centers require assessors to observe, process, and evaluate an enormous amount of behavior. The cognitive overload created by these conditions has been postulated as a potential source of error in assessment center ratings and may explain why assessment centers do not seem to work according to design (Sackett & Dreher, 1982; Silverman et al. 1986; Bycio et al., 1987) This explanation seems consistent with Sackett and Hakel's (1979) and Slovic and Lichtenstein's (1971) findings that assessors tend to use only a few dimensions to arrive at overall ratings even when they are trained to use many different dimensions; assessors may resort to a more economical list of dimensions to deal with overload. Klimoski and Brickner (1987) suggest that limiting the number of dimensions to be assessed may improve the potential for finding construct validity.

Gaugler and Thornton (1989) have examined the effects of using different numbers of dimensions. Assessor subjects using fewer dimensions classified and rated dimensions more accurately than subjects using a larger number of dimensions. Thus, some accuracy seems to be

gained with a more economical list of dimensions. They also examined convergent and discriminant validity, yet some caution is necessary in interpreting the construct validity results, since assessee behavior was more homogeneous than is typical due to the experimental manipulation. Assessee performance was manipulated so that assessees were either good, moderate, or poor candidates across all dimensions, thereby creating high interdimension correlations within an exercise. The data was adjusted statistically in order to proceed with the convergent and discriminant validity analyses. Using the ANOVA procedure, they reported evidence for convergent validity across dimensions and exercises for all conditions with better convergence corresponding to fewer numbers of dimensions. Also, similar to previous studies, halo was pervasive across all dimensions and exercises as evidenced by mean heterotrait-heteromethod correlations of .83, .77, and .69 respectively for the three, six, and nine dimension conditions. Again, overall halo of this magnitude tends to inflate the estimates of convergent validity. They did not find evidence for discriminant validity in any of the conditions.

Independence of dimensions. The extent to which assessment center dimensions represent independent clusters of behavior has bearing on the probability of finding discriminant validity. Some intercorrelations between dimensions is expected since dimensions are not designed to be orthogonal (Sackett & Dreher, 1982), yet as assessors are required to rate a large number of redundant dimensions, they likely will have difficulty classifying and evaluating behaviors (Gaugler & Thornton, 1989). Using classes of dimensions that have been shown to be uncorrelated in past studies may improve the probability of finding

discriminant validity.

Previous factor analysis studies in assessment centers revealed some consistency in the kinds of factors underlying assessment center ratings. In a study of assessor ratings of 17 dimensions, Sackett and Hakel (1979) found that virtually all assessors used two common factors: leadership and organizing; and planning and decision making. Schmitt (1977) found three factors from 17 dimensions: administrative skills, interpersonal skills, and activity level. Russell (1985) found two assessment center factors, interpersonal skills and problem solving skills. Archambeau (1979) found two factors which he labelled outcome orientation (leadership, control & follow-up, organizing and planning, decision making, and decisiveness) and process orientation (perceptual and analytical, interpersonal, flexibility, oral communication, and written communication). Common factors to all of these studies seem to be administrative and decision making skills, interpersonal skills, and activity level. A center selecting dimensions reflecting similar factor patterns should have an increased chance of finding evidence of discriminant validity.

Checklist scoring systems. An additional point involves the use of scoring guidelines and the resulting effect on the decision making load placed on assessors. In addition to reducing the number of dimensions and making dimension categories more distinct, the cognitive load on assessors can be reduced by using certain kinds of scoring formats. In more traditional scoring systems, assessors take extensive notes during exercises and categorize and rate behavioral incidents from the notes immediately after the exercise or much later during a consensus discussion. Checklist scoring systems where assessors check behaviors

previously categorized according to dimensions by subject matter experts may not only reduce the load on assessors by eliminating the note taking and categorization task but also may eliminate errors in categorizing behaviors that likely lead to halo or a lack of discriminate validity.

In summary, reducing the cognitive load on assessors by reducing the number of dimensions to a few, distinct dimensions and by using checklist scoring formats may resolve some problematic issues. First, discriminant validity is more likely to be found due to a reduced level of redundancy among dimensions. Second, the use of fewer dimensions is apt to reduce the cognitive demands on assessors. Third, a checklist scoring system may not only reduce load but may also eliminate categorization errors that result in halo.

Rater effects

Rater effects are a commonly offered explanation of the findings (Robertson, et. al., 1987; Sackett & Dreher, 1982). First, a lack of inter-rater reliability may be a source for the lack of consistency in monotrait-heteromethod correlations. For example, Sackett and Dreher (1982) note that an exercise effect could be the result of different rating tendencies for different assessors (i.e. unreliability), since it is common for different, single, assessors to observe different exercises. Previous research, however, suggests that inter-rater reliability is at an acceptable level in assessment centers. For example, Schmitt (1977) reports inter-rater alphas ranging from .77 to .97 before discussion and from .92 to .98 after discussion. It should be noted that Schmitt's reliabilities are calculated on ratings made by multiple assessors based on the verbal report of a lone observer of the exercise. Borman (1982) found inter-rater reliabilities ranging from

.44 to .92 with a median reliability of .76 in a center where assessor pairs observed and independently rated performance in specific exercises. These findings have led some researchers to conclude that inter-rater unreliability is not the sole or major contributor to the failure to demonstrate construct validity (Bycio, et al. 1987; Sackett & Dreher, 1982). Second, an exercise effect could result if particular assessors are nested within exercises; differences in ratings across exercises may reflect different assessor tendencies. This effect can be controlled by having assessors rate all exercises such that assessors and exercises are uncorrelated (Bycio, et al. 1987).

Thus, two steps can be taken to ensure that method variance is not due to an assessor effect. First, multiple assessors can be used and inter-rater reliabilities can be estimated. Second, assessors and exercises should be uncorrelated.

Conclusions--design interventions

Summarizing all the points highlighted above, the following are important considerations in designing an assessment center to improve construct validity. First, assessors need a clear understanding of their rating task. Assessors should be trained in all the aspects of their job including avoiding common rater errors. Dimensions should be observable, carefully matched with exercises, and clearly defined by behavioral based scoring guidelines. Second, center designers need to be sensitive to the cognitive limitations of assessors and the ecological correlations among performance dimensions. The list of dimensions should be as economical and as independent as possible. Scoring checklists may also reduce the cognitive demands on assessors and eliminate the possible ill-effects of overloading assessors. Third,

multiple assessors should be used to estimate inter-rater reliability and assessors and exercises should be uncorrelated to ensure that exercise effects are not a function of assessor assignment.

Method Variance as a Function of Exercise Design

Behavior/exercise confounds

Following the Campbell and Fiske (1959) logic, an appropriate test of construct validity in assessment centers depends on the comparison of measurements of similar dimensions or similar clusters of observable behavior across independent methods (exercises) of measurement.

Researchers have acknowledged, however, that different exercises may not necessarily measure the same behaviors within a dimension (Robertson et al., 1987). This phenomenon has been conceptually explained by Byham (1977) who likens the process of combining observations from different exercises to taking multiple pictures of a city from an airplane. Byham suggests that while snapshots overlap with other pictures taken, each provides some uniqueness to the overall picture of the city.

Byham's notion seems to be born out empirically by Neidig et al. (1979). Using a series of regression analyses with dimension ratings as predictor variables and final consensus skill ratings as dependent variables, Neidig et al. concluded that each of their six exercises (a Background Interview, an assigned role and a non-assigned role Leaderless Group Discussion, an Interview Simulation, an Analysis Problem, and an In-basket) accounted for unique variance in each of the skill dimensions they were designed to measure. Neidig and Neidig (1984) expressed little surprise or concern about the low trait inter-correlations across exercises. They explained the findings by

suggesting that two exercises rarely represent parallel forms of measurement, even when two highly similar exercise forms such as group discussions are used. If exercises are uniquely sampling dimension-related behaviors that are not similar to the behavior sampled by other exercises, the finding of an exercise effect is not so surprising. D. R. Ilgen (personal correspondence, July 4, 1989) refers to this problem as the behavior/exercise confound.

An appropriate test of construct validity must involve a comparison of ratings across exercises designed specifically to measure the same behaviors. In the remainder of this section, I introduce exercise form and content as terms for classifying exercises that should facilitate appropriate comparisons of ratings across exercises. First, I define exercise form and content and then discuss the implications for construct validity.

Exercise form and content

Some researchers have made a distinction between the content and the form of a test. For example, Benson (1981) suggests that the universe of behaviors sampled by a test instrument depends on both the subject matter (content) and the structure (form) of the instrument. Since he is describing paper and pencil, school achievement tests, content refers to item content which represents some domain of knowledge. Form defined as "the stimulus to which the examinee responds" (p. 794) refers to features such as item format, item readability, and test directions. I suggest that form and content can be imported into an assessment center context as two dimensions according to which exercises can be classified.

I propose that the form of an exercise is referenced by its common title--Role play, In-basket, Group discussion, etc. Each of these exercise forms defines some of the situational parameters for the assessee such as whether the assessee will perform individually or in a group, the time allotted to the task, the primary mode of response (i.e. speaking, writing, etc.). These situational parameters, in turn, proscribe certain kinds of behavior that can be demonstrated in the exercise.

Previous construct validity studies have been conducted by comparing dimension ratings across different kinds of exercise forms. For example, Sackett and Dreher (1982) compared oral presentation in a group discussion and in a speech exercise. While these exercises have some oral presentation behaviors in common (e.g. speaking clearly, articulately, using appropriate grammar, etc.), there are a number of oral presentation behaviors unique to each exercise. For example, a speech exercise involves giving a prepared and rehearsed talk from notes with visual aids whereas a group discussion involves speaking in dialogue without rehearsal. It is likely that performance in these different exercise forms requires very different behaviors.

Similar comparisons can be found in other construct validity studies. Bycio et al. (1987) compared organizing and planning in an in-basket exercise, a role play, and a group exercise; it is also likely that these different exercises stimulated some similar and some very different kinds of organizing and planning behavior. Turnage and Muchinsky (1982) tested for similarity in interpersonal skills dimension ratings (human relations ability, communication skill, leadership) across exercises by comparing performance in the in-basket exercise,

primarily a written exercise, with performance in face-to-face, group exercises. In all these cases, the comparison across different forms may be a source of method variance since different forms stimulate different kinds of dimension-related behavior. The leaderless group discussion and the role play exercise (also called interview simulation) are the form manipulations in this study.

It is clear, however, that comparing across exercise forms is not the sole explanation of method variance. Bycio et al. (1987), discussing exercise structure (form) and content, note that the highest correlation in their data between two exercises similar in structure (form) was .50 sharing, at most, 25% exercise variance. They suggest that the remaining crucial difference between the exercises was content.

Since no system or language for classifying exercise content is available, I am proposing that exercise content can be described according to the design of the tasks performed in the exercise. For example, exercises are commonly designed with superior-subordinate counseling tasks, superior-subordinate confrontational tasks, and problem analysis tasks. For this study, I have chosen two commonly used task designs for group discussion exercises: cooperative and competitive tasks (Thornton and Byham, 1982). Exercises of similar content should stimulate a narrower, more similar range of behaviors than exercises that differ in content so higher correlations should result. Comparing how ratings correlate across cooperative and competitive exercises will allow for an estimation of variance accounted for by exercise content.

This study, then, explores whether exercise content and form represent a reasonable system for thinking about exercise design. The design of the exercises are manipulated in the assessment center in this

study so that assessee performance variance due to exercise form and content can be estimated. The conceptual basis for both the specific content and form manipulations follows.

Exercise forms--The leaderless group discussion and role play exercise

The leaderless group discussion exercise and the role play exercise were chosen as the form manipulations for this study. Thornton and Byham (1982) report that the leaderless group discussion was used in 45 percent of the centers in their review of approximately 500 assessment centers. In this study, both group exercises were designed for four participants. Specific goals were given to motivate and direct the participants in the group. No status or task-related roles were assigned to participants, and participants were instructed to interact as equals or peers.

Thornton and Byham (1982) report that the role play exercise (referred to as the interview simulation) was used in 75 percent of the assessment centers in their review. Role play exercises simulate the kinds of one-on-one encounters that occur in many jobs. The participant interacts with a "role player" who is a cohort of the assessment center and who is trained to play a standardized role with all participants. In this study, the participants were given specific goals to motivate and direct their behavior during the role play exercises. Participants were not given an assigned role to play in the exercise and were instructed to assume an equal or peer relationship with the role player.

One difference between the group and role play exercise is worth noting. The role play exercise holds more of the "situation" constant across participants than the group exercise, since participants encounter a single individual playing a standardized role. The

situation stimuli in group exercises in the form of the performance of other group members may vary widely across participants, since group members do not play a standardized role.

In sum, the two form manipulations, the leaderless group exercise and the role play exercise, represent commonly used assessment center exercise forms. Both types of exercise form require participants to interact face-to-face with others and resolve task-oriented and group process-oriented problems during the course of the exercises. As such, the exercise forms are similar enough so that they will stimulate comparable dimension-related behavior, eliminating a possible behavior-exercise confound in the design of the study.

Exercise content--cooperation and competition

Cooperation and competition represent common task designs for group discussion exercises (Thornton & Byham, 1982). In cooperative group discussion exercises, information is typically distributed among assessees so that each does not receive the same information about the problem, and therefore, cooperation is important for effectiveness. In competitive group discussion exercises, groups are typically assigned to distribute scarce resources among the individual members; assessees are pitted against each other so that one assessee's gain is another's loss. I have selected these task designs for use in both the group and role play exercise conditions. The paragraph that follows provides some background regarding the design of cooperative and competitive tasks in experimental research.

Group performance researchers have refined the distinctions between cooperative and competitive group task designs. Deutsch (1949) made an early distinction between cooperative and competitive groups. He

defined cooperative groups as groups in which individuals' attainments of goals are positively correlated. A negative correlation among the goal attainments of individuals exists in competitive groups such that goal attainment by one individual precludes the achievements of others. Miller and Hamblin (1963) fine tuned Deutsch's definition by considering the degree of task interdependence. Under conditions of high task interdependence, problems are solved most efficiently by the group as a whole. Under conditions of low task interdependence, each individual works more or less independently in the group. In the extreme case of the latter, individuals have no opportunity to impede or facilitate each other's performance. Mitchell and Silver (1990), synthesizing the work of a number of researchers, advanced cooperative versus competitive distinctions further by adding goal interdependence. High goal interdependence occurs when individuals in groups all share a common goal. Consistent with the group performance research, the cooperative exercise in this study was designed with a positive correlation between the goal attainments of individual students, high task interdependence, and high goal interdependence. The competitive exercise was designed with a negative correlation between the outcomes of individual students, low task interdependence, and high goal interdependence.

Exercise form and content--conclusions

In sum, it follows that assessment center exercises can differ in content and form, and both the content and form of an exercise circumscribe a particular set of behaviors that can be demonstrated in that exercise. Exercise form, referenced by its common title, and exercise content, or task design, specify the kinds of behaviors an assessee can demonstrate in the exercise. Both exercise form and

content may be factors explaining variability in assessee performance.

This study aims to determine whether exercise content and form represent reasonable dimensions for classifying exercise task design. First, the factor structure of the ratings was examined for evidence of form and content exercise factors. Second, the portion of method variance that can be accounted for by form and content will be estimated to determine to what extent the dominant method variance in assessment centers is a function of exercise form and content. Comparisons of exercises that are similar in content and form should result in dimension ratings that correlate higher than in previous studies since exercises similar in content and form narrow and make similar the range of dimension-related behavior sampled in exercise conditions. This narrowing should eliminate a portion behavior/exercise confound effect by ensuring that the same clusters of behaviors are measured by the different exercises.

Hypotheses

A confirmatory factor analysis (CFA) procedure similar to the procedure outlined by Widaman (1985) was used to test the hypotheses in this study. Hypotheses were tested by comparing a series of hierarchically nested models which are illustrated in Table 2.

First, it was hypothesized that design interventions suggested by previous researchers would result in ratings that demonstrate better convergent validity and discriminant validity. Evidence for construct validity was tested by comparing model I with models II and III from the A, B, and C columns (see Table 2). If model I (method factors only) fit the data as well as models II and III, there would be little evidence for retaining trait factors in the model. Evidence for discriminate

Table 2

Hierarchically Nested Models for Hypothesis Testing

	A	B	C	D
I	Null 12 factors (orthogonal) df=63	4 exercises (orthogonal) df=54	4 exercises (oblique) df=48	2 content (oblique) 2 form (oblique) df=41
II	1 trait df=54	1 trait 4 exercises (orthogonal) df=42	1 trait 4 exercises (oblique) df=32	
III	3 traits (orthogonal) df=54	3 traits (orthogonal) 4 exercises (orthogonal) df=42	3 traits (orthogonal) 4 exercises (oblique) df=36	
IV	3 traits (oblique) df=51	3 traits (oblique) 4 exercises (orthogonal) df=39	3 traits (oblique) 4 exercises (oblique) df=33	3 traits (oblique) 2 content (oblique) 2 form (oblique) df=25

Note. Degrees of Freedom (df)

validity was tested by comparing models II, III, and IV from the A, B, and C columns. If model II (1 general trait) fits the data as well as model III (three orthogonal traits), there would be little evidence that assessors were discriminating among the dimensions. Similarly, if model IV (three oblique trait factors) fits the data better than model III, there would be additional evidence that assessors were not discriminating among the dimensions.

Second, it was hypothesized that the factor structure in the ratings would reflect the form and content manipulations of the exercises. This hypothesis was tested by comparing models IB, IC, and ID (see Table 2). If models IB (four orthogonal exercise factors) and IC (four oblique exercise factors) fit the data as well as model ID (two form and two content factors), there would be little support for content and form exercise factors.

Third, model IVD is also of interest, for other than hypothesis testing reasons. The LISREL estimates from model IVD can be used to estimate the effects of the various facets of the design--traits, exercises, content and forms.

METHOD

Description and sample

The data for this study was obtained from an assessment center administered in cooperation with the Department of Education (DE) in a Midwestern state. The subjects for the study were juniors from a suburban/rural high school in the state. Student volunteers were recruited through presentations to the entire junior class in history and social studies classes and through individual contacts initiated by the principal, vice principals, counselors, and teachers. Students and a parent or legal guardian consented for participation, and students received oral and written development feedback in return for their participation. The letter introducing the project to students and the consent form can be found in Appendix A. The center was administered at the high school; students were excused from classes while participating in the center.

The center was scheduled to run 12 times with eight students participating in each session so a total of 96 students were targeted to participate. Students were scheduled into one of the twelve sessions by the vice principal and were assigned randomly to exercise conditions, once they arrived at the center. Five students were absent on the day they were scheduled to participate, and one student was dropped from the study because her consent form was misplaced by the high school. The exercises were videotaped for later rating by assessors, and one student was dropped from the sample because two of his exercises were not captured on video tape. Technical failure also occurred in a single exercise for eight other students, and ratings made on site by the center administrator or by role players were inserted in these eight

cases. In total, data from eighty-nine students were available for the study.

Dimensions

The process of selecting dimensions began with a list of 66 Employability skills dimensions which represented the dimensions of interest to the DE. The dimensions, generated by the Governor's Employability Skills Task Force, were considered important personal qualities for students' future employment and represented a wide range of ability and skill constructs in the areas of academic skills, personal management skills and teamwork skills. Problem solving skills, interpersonal skills, and initiative skills were selected from the list on the basis of previous assessment center research. First, previous factor analyses of assessment center ratings have demonstrated that ratings on multiple dimensions can be represented by a smaller list of two to four factors (Archambeau, 1979; Russell, 1985; Sackett and Hakel, 1979; Schmitt, 1977). The consistent finding across these studies is that assessment center ratings reflect a decision making/administrative factor, an interpersonal factor, and a participation/activity dimension. Problem solving skills, interpersonal skills, and initiative skills measures were developed to reflect these three factors. Second, three dimensions were selected on the basis that fewer dimensions may reduce the possible ill-effects of overloading assessors (Gaugler & Thornton, 1989).

The dimensions were defined with behavioral examples generated from dimensions used in other assessment center research and from the researcher's previous experience with assessment centers. Three representatives from the DE reviewed a draft of the dimensions, and

Table 3

Dimensions

PROBLEM SOLVING SKILLS:

- Seeking all available information about problems.
- Integrating available information to solve "whole" problems.
- Recommending many alternative solutions to problems.
- Recommending solutions and making decisions through logical reasoning and analysis.

INTERPERSONAL SKILLS:

- Reinforcing and supporting others' ideas.
- Promoting teamwork through communication.
- Promoting positive rapport among team members.
- Compromising in support of team objectives.
- Non-verbally demonstrating warmth and openness.

INITIATIVE SKILLS:

- Demonstrating motivation through active participation.
 - Stating ideas confidently and directly.
 - Setting direction and goals for others
 - Emphasizing the importance of achieving assignments
-

their suggestions were incorporated into the final definitions. The dimensions and a concise definition can be found in Table 3. The detailed dimension definitions used in the administration of the center can be found in Appendix B.

Exercises

Four exercises conditions were used to achieve the exercise form and content manipulations in this study. As for the form manipulation, two exercises were group exercises in which students worked with each other in groups of four, and two exercises were role play exercises in which individual students worked with confederates who were trained to play a standardized role. Students rotated between the exercise conditions so that four of eight students participated in the group exercises while four students participated in role play exercises. Each exercise took approximately twenty-five minutes including preparation time.

Instructions for each exercise were read aloud to the student by a staff member, and students were encouraged to ask any questions regarding unclear directions.

The group discussion and role play exercise forms were chosen because research suggests that they provided opportunity to observe the three dimensions of interest in this study. Thornton and Byham (1982) published figures regarding the opportunity to observe certain dimensions in non-assigned role, leaderless group discussions. Based on their sample, initiative behaviors were observed in 95 percent of non-assigned role, leaderless group discussion exercises; interpersonal skills (oral communication and sensitivity) in 90-95 percent; and problem solving (judgement and decisiveness) in 90 percent. Initiative behaviors were observed in 90 percent of certain types of role play

exercises; interpersonal skills (oral communication and sensitivity) in 90-95 percent; and problem solving skills (judgment and decisiveness) in 90 percent. As indicated previously, the matching of dimensions and exercises is critical, and based on the Thornton and Byham findings, the selection of these dimensions does not contribute to the method variance problem. According to Thornton and Byham, all three dimensions can be observed relatively frequently and equally in group discussion and role play exercises.

Exercise content was manipulated by altering the task design in the exercises. Competitive exercises were designed so that: students were given individual goals; the achievements of different students were negatively correlated; and students could perform tasks with low interdependence on other participants. Cooperative exercises were designed so that: students were given group goals; a student's goal achievement was correlated positively with the goal achievements of other students in the group; and students were highly interdependent on others for successful task performance. A competitive task design was used in one group discussion exercise and one role play exercise, and a cooperative task design was used in one group discussion and one role play exercise.

The setting for the problems presented to the student in each exercise was selected with careful consideration of the kinds of knowledge that students would need to resolve the problems and issues involved in the exercises. For the most part, exercises were self-contained in that all of the material necessary to solve the problem was included in the exercise directions and background information. Yet, exercise problems were also set in context thought to be familiar

territory for students such as a retail sales environment, a manufacturing environment, and a high school environment. Exercise materials can be found in Appendix C. Descriptions of each exercise follow.

Grant Allocation Exercise (competitive-group discussion). In this exercise, students were assigned to serve on a committee whose purpose was to decide how to allocate a \$15,000 grant awarded to their high school. During the exercise, each student individually prepared and presented an idea for allocating the money, then all four students discussed the ideas and developed a plan for allocating the money among the ideas. A number of generic ideas were laid on the table for use by any students who decided not to create an idea on their own.

The competitive manipulation in this exercise was introduced by the allocation rules. The rules stated that money could be allocated entirely to one idea or to any combination of two ideas as long as the amounts were not equal. Thus, individual student's outcomes were negatively correlated; money allocated to one project represented money lost by other projects. The exercise was low on task interdependence since students worked independently to create and earn money for their idea. Students were given the individual goal of winning the highest amount of money for their own project (low goal interdependence).

Team Manufacturing Exercise (cooperative-group discussion). In this exercise, students worked in groups of four to build tinker toy models according to a set of plans in the exercise instructions. Students were assigned to build tinker toy models and sell them to the center administrator. A number of errors and missing information were built into the plans to make the exercise a discussion exercise rather than a

strict manufacturing exercise. Given the problems encountered in the plans, the students spent the bulk of their time in collaborative problem solving.

The cooperative manipulation was designed into the exercise through the distribution of information and through group goal setting. Information and blueprints were presented to students in six folders placed on the work table. Each folder contained unique information about the assembly of the product, creating a performance task which could not be accomplished unless individuals pooled their information (high task interdependence). Students were given the group goal of earning \$2,000 during the allotted time period (high goal interdependence), and earnings for sales were credited to the group as a whole, creating a positive correlation between individual outcomes.

Customer Service Exercise (cooperative-role play). In this exercise, students worked with a confederate role player to review and recommend solutions regarding a customer service incident in a department store. Students received information about the incident from a variety of sources including the customer, the sales clerk, the clerk's supervisor, and an eyewitness. The confederate role player also had information about the problem some of which was identical to the student's information and some of which presented a slightly different perspective about the incident. Students were assigned to discuss the information with the role player, reach a consensus about what happened, and recommend solutions. Since students were not likely experienced in handling good and bad performing employees, a menu of employee commendation and disciplinary actions was included to prompt students about the kinds of action steps they could take.

The cooperative manipulation was achieved by giving the student and the role player the joint goal of discussing and arriving at specific solutions (high goal interdependence). Student and role player outcomes were positively correlated, since the success or failure in understanding and resolving the situation was shared by both equally. High task interdependence was created by the distribution of the information. Both the student and role player were given unique information about the incident. Any attempt by the student to resolve the problem without understanding the role player's information resulted in conflicting perspectives whereas a relatively conflict-free view of the incident could be achieved through sharing information.

Team Selection Exercise (competitive-role play). In this exercise, students assumed the role of a team leader in a food products company and were assigned to select three employees to fill three jobs on their team. Confederate role players also assumed the role of a team leader who needed to fill the same three jobs on their team. The student and the role player were given a description of two candidates for each job, creating a negotiating and decision making task where one of the two candidates would be assigned to fill one of the jobs on the student's team and the other would be assigned to fill the same job on the role player's team. The candidate descriptions included verbal descriptions of the candidates' strengths and weaknesses and a rating on a five point scale representing the candidates' overall level of qualification for the job. Thus, students could compare and select the candidates on the basis of both verbal and quantitative information. Candidates tended to differ on personality qualities more than on technical qualifications so that students could use their experience in choosing friends or

colleagues in school projects or other extracurricular activities.

The competitive manipulation was achieved by manipulating the qualifications of the two job candidates so that the student and role player were negotiating for scarce resources. The student was given the individual goal of selecting the best possible team (low goal interdependence), and the descriptions and numerical ratings of job candidates established the superiority of one candidate in each pair for a given job. This created a win/lose atmosphere for the negotiation of each job, and the outcomes achieved by the student and the role player in each negotiation were negatively correlated. Since candidates were typically discussed and assigned one job at a time, the exercise consisted of three separate negotiations, giving students three relatively distinct opportunities to "win" or compromise. Students were instructed that ratings of candidates assigned to a team could also be added to determine the relative strength of the entire team, and role players added the team totals aloud after each negotiation, further emphasizing the win/lose nature of the exercise. Low task interdependence by giving the student and the role player identical information and by leaving the student to his or her own devices in the negotiating process. Role players, playing a standardized role, negotiated for the same job candidates regardless of the selections made by the students.

Pre-test of exercises

The exercises were pre-tested with a sample of approximately forty undergraduate students who participated in a single exercise in return for course credit. Undergraduates were selected because of their relative closeness in age to high school students. The pre-test was

performed to ensure that instructions and background materials would be clearly understood by participants, to test the timing of the exercises, and to ensure that the exercises were stimulating the kinds of dimension-related behaviors that were expected. The pre-test also served as training for the role players and the center administrator. The exercises were also administered to a sample of 12 high school seniors from the same school from which the research sample was drawn. This sample was used primarily to test the readability of instructions and exercise materials. Adaptations in the exercises were made based on the pre-tests.

Role players and training

Nine individuals were trained by the researcher to play the standardized roles in the role play exercises. Eight role players were undergraduates who participated in the study in return for independent study course credits, and one was a graduate student. Undergraduates were used because of their relative closeness in age to the high school students. All role players received approximately forty hours of training prior to the administration of the center. Role players read selections from the assessment center literature to become familiar with the method and were introduced to the specific content of the role play exercises. Role players were also given a set of scripts to guide and to standardize their performance. The training agenda, scripts, and instructions to role players can be found in Appendix D.

The bulk of their training consisted of behavioral modeling training. Role players watched video taped examples of the roles and then were given the opportunity to practice the roles with each other. These practice sessions were video taped, and role players watched

themselves and others and received feedback from the trainer. After achieving a level of mastery of the roles, role players performed their roles in the pre-test of the exercises. These performances were also video taped, and role players were given feedback. All of the behavioral modeling training was done with groups of role players to foster the standardization of the roles.

Assessors and assessor training

Twenty one assessors were recruited and paid for their participation in the study by the DE. 12 assessors were teachers; 6 were school administrators, school counselors or curriculum administrators; and 3 were graduate students in psychology. None of the assessors had previous experience with assessment centers or assessor tasks.

Assessors were trained in a training program designed according to the guidelines established by the Task Force on Assessment Center Standards (1980). In a one day training session, assessors received training in understanding dimensions and exercises, observing and rating behavior, and avoiding halo, leniency, and central tendency. Assessor training included practice sessions during which assessors observed, used the scoring guidelines and rated examples of each exercise. Observations and ratings were discussed to achieve a measure of standardization across assessors. Materials from the assessor training sessions can be found in Appendix E.

A majority of the assessors had no previous contact with the students, yet eight of the assessors were teachers in the high school from which the sample was drawn and had varying levels of previous contact with the student subjects. As a result, three questions measuring the amount of previous school and community contact with

Table 4

Questions Regarding Previous Contact with Students

HOW MUCH SCHOOL-RELATED CONTACT HAVE YOU HAD WITH THIS STUDENT?

- 0^a ___ None at all.
- 1 ___ Brief, intermittent contact (ex. Like a librarian would have).
- 2 ___ In class or extracurricular activity for 1 quarter/semester.
- 3 ___ In class or extracurricular activity for 1 quarter/semester to 1 year.
- 4 ___ In class or extracurricular activity more than 1 year.

HOW MUCH NON-SCHOOL CONTACT HAVE YOU HAD WITH THIS STUDENT?

- 0 ___ None at all.
- 1 ___ Know casually from community (ex. Church).
- 2 ___ Know well from community (ex. Family friend).

OVERALL, HOW WELL DO YOU KNOW THIS STUDENT

- 0 ___ Not at all, don't even recognize.
- 1 ___ recognize student, know his/her name.
- 2 ___ know name, know student from reputation.
- 3 ___ know student well from school/community contact.
- 4 ___ know student extremely well from school/community contact.
-

Note. ^a Numbers in blanks are the coded values for data analysis

students were completed by these teachers every time they rated a student. These three questions, tapping previous school and community contact and overall association with the student, can be found in Table 4. Summary statistics for the responses will be reported in the Results section.

Rating procedure and scoring guidelines

The exercises in this study were videotaped at the high school for later viewing by assessors. Each exercise was rated by two independent assessors, allowing for estimates of inter-rater reliability. Assessors were rotated so that they were paired with different partners and that reliability estimates were not a function of the pairing assignments. Inter-rater reliability was estimated using inter-rater correlations and/or alpha (Borman, 1982; Schmitt, 1977). Assessors typically viewed the video tapes alone or in pairs. Assessors were instructed not to discuss observations or evaluations of students when working in pairs, and the researcher monitored the assessors to ensure that no discussion was taking place. Assessors observed tapes in distraction free environments, often in private rooms, and assessors' work space included a large table on which they could keep multiple pages of scoring guidelines in view at all times. Assessors tended to work 4 to 6 hour days and were encouraged to take frequent breaks to avoid fatigue.

The assessor assignments to exercises were made to eliminate potential assessor effects. Assessors rated all exercise conditions (assessors and exercises were uncorrelated). The rating procedure was also designed so that assessors only observed an individual student only one exercise, thus eliminating the possibility that correlations of ratings across exercises would occur because of an assessor's tendency

to rate a student consistent with previous ratings of the student. This was not achieved perfectly in practice, however; 19 of the eighty-nine students were rated twice by the same assessor. To minimize the effect of a potential memory or consistency effect, at least a week passed between viewings of the same student by one assessor.

Scoring guidelines in the form of behavioral checklists were developed for each exercise condition. Examples of high, typical, and poor behaviors were created for each dimension and for each exercise. Assessors were trained to make a check mark next to the behaviors on the checklist as many times as they saw them on the video. Assessors were also trained to add and check behaviors that they considered relevant to the dimension but that were not represented on the checklist. Assessors were instructed to consider the pattern of their check marks and assign an integer rating on a five point rating scale (5=high; 1=low) immediately after observing the exercise. The use of these checklists was intended to reduce the cognitive load on assessors since amount of attention devoted to note taking and categorization was reduced. The checklists for each exercise are in Appendix F.

The practices of rating dimensions after each exercise and of using a checklist scoring system are not used by all assessment centers. Ratings made after single exercises represents the procedure used in all of the construct validity studies cited earlier. It is necessary to collect ratings in this fashion to allow examination of the exercise effects found in previous studies. Scoring checklists are experiencing more widespread use in assessment center applications, even though their use is not widely represented in the published literature (Hollenbeck, 1990).

Design issues

Since the exercise designs were manipulated in this study, a number of steps were taken to ensure that exercise design was not confounded with assessors, role players, or other exercise effects. First, an order effect for exercises was controlled by changing the assessment center schedule every other day so that a relatively equal number of students participated in the exercises in one of four different orders. Second, role players were assigned so that students would interact with a different role player in the two different role play exercise conditions. Similarly, group exercise assignments were made so that students did not participate with the same three students in both group exercises. Third, two independent assessors were used so that inter-rater reliability could be estimated and so that reliability could be enhanced. The assignment of assessors to a particular student only once controlled for a rater-exercise confound. Fourth, role players played both role play roles, and assessors rated all exercise conditions so that neither specific assessors nor role players were nested within a particular exercise condition.

Statistical analyses

LISREL (Joreskog & Sorbom, 1986) was used for the confirmatory factor analysis procedure in this study. LISREL generates an estimated correlation matrix using maximum likelihood estimates of parameters corresponding to the hypothesized factor structure and hypothesized pattern of factor intercorrelations as a guide. This estimated matrix is then compared with the actual correlation matrix as a means of evaluating the hypotheses. CFA via LISREL differs from traditional factor analysis in that not all of the loadings between observable

Table 5

Model IVD: Three Trait Factors (Oblique); Two Form Factors (Oblique);
Two Content Factors (Oblique).

HYPOTHESIZED FACTOR MATRIX^a

		Traits			Form		Content		Uniqueness
Source		PS	IP	IN	GR	RP	CM	CO	
CMGR	PS	?	0	0	?	0	?	0	?
	IP	0	?	0	?	0	?	0	?
	IN	0	0	?	?	0	?	0	?
COGR	PS	?	0	0	?	0	0	?	?
	IP	0	?	0	?	0	0	?	?
	IN	0	0	?	?	0	0	?	?
CMRP	PS	?	0	0	0	?	?	0	?
	IP	0	?	0	0	?	?	0	?
	IN	0	0	?	0	?	?	0	?
CORP	PS	?	0	0	0	?	0	?	?
	IP	0	?	0	0	?	0	?	?
	IN	0	0	?	0	?	0	?	?

HYPOTHESIZED PATTERN OF FACTOR INTERCORRELATIONS

	PS	IP	IN	GR	RP	CM	CO
PS	1.0						
IP	?	1.0					
IN	?	?	1.0				
GR	0	0	0	1.0			
RP	0	0	0	?	1.0		
CM	0	0	0	0	0	1.0	
CO	0	0	0	0	0	?	1.0

Note. ^a Fixed values are represented by 0 and 1. Free parameters are represented by ?. Abbreviations include problem solving (PS), interpersonal skills (IP), initiative (IN), competitive exercise (CM), cooperative exercise (CO), role play exercise (RP), group exercise (GR).

variables and factors are estimated. Hypothetical models are specified by the researcher by fixing factor loadings at zero or one or by allowing them to be freely estimated. When the LISREL algorithm generates a solution, the fixed loadings (0, 1.0) remain at their fixed level and the free loadings are estimated as in traditional factor analysis. Relationships are hypothesized in three different matrices. Lambda X is a model of the relationships between observed variables and latent constructs; Phi includes estimates of the relationships between latent constructs; and Theta Delta represents estimates of the uniqueness or error associated with each observed variable. Table 5 shows the hypothesized relationships for all three matrices for Model IVD as an example.

LISREL calculates four indices of fit for determining which model is the best representation of the factor structure in the ratings (Joreskog & Sorbom, 1986). The four indices are chi-square, the goodness of fit index, the adjusted goodness of fit index, and the root mean square residual. Chi-square (χ^2) can be used as a statistical test to judge how well a model fits the data. A significant chi-square suggests that a model should be rejected since this suggests significant differences between the model and an unconstrained variance-covariance matrix. A non-significant chi-square suggests a good fitting model. Chi-square can also be used to compare the relative fit of two different models by testing the significance of the differences between the chi-square values relative to the difference between the number of degrees of freedom for the two models. The goodness of fit index (GFI) is a measure of the degree to which the variance-covariance matrix is accounted for by the model. The adjusted goodness of fit index (AGFI)

is a similar index that adjusts for the degrees of freedom in the model by using mean squares rather than sums of squares in the calculation. Observed GFI and AGFI values range from 0 to 1, though values theoretically could be negative. Larger values approaching 1.0 represent a better fitting model. No standards are available for judging fit, since the statistical distribution of these indices is unknown. To aid the interpretation of fit indices, Bentler and Bonett (1980) suggested that a model represents a reasonable fit when rho, a similar index to GFI, is at least .90, and Widaman (1985) has suggested that differences in rho of .01 or more were important on practical grounds. The root mean square residual (RMSR) is a measure of the average of the residual variances and covariances. Smaller values approaching zero represent better fitting models. Joreskog and Sorbom (1986) suggest that the size of RMSR should be judged relative to the observed variances and covariances.

Research has shown that sample size is an important consideration when interpreting indices of fit (Mulaik, James, Van Alstine, Bennett, Lind, & Stilwell, 1989; Marsh, Balla, & McDonald, 1988). Chi-square seems particularly vulnerable since chi-square is frequently non-significant, suggesting a good fitting model, with small sample sizes. Even a well fitting model can be rejected (i.e. a significant chi-square) with a large sample size. Despite claims to the contrary by Joreskog and Sorbom, research by Marsh et al. (1988) suggests that the indices of fit computed by LISREL are substantially affected by sample size. Marsh et al. examined 33 different indices of fit by varying the sample sizes ($N = 25, 50, 100, 200, 400, 800, \text{ and } 1600$) for four different data sets. Among the many indices examined, the index

proposed by Tucker and Lewis (1973) and other indices based on the Tucker and Lewis formula proved to be the least sensitive to sample size. Based on these results and considering the moderate sample size in this study, the Tucker-Lewis index was calculated for each model in the analysis. The range for the Tucker-Lewis index is also 0 to 1.0 with larger values approaching 1.0 representing the best fitting model.

RESULTS

The results pertaining to the hypotheses are reported in this section. First, descriptive statistics of rating distributions and inter-rater reliabilities are reported. Second, the results of the hypothesis testing through the confirmatory factor analysis procedure are reported.

Rating distributions and inter-rater reliability

The descriptive statistics of interest are reported in Tables 6 and 7. Table 6 reports the mean scores, standard deviations, and alphas for each dimension within the exercises. The results suggest that the ratings were relatively uniform across dimensions and exercises in terms of their means and standard deviations. The alphas were within an acceptable range, suggesting that independent assessors rated consistent with each other. The magnitude of the alphas was consistent with previous assessment center research, though one value, .62, for the interpersonal dimension in the cooperative group exercise was somewhat low. Given the acceptable levels of reliability the ratings, the average of the two assessors ratings was used to represent the students' score for each dimension within exercises. These scores were used in the confirmatory factor analysis procedure.

Table 7 reports assessor statistics including the number of students rated, the means and standard deviations for the questions regarding the previous contact with students, and the means and standard deviation for each dimension. Regarding previous contact with students, the mean and standard deviations for the non-school/community question is reported in the column headed PCCM. The questions regarding previous school related contact and the overall familiarity question were highly

Table 6

Dimension Means, Standard Deviations, and Alphas by Exercise

Exercise	Mean(SD)	Alpha
Competitive Group		
PS	3.07 (1.07)	.79
IP	2.95 (.92)	.68
IN	2.74 (1.15)	.82
Cooperative Group		
PS	3.02 (1.08)	.79
IP	2.99 (.95)	.62
IN	3.10 (1.15)	.82
Competitive Role Play		
PS	3.15 (1.22)	.91
IP	3.03 (1.02)	.80
IN	2.74 (1.10)	.81
Cooperative Role Play		
PS	3.17 (1.21)	.88
IP	3.16 (1.06)	.78
IN	2.82 (1.21)	.86

Note. Dimension titles are abbreviated as follows: Problem solving skills (PS), Interpersonal skills (IP), and Initiative skills (IN).

Table 7

Means and Standard Deviations for Previous Contact Scales and for
Dimension Ratings by Rater

R	N	PCCM (0-2) ^a	PCOV (0-4)	PS (1-5)	IP (1-5)	IN (1-5)
1	66	.00	.00	3.48(1.17)	3.58(1.19)	3.30(1.24)
2	64	.00	.00	3.37(1.00)	3.11(.89)	3.05(1.03)
3	64	.00	.00	3.28(.92)	3.23(.90)	3.12(1.09)
4	62	.00	.00	3.14(1.27)	3.02(1.11)	2.81(1.29)
5	61	.00	.00	3.23(1.01)	3.08(.86)	3.08(1.07)
6	61	.10(.58)	.18(.66)	2.98(1.26)	2.67(1.09)	2.39(1.21)
7	60	.00	1.41(1.51)	3.07(1.47)	3.13(1.36)	2.87(1.41)
8	49	.00	.22(.74)	3.10(1.00)	3.26(.93)	3.10(1.14)
9	19	.12(.48)	.41(1.18)	2.05(1.47)	2.10(1.05)	1.84(1.34)
10	23	.00	.00	2.87(.87)	3.00(1.00)	2.74(1.01)
11	20	.00	.00	3.50(1.23)	3.30(1.17)	3.00(1.38)
12	17	.00	1.00(1.21)	1.94(.97)	2.00(.79)	1.65(.79)
13	14	.00	.00	3.07(1.64)	2.79(1.31)	2.79(1.31)
14	14	.00	.00	3.43(1.40)	3.36(.93)	2.93(1.49)
15	14	.00	.00	2.57(1.28)	2.93(1.03)	2.79(1.37)
16	14	.00	.00	3.36(1.34)	3.00(1.30)	2.86(1.41)
17	14	.13(.35)	1.00(1.19)	2.93(1.33)	2.50(1.09)	2.71(1.14)
18	14	.85(.44)	.77(1.30)	2.29(1.27)	2.79(1.25)	2.07(1.14)
19	14	.00	.00	3.43(1.34)	3.57(1.50)	3.79(1.19)
20	13	.08(.28)	.46(1.13)	2.61(1.50)	3.00(1.00)	1.92(1.26)
21	12	.00	.00	2.36(1.45)	1.79(.70)	1.71(.73)

Note. ^a Scale ranges are in parentheses under each heading. Values in parentheses in body of table are standard deviations. Raters are listed in order of number of students rated. Abbreviations include: Rater (R), Number of students rated (N), Previous non-school/community contact (PCCM) Previous overall contact (PCOV), Problem solving (PS), Interpersonal skill (IP), Initiative (IN).

correlated (.93) so only the overall question mean and standard deviation is reported in the column headed PCOV. The bulk of the assessors had no previous association with the students, and for those assessors with a previous association, the means were relatively small, reducing the concern that previous association with the students had an effect on the result. The level of concern was also not high considering any one of these eight assessors observed students in only one exercise. The means and standard deviations for dimension ratings reflect varying levels of central tendency and leniency across assessors, yet assessors did not necessarily view a random sample of the participants.

Multitrait-Multimethod Matrix

The correlations between dimension ratings are reported in multitrait-multimethod form in Table 8. On first glance, method variance seems dominant, and there appears to be very little discrimination among traits. The correlations within monomethod triangles are of similar magnitude and uniformly higher than the values on the validity diagonals (monotrait-heteromethod). This is confirmed further by examining the mean HTMM, MTHM, and HTHM values across the matrix. A comparison with Table 1 suggests the data is remarkably similar to previous assessment center data. However, the high level of overall halo in the Turnage and Muchinsky study, the third data set in the Sackett and Dreher study, and the Gaugler and Thornton study is not present in these data (mean HTHM = .22). A glance at the table also allows for a preliminary check of the content and form manipulation. The highest monotrait-heteromethod and heterotrait-heteromethod correlations occur in the block with the role play form in common

Table 8

Multitrait-Multimethod Correlations

	CMGR			COGR			CMRP			CORP		
	PS	IP	IN	PS	IP	IN	PS	IP	IN	PS	IP	IN
<u>CMGR</u>												
PS	(.78)											
IP	.66	(.68)										
IN	.81	.60	(.82)									
<u>COGR</u>												
PS	<u>.15</u>	.11	.19	(.79)								
IP	.26	<u>.22</u>	.28	.60	(.62)							
IN	.26	.19	<u>.33</u>	.76	.69	(.82)						
<u>CMRP</u>												
PS	<u>.27</u>	.11	.28	<u>.32</u>	.27	.39	(.91)					
IP	.25	<u>.07</u>	.28	.06	<u>.20</u>	.13	.70	(.80)				
IN	.13	-.06	<u>.16</u>	.14	.13	<u>.25</u>	.77	.65	(.91)			
<u>CORP</u>												
PS	<u>.21</u>	.17	.22	<u>.27</u>	.26	.19	<u>.34</u>	.35	.25	(.89)		
IP	.24	<u>.17</u>	.29	.17	<u>.31</u>	.18	.33	<u>.49</u>	.30	.78	(.78)	
IN	.25	.19	<u>.25</u>	.16	.23	<u>.14</u>	.35	.44	<u>.32</u>	.84	.79	(.86)
Mean Values across the matrix												
HTMM	.72			MTHM	.25		HTHM	.22				

Note. Alpha reliabilities in parentheses. Validities underlined. Competitive exercise (CM), Cooperative exercise (CO), Group exercise (GR), Role play exercise (RP). Heterotrait-monomethod (HTMM), Monotrait-heteromethod (MTHM), Heterotrait-heteromethod (HTHM).

(.25 to .49) whereas the lowest values occur where the exercises have competitive content in common (-.06 to .28).

Convergent and discriminant validity

The test of convergent and discriminant validity here and the tests of the remaining hypotheses are based on the LISREL analysis. Regarding individual models, a judgment of good fit can be made to the extent that X^2 is not significant, GFI, AGFI, and T/L approach one, and RMSR approaches zero. Comparing the relative fit of hierarchically nested models is also important for hypothesis testing. This comparison is achieved by a Chi-square test of significance of the differences in X^2 and the degrees of freedom between the more restricted and less restricted model. In other words, a less restricted model can be accepted as a better fit if it results in a significant decrease in the X^2 value given the number of degrees of freedom that are lost by allowing more parameters to be estimated.

Evidence for convergent and discriminant validity were tested by comparing models I, II, III, IV for columns A, B, and C (see Table 2). Model IVA resulted in a solution in which factor intercorrelation estimates were greater than 1.0, so the data from this model is not reported. The indices of fit for the column A models, reported in Table 9, are very poor suggesting that models hypothesizing traits only are not a good fit of the data. Tables 10 and 11 report the LISREL estimates for models IIA and IIIA. The uniqueness values in both of these tables are quite high suggesting much variance is unexplained by traits only models.

Column B models allow for both trait and method factors and as can be seen in Table 12, are much better approximations of the data. The

Table 9

Convergent and Discriminant Validity: Column A Indices of Fit

Model	df	χ^2	GFI	AGFI	RMSR	T/L
Null 12 Factors (orthogonal)						
IA	66	734.34	.382	.270	.352	
1 trait						
IIA	54	442.48	.545	.343	.210	.290
3 traits (orthogonal)						
IIIA	54	642.19	.487	.259	.330	-.075

Note. Abbreviations include: Degrees of freedom (df), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Root Mean Square Residual (RMSR), Tucker and Lewis Index (T/L).

Table 10

LISREL Estimates for Model IIA: One trait factor

FACTOR MATRIX			
	Traits		Uniqueness
Source	TR		
CMGR	PS	.32	.89
	IP	.24	.94
	IN	.35	.88
COGR	PS	.27	.93
	IP	.34	.88
	IN	.27	.93
CMRP	PS	.46	.78
	IP	.52	.72
	IN	.39	.85
CORP	PS	.89	.21
	IP	.86	.25
	IN	.86	.18
FACTOR INTERCORRELATIONS			
	TR		
	TR	1.0	

Note. Fixed values are represented by 0 and 1. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), competitive exercise (CM), cooperative exercise (CO), role play exercise (RP), group exercise (GR), trait factors (TR).

Table 11

LISREL Estimates for Model IIIA: Three Trait Factors (Orthogonal)

FACTOR MATRIX

		Traits			Uniqueness
Source		PS	IP	IN	
CMGR	PS	.39	0	0	.85
	IP	0	.21	0	.96
	IN	0	0	.51	.74
COGR	PS	.47	0	0	.78
	IP	0	.37	0	.86
	IN	0	0	.50	.75
CMRP	PS	.67	0	0	.55
	IP	0	.57	0	.68
	IN	0	0	.49	.76
CORP	PS	.53	0	0	.72
	IP	0	.84	0	.28
	IN	0	0	.47	.78

FACTOR INTERCORRELATIONS

	PS	IP	IN
PS	1.0		
IP	0	1.0	
IN	0	0	1.0

Note. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), competitive exercise (CM), cooperative exercise (CO), role play exercise (RP), group exercise (GR).

LISREL estimates for these models are reported in Tables 13 through 15. Regarding convergent validity, a significant ($\chi^2 = 48.08$; $df = 2$; $p < .05$) gain in fit occurs when adding a general trait (model IIB) to the exercises only model (model IB). Indeed, model IIB seems to be a good fit of the data. According to Widaman (1985), evidence that a model hypothesizing traits fits better than a methods only model is support for convergent validity. The fit of model IIB over model IB, however, only suggests convergence on a general trait and not convergence on three traits. Regarding discriminant validity, all of the indices of fit for the model hypothesizing three orthogonal trait factors (model IIIB) suggest a poorer fit than model IIB. A reasonable conclusion is that assessors were not highly discriminating in their rating of different dimensions. Model IVB, allowing for oblique relations among trait factors, seems to be the best fitting model in the column and represents an incremental yet statistically significant fit over model IIB ($\chi^2 = 14.49$; $df = 3$; $p < .05$) and over model IIIB ($\chi^2 = 33.63$; $df = 3$; $p < .05$). GFI, AGFI, and T/L values are greater for model IVB; RMSR, however, is smaller suggesting that model IIB is a better fit. The support of models IIB and IVB relative to model IIIB warrants a conclusion that assessors were not discriminating among traits. Even when three traits are hypothesized, the correlations among trait factors in Table 15 are very high, ranging from .78 to .93.

An examination of the factor loadings in Table 15 offers further insight. Exercises in model IVB seem to account for a higher proportion of variance than traits, particularly for the group exercises and for the cooperative role play. The highest trait loadings occurred for the competitive role play exercise including trait loadings for PS and IP

Table 12

Convergent and Discriminant Validity: Column B Indices of Fit

Model	df	χ^2	GFI	AGFI	RMSR	T/L
4 exercises (orthogonal)						
IB	54	103.17	.832	.757	.208	.910
1 trait, 4 exercises (orthogonal)						
IIB	42	55.09(ns)	.907	.827	.064	.969
3 traits (orthogonal), 4 exercises (orthogonal)						
IIIB	42	74.23	.872	.763	.199	.924
3 traits (oblique), 4 exercises (orthogonal)						
IVB	39	40.60(ns)	.930	.861	.072	.996

Note. Abbreviations are as follows: Degrees of Freedom (df); Goodness of Fit Index (GFI); Adjusted Goodness of Fit Index (AGFI); Root Mean Square Residual Index (RMSR); Tucker and Lewis Index (T/L). (ns) = not significant at $p < .05$

Table 13

LISREL Estimates for Model IIB: One Trait Factor; Four Exercise Factors
(Orthogonal)

FACTOR MATRIX

		Traits		Exercises			Uniqueness
Source		TR	CMGR	COGR	CMRP	CORP	
CMGR	PS	.42	.83	0	0	0	.13
	IP	.24	.67	0	0	0	.50
	IN	.47	.73	0	0	0	.24
COGR	PS	.23	0	.79	0	0	.33
	IP	.41	0	.64	0	0	.42
	IN	.32	0	.87	0	0	.13
CMRP	PS	.57	0	0	.68	0	.22
	IP	.68	0	0	.46	0	.32
	IN	.45	0	0	.76	0	.23
CORP	PS	.52	0	0	0	.77	.14
	IP	.66	0	0	0	.57	.24
	IN	.59	0	0	0	.70	.16

HYPOTHESIZED PATTERN OF FACTOR INTERCORRELATIONS

	TR	CMGR	COGR	CMRP	COGR
TR	1.0				
CMGR	0	1.0			
COGR	0	0	1.0		
CMRP	0	0	0	1.0	
CORP	0	0	0	0	1.0

Note. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), Competitive exercise (CM), Cooperative exercise (CO), Role play exercise (RP), Group exercise (GR), Trait factors (TR).

Table 14

LISREL Estimates for Model IIIB: Three Trait Factors (Orthogonal); Four Exercise Factors (Orthogonal)

FACTOR MATRIX

Source	Traits			Exercises				Uniqueness
	PS	IP	IN	CMGR	COGR	CMRP	CORP	
CMGR	PS	.02	0	0	.93	0	0	.13
	IP	0	-.05	0	.71	0	0	.50
	IN	0	0	-.18	.85	0	0	.22
COGR	PS	-.32	0	0	0	.76	0	.26
	IP	0	.31	0	0	.75	0	.37
	IN	0	0	-.33	0	.93	0	.02
CMRP	PS	-.20	0	0	0	0	.91	.14
	IP	0	.41	0	0	0	.75	.23
	IN	0	0	-.08	0	0	.84	.28
CORP	PS	-.25	0	0	0	0	0	.90
	IP	0	.28	0	0	0	0	.80
	IN	0	0	.05	0	0	0	.92

FACTOR INTERCORRELATIONS

	PS	IP	IN	CMGR	COGR	CMRP	CORP
PS	1.0						
IP	0	1.0					
IN	0	0	1.0				
CMGR	0	0	0	1.0			
COGR	0	0	0	0	1.0		
CMRP	0	0	0	0	0	1.0	
CORP	0	0	0	0	0	0	1.0

Note. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), Competitive exercise (CM), Cooperative exercise (CO), Role play exercise (RP), Group exercise (GR), Trait factors (TR).

Table 15

LISREL Estimates for Model IVB: Three Trait Factors (oblique); Four Exercise Factors (orthogonal)

FACTOR MATRIX

Source	Traits			Exercises				Uniqueness
	PS	IP	IN	CMGR	COGR	CMRP	CORP	
CMGR	PS	.34	0	0	.85	0	0	.13
	IP	0	.13	0	.68	0	0	.50
	IN	0	0	.40	.75	0	0	.23
COGR	PS	.42	0	0	0	.71	0	.30
	IP	0	.51	0	0	.65	0	.38
	IN	0	0	.52	0	.79	0	.12
CMRP	PS	.76	0	0	0	0	.59	.08
	IP	0	.68	0	0	0	.53	.27
	IN	0	0	.54	0	0	.66	.27
CORP	PS	.47	0	0	0	0	0	.10
	IP	0	.60	0	0	0	0	.18
	IN	0	0	.49	0	0	0	.17

FACTOR INTERCORRELATIONS

	PS	IP	IN	CMGR	COGR	CMRP	CORP
PS	1.0						
IP	.78	1.0					
IN	.93	.87	1.0				
CMGR	0	0	0	1.0			
COGR	0	0	0	0	1.0		
CMRP	0	0	0	0	0	1.0	
CORP	0	0	0	0	0	0	1.0

Note. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), Competitive exercise (CM), Cooperative exercise (CO), Role play exercise (RP), Group exercise (GR), Trait factors (TR).

that were higher than the exercise loadings. This is consistent with the design of this exercise; first, as a role play, more situational variance is controlled than in group exercise, and second, given the low task interdependence in the exercise, students were challenged to rely on their own skills much more than in the cooperative, high task interdependent, role play where their performance was somewhat dependent on the role player. Given the conditions of performing alone on an independent task, it follows that more variance in ratings is due to the person (i.e. traits). This also suggests that an important part of the situation are the persons in the groups with whom assessee are evaluated. In sum, there is some evidence that ratings converge on a general trait, yet overall, method variance is dominant and there is little evidence supporting discriminate validity.

A number of problems occurred in the LISREL procedure while analyzing the column C models. Some parameter estimates were greater than 1.0 in the Phi and Theta Delta matrices, and some the matrices in models IIC and IIIC were indeterminant. The analysis of model IVC failed to converge after 250 iterations. These problems did not allow for any hypothesis testing with the models in this column.

Exercise form and content

Support for exercise form and content was tested by comparing row I models (see Table 2); the indices of fit for these models are in Table 16 and the LISREL estimates are reported in Tables 17 through 19. In order for the LISREL procedure to converge on a solution, one of the paths in the Lambda X factor matrix was fixed at zero. The loading of the IP rating in the competitive group on the competitive factor was chosen to be fixed at zero since previous analyses showed this path to

Table 16

Exercise Form and Content: Row I Indices of Fit

Model	df	χ^2	GFI	AGFI	RMSR	T/L
Null 12 Factors (orthogonal)						
IA	66	734.34	.382	.270	.352	
4 exercises (orthogonal)						
IB	54	103.17	.832	.757	.208	.910
4 exercises (oblique)						
IC	48	63.90(ns)	.890	.821	.065	.967
2 content (oblique) 2 form (oblique)						
ID	41	48.27(ns)	.916	.840	.042	.982

Note. Abbreviations are as follows: Degrees of Freedom (df); Goodness of Fit Index (GFI); Adjusted Goodness of Fit Index (AGFI); Root Mean Square Residual Index (RMSR); Tucker and Lewis Index (T/L). (ns) = not significant at $p < .05$

Table 17

LISREL Estimates for Model IB: Zero traits; Four Exercise Factors
(Orthogonal)

FACTOR MATRIX

Source		Exercise				Uniqueness
		CMGR	COGR	CMRP	CORP	
CMGR	PS	.94	0	0	0	.12
	IP	.70	0	0	0	.51
	IN	.86	0	0	0	.26
COGR	PS	0	.81	0	0	.34
	IP	0	.74	0	0	.46
	IN	0	.94	0	0	.12
CMRP	PS	0	0	.91	0	.18
	IP	0	0	.77	0	.40
	IN	0	0	.85	0	.28
CORP	PS	0	0	0	.91	.17
	IP	0	0	0	.85	.27
	IN	0	0	0	.92	.14

FACTOR INTERCORRELATIONS

	CMGR	COGR	CMRP	CORP
CMGR	1.0			
COGR	0	1.0		
CMRP	0	0	1.0	
CORP	0	0	0	1.0

Note. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), Competitive exercise (CM), Cooperative exercise (CO), Role play exercise (RP), Group exercise (GR).

Table 18

LISREL Estimates for Model IC: Zero Trait Factors; Four exercise
Factors (Oblique)

FACTOR MATRIX

		Exercises				Uniqueness
Source		CMGR	COGR	CMRP	CORP	
CMGR	PS	.92	0	0	0	.14
	IP	.70	0	0	0	.51
	IN	.87	0	0	0	.24
COGR	PS	0	.80	0	0	.36
	IP	0	.73	0	0	.46
	IN	0	.95	0	0	.10
CMRP	PS	0	0	.93	0	.14
	IP	0	0	.77	0	.41
	IN	0	0	.83	0	.32
CORP	PS	0	0	0	.91	.17
	IP	0	0	0	.86	.27
	IN	0	0	0	.93	.14

FACTOR INTERCORRELATIONS

	CMGR	COGR	CMRP	CORP
CMGR	1.0			
COGR	.32	1.0		
CMRP	.28	.37	1.0	
CORP	.29	.23	.43	1.0

Note. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), Competitive exercise (CM), Cooperative exercise (CO), Role play exercise (RP), Group exercise (GR).

Table 19

LISREL Estimates for Model ID: Zero Trait factors; Two form factors
(Oblique), Two Content Factors (Oblique)

FACTOR MATRIX

		Form		Content		Uniqueness
Source		GR	RP	CM	CO	
CMGR	PS	.90	0	.19	0	.15
	IP	.72	0	0 ^a	0	.48
	IN	.85	0	.20	0	.24
COGR	PS	.12	0	0	.81	.32
	IP	.26	0	0	.70	.45
	IN	.24	0	0	.90	.13
CMRP	PS	0	.31	.89	0	.12
	IP	0	.46	.65	0	.38
	IN	0	.28	.78	0	.32
CORP	PS	0	.88	0	.26	.18
	IP	0	.84	0	.22	.26
	IN	0	.93	0	.18	.13

FACTOR INTERCORRELATIONS

	GR	RP	CM	CO
GR	1.0			
RP	.29	1.0		
CM	0	0	1.0	
CO	0	0	.41	1.0

Note. ^a denotes parameter fixed at 0 for LISREL to converge. Fixed values are represented by 0 and 1.0. Abbreviations include Problem solving (PS), Interpersonal skills (IP), Initiative (IN), Competitive exercise (CM), Cooperative exercise (CO), Role play exercise (RP), Group exercise (GR).

be non-significant. Once this path was fixed at zero, the LISREL procedure converged without difficulty.

Examining the fit of these models suggests that method or exercise is a dominant factor in this center as it has been in other centers. Models IB and IC hypothesizing no trait factors and four exercise factors representing the four different exercises are both reasonable fits of the data. Model IC, allowing for correlations among exercise factors, represents an incremental yet statistically significant increase in fit ($\chi^2 = 39.27$; $df = 6$; $p < .05$) over model IB. This suggests that exercises are not completely independent; Table 18 shows that the correlations between exercise factors range from .28 to .32. The highest factor intercorrelation in model IC, Table 18, occurs between the exercises which have the role play form in common, confirming further this phenomena that was observed in the MIMM matrix. Model ID, hypothesizing form and content exercise design factors, also represents an incremental yet statistically significant fit over model IB ($\chi^2 = 54.90$; $df = 13$; $p < .05$) and model IC ($\chi^2 = 15.63$; $df = 7$; $p < .05$), suggesting there is some empirical support for the existence of these factors in exercise design. All of the indices of fit for model ID suggest incremental gains over the exercises only models IB and IC. It should be emphasized that the improvement in fit is very small, and a consideration of parsimony lends support to Model IC. The specific loadings in model ID, Table 19, also lend support for an interpretation that each exercise represented a different factor. Table 19 shows four distinct clusters of loadings (under group for the competitive group, under cooperative for the cooperative group, under competitive for the competitive role play, and under role play for the

cooperative role play), and these four clusters do not occur in a way that is consistent with the exercise design manipulations. This further reinforces the predominance of other method factors over the effects of the content and form of an exercise.

Estimates of Variance Attributable to Facets in the Design

The LISREL procedure for the model IVD hypothesizing 3 traits and two form and content factors did not converge after 250 iterations. The ANOVA procedure (e.g. Stanley, 1961; Kavanagh et al. 1971) was used to calculate the portions of variance attributable to the different facets in the model. Table 20 reports these values. Consistent with the previous models examined, the ANOVA procedure confirms the dominant effect of method or exercise on assessee performance. Exercise form seems to account for a small portion (16%) of the method variance whereas the effect of exercise content is negligible. In fact, there is evidence of a higher average correlation across all exercises than across exercises similar in content. Eighty three percent of the method variance remains unexplained.

Table 20

Variance Attributable to Facets in the Design

Facet	Variance
Person	.222
Trait	.026
Method	.499
Form	.082
Content	.000 ^a
Error	.253

Note. ^a Mean correlation across exercises with content in common was less than the mean correlation across all exercises.

DISCUSSION

This study set out to achieve two purposes. In pursuit of the first purpose, a number of features based on recommendations of previous assessment center researchers were designed into this center to increase the probability of finding evidence of convergent and discriminant validity. First, a number of design interventions were implemented based on the suggestions of previous researchers: dimensions were created to be uncorrelated; exercises were carefully selected to sample the dimensions frequently and equally; assessors were trained according to recommended guidelines including avoiding halo error; and dimension definitions and scoring guidelines were systematically developed. Second, many steps were taken to reduce the cognitive load on assessors including reducing the number of dimensions, streamlining the rating task with scoring checklists, and assigning assessors to rate only one assessee. Third, steps were taken to control for assessor reliability and for assessor and exercise confounds. In pursuit of a second purpose, exercise content and form were manipulated to determine the effects of these exercise features on assessee performance. The results of this study further confirm the robust and durable finding of a dominant exercise factor in assessment center ratings. Correlations between ratings of dissimilar dimensions within exercises were higher than correlations between ratings of similar dimensions across exercises. The implications of these findings in light of the purposes of this study follow.

The center design interventions in this study did not result in better evidence of convergent or discriminant validity. Previous

construct validity studies have been performed on archived data and under conditions where the researchers had no control over the design of exercises, dimensions, training, or scoring procedures. This study found highly similar results even though many of the features were designed on the basis of the recommendations of previous researchers and were under the complete control of the researcher. While this study's manipulation of these design characteristics was by no means exhaustive, the findings raise little hope that efforts aimed at improving the mechanics of assessment centers will be able to make a significant impact on construct validity. This conclusion is similar to conclusions drawn from a broader base of research within performance appraisals; years of research on scale formats and assessor training has not led to real improvements in the appraisal process (Landy & Farr, 1980).

Human judgment processes are at the root of measurement in assessment centers, and researchers have cited cognitive processes as an explanation of method variance. Recognizing that assessors may be overloaded by the demands of assessor tasks, the center in this study was designed to reduce the cognitive processing demands on assessors and thereby, reduce the likely negative impact of overload. These design interventions included reducing the number of dimensions, reducing the redundancy among dimensions, reducing the note taking and categorization demands by using scoring checklists, assigning assessors to observe only one assessee, providing assessors with a comfortable and distraction-free rating environment, and allowing assessors to take frequent rest breaks as necessary to avoid fatigue. These interventions did not make a significant impact on construct validity, however. Gaugler and Thornton (1989), under stricter experimental conditions, similarly found

that reducing the cognitive load on assessor by reducing the number of dimensions did not result in evidence of construct validity. Gaugler and Thornton's results combined with the current study represent a body of support that reducing cognitive demands on assessors is also not a solution to the construct validity problem.

Other assessor cognitive processing issues, though not explored in this study, remain as possible explanations for the findings. Understanding how assessors observe and what kinds of assessee behaviors tend to capture the attention of assessors may be a productive research pursuit. This research has been called for (Russell, 1985; Zedeck, 1986) within assessment centers and has been pursued within performance appraisals. For example, within performance appraisals, there is evidence that purpose (Zedeck & Cascio, 1982), individual differences in selective attention (Cardy & Kehoe, 1984), and initial impressions (Balzer, 1986) affect information acquisition by assessors. It is generally assumed within assessment centers that assessors attend to behavior at a discrete, elemental level where the observation of behavior is a "yes, the behavior of interest occurred"/"no, it did not occur" decision and where the classification of behavior is a matter of assigning the behavioral incident to the appropriate category. A similar assumption was made in this study; namely, that assessors attend to behaviors at a level of detail equivalent to the elemental and discrete level of the behaviors defined in the scoring guidelines. Social cognition researchers, however, suggest human perceptual processes and attention is much more molar, impressionistic, and inferential driven by complex attributional processes. Rather than mere recorders of objective reality, "people create meaning and add it on to

raw data of the objective world" (Fiske & Taylor, 1984, p. 141).

Zedeck (1986) interprets assessor behavior within a social cognition perspective. Rather than considering observation a matter of a yes/no decision, he points out that assessee behaviors differ in their representativeness of the prototypical behaviors defined by scoring guidelines. Rather than considering categorization a matter of assigning observed behaviors to discrete categories, he suggests that the boundaries between dimensions are often fuzzy. The result, and a potential explanation of the high correlations among dimensions according to Zedeck, is that assessors participate in a decision making process that is probabilistic in nature, and that certain behavioral examples may load on more than one dimension. While this study attempted to eliminate some of the fuzziness in categorization with scoring checklists, the results were vulnerable to any assessor tendencies not to observe behaviors at an elemental and discrete level. Given the relatively short time period of an exercise, only a few critical behavioral clusters or incidents may govern a broad range of assessor checkmaking and rating behavior.

The previous discussion is not to say that raters are in error in the way they observe and rate participants in assessment centers. On the contrary, the criterion-related validity of assessment centers suggests that raters, despite not conforming to the rating process as expected, assess candidates similar to the way employees are assessed on important job criteria. In addition, the finding of acceptable levels of inter-rater reliability in this study and in other assessment center studies, suggests assessors seem to view assessees relatively consistently, regardless of the assessor observational and cognitive

processes that are occurring.

Regarding the form and content of exercises, this study suggests that form accounts for 16 percent of method variance whereas the effect of exercise content, defined as cooperative and competitive task designs, is negligible. Thus, exercise form represented a method construct that discriminated among assessees whereas exercise content did not. These findings have practical implications for centers defended on the basis of their content validity. On one hand, the finding that form accounts for variance in this study emphasizes the importance of selecting the form of an exercise in a way that models the kinds of interpersonal relations an assessee will encounter on the job. If a job demands that incumbents work in one on one situations or work on teams, the form of exercises should reflect these distinctions (i.e. role plays vs. group discussions). On the other hand, it seems less critical, based on these findings, to create cooperative and competitive tasks in exercises to simulate job characteristics. While content validation may suggest competitive exercises for assertive, confrontational jobs (i.e. salesperson) and a cooperative task design for team oriented, collaborative jobs (i.e. manager), these exercise task designs do not seem to result in the expected variance in assessee performance. An additional, practical, implication of exercise design research involves the creation of parallel forms. These data suggest that the creation of parallel forms assessment center exercises should consider the form of the exercise yet not the content or task design of exercises.

Further research determining how different exercise characteristics result in assessee performance variance would assist the exercise

creation and creation of parallel forms. It is common in the development of traditional, paper and pencil, intelligence and personality tests to discard items that do not discriminate among assesseees and maintain test items that result in the greatest amount of variance among assesseees. Similarly within assessment centers, more research manipulating exercise designs may aid in the understanding of the key design determinants of assessee performance variance. This kind of research may identify and distinguish essential exercise characteristics for creating performance variance (i.e. form) from non-essential exercise characteristics that do not account for performance variance (i.e. content). Some of the uncontrolled exercise features in this study might form the basis for a taxonomy of key exercise design features including: the organizational context of the problem (retail, school, and manufacturing); the types of problem solving encountered (mechanical, interpersonal, negotiation, and selling); number and kind of other participants (one standardized role player and three assesseees); and level of goal ambiguity/amount of structure (specific plans and action steps, and creative, open ended goals). These features may well account for some of the unexplained method variance in this study and may represent areas for further research in exercise design.

Finally, this study revealed higher trait intercorrelations across role plays. This is not surprising given that the role play held much more situational variance in check compared to the group exercises. The confederate playing a standardized role (as opposed to group exercises in which participants interacted with each other) represented a form of control of situational variance, and may have allowed for assesseees personal characteristics to have a greater effect on performance

relative to situational variables. Other participants in a group exercise may represent a large source of situational variance thus dwarfing any trait variance. There is also suggestive evidence that the low task interdependent role play provided better evidence of trait factors suggesting that more trait variance may be generated by exercises where assessees perform independently and alone.

It makes conceptual sense that controlling sources of extraneous situational variance in assessment center exercises should increase the likelihood that performance differences will be due to individual characteristics. Evidence for stable individual characteristics such as general cognitive ability has come from tests that are highly standardized in administration and scoring. Introducing a similar level of standardization and control may improve the measurement capability of assessment centers. For example, Schmitt, Schneider, & Cohen (1990) found that different center administration practices across different geographical locations can moderate validity; this finding emphasizes the importance of standardization and control in the administration a center. It follows from this study that exercise designs which promote the standardization of situations across assessees, particularly by controlling the extraneous influence of other participants, may well result in improved measurement of personal qualities in assessment centers.

APPENDICES

APPENDIX A
CORRESPONDENCE TO PARTICIPANTS

Dear Parent:

We are offering your child the opportunity to participate with 100 other students from _____ in an Employability Skills training workshop. During the workshop, all of the students will participate in four employability skills exercises in which they will perform tasks similar to the kinds of tasks that they will encounter in a first job. All students will individually receive helpful suggestions aimed at improving their job skills as a result of their participation in the exercises. Participation in the program will be according to a first come, first serve, volunteer basis. All participants can benefit greatly.

The project is jointly sponsored by the _____ and _____. The exercises will be videotaped and viewed as part of a research study about employability skills exercises. Your child's participation will have no bearing on your his/her future academic or employment status. All results will be treated with strict confidence and your child will remain anonymous in any report of research findings. The overall results of the study will be available to you on request within the limits of confidentiality.

The exercises take approximately three hours and will take place at _____ between April 12 and May 18, 1990. Participation is supported by the schools' administration, principal, and teachers, so your child will be excused from classes. There will be no penalty for participating or for refusing to participate in the study. Your child's participation is entirely voluntary and he/she can choose not to participate or to stop participating at any time during the workshop without penalty. Your child will be informed of this right at the workshop. The purpose of the study will be explained in even more depth after the exercises, and the researchers will be glad to answer any questions during that time or at any other time before or after the workshop.

If you have any questions call Jeff Schneider at Michigan State University (517-355-2171).

Sincerely:

Principal
High School

I have read and understand the above statement. My child has my full consent to participate in this research. I understand that he/she may discontinue participation at any time without penalty. I have explained and discussed the research and the rights associated with voluntary participation with my child. I authorize the researchers from Michigan State University to utilize my participation in the study for teaching and research purposes. Part of the experiment will require video-taping, and I authorize video-taping these sessions. I understand that my child's identity will be held confidential unless I authorize its use.

Signature of Student

Signature of Parent or Legal Guardian

APPENDIX B
DIMENSIONS AND DEFINITIONS

Problem Solving Skills:

- seeking all available information about problems
 - asking incisive questions
 - shows natural curiosity for information
- integrating available information to solve "whole" problems
 - puts information together like pieces in a puzzle
 - sees relationships among items in background information
- recommending many alternative solutions to problems
 - brainstorms many ideas
 - shows creativity by offering many diverse yet workable ideas
- recommending solutions and making decisions through logical reasoning and analysis
 - offers logical support for ideas and decisions
 - evaluates solutions in terms of pro's and con's, short and long term implications and benefits.

Interpersonal Skills:

- reinforcing and supporting others' ideas
 - saying "good job," "I like your idea," "you do good work"
- promoting teamwork
 - asking questions to understand others' ideas and points of view
 - listening to others with an open-mind and without criticism
- promoting positive rapport among teammembers
 - showing enthusiasm for working with others
 - engaging in small talk to break the ice when first meeting people.
 - encouraging the participation of quiet teammembers
 - using sense of humor
- compromising in support of team objectives
 - supporting solutions that meet needs of all teammembers
 - promoting a spirit of compromise rather than self-centered, arrogance
- non-verbally demonstrating warmth and openness
 - maintaining eye contact when listening or speaking
 - smiling, warm facial features
 - maintaining an attentive body posture--leaning forward, etc.

Initiative Skills:

- demonstrating motivation through active participating
 - remaining active and involved for the duration of exercises
 - consistently working at an above average pace
- stating ideas confidently and directly
 - using "I" language--"I think," "I feel," etc.
 - committing decisively to a position or point of view
 - sticking to good ideas even in the face of criticism
- setting direction and goals for others
 - providing direction to group regarding how to complete tasks
 - helping groups get unstuck by suggesting new directions
- emphasizing the importance of achieving assignments
 - monitors progress toward goals
 - pays attention to time considerations

APPENDIX C

EXERCISES

GRANT ALLOCATION EXERCISE

This exercise is made up of two segments. First, there will be a 10 minutes period where each of you, individually, will prepare for a meeting. Then, all four of you will meet together for 18 minutes to complete the following assignment.

LET ME GIVE YOU SOME BACKGROUND INFORMATION

Your School has been awarded a \$15,000 grant that can be used to for any student-related purpose. All of you are members of a committee who will decided how the money will be spent.

As a committee member, each of you are required to present an idea about how to use the \$15,000. You have three options for selecting an idea:

- 1.) You may come up with your own idea
- 2.) You may each pick one of the ideas on the table and present it, or
- 3.) You may pick an idea from the table, make changes, and present your new version of the idea.

You are encouraged to use your imagination.

You should use the rest of the preparation period to prepare a 2 minute oral presentation about the idea which you will give to the other committee members. Remember you should build a good case for the idea since you will have to convince the other group members to accept it.

I have set a timer so that you can keep track of time. The timer will beep when you have five minutes left.

(AFTER THE 10 MINUTE PREPARATION PERIOD)

You will now begin the meeting. Each of you should give your presentation, then together, all of you should discuss the ideas and decide how to allocate the money.

There are some rules for allocating the money.

--The group may allocate all the money to one idea.

or

--The group may distribute the money among two ideas, yet both ideas cannot receive the same amount of money.

--YOU MAY NOT allocate money to more than two ideas.

You have 18 minutes. The timer will beep when you have 10 and five minutes left.

Grant Allocation Exercise

GRANT ALLOCATION IDEA #1

STUDENT ACTIVITIES FUND

You believe that the best way to spend the \$15,000 will be to establish a general fund that can be used to pay for special student activities.

Any student groups who want money from the fund would apply by submitting an explanation of the project and a budget to a Student Activities Fund (SAF) committee. The SAF committee would be responsible for distributing the money.

Money could go toward sponsoring:

- A social event such as a school dance or picnic.
- A guest speaker--famous author, athlete, etc.
- Clubs or Teams--Debate, Sports, Band, Choir.

(Think of other examples to add to this list).

Strengths of this idea

- All students could benefit from the use of the money.
- Students applying will learn how to sell ideas and budget money.
- Committee members will learn about responsible decision making.
- Interest could be earned on money until it is spent.

Weaknesses of the idea

- No long term benefits--money could run out in a year.
- Arguments over the fairness of committee decisions could result.

Grant Allocation Exercise

GRANT ALLOCATION IDEA #2

SNACK BAR

You believe that the money should go toward building a snack bar.

The snack bar would offer students alternatives to the regular lunch plan including cold sandwiches, pizza, soda pop, soup, salads, french fries, popcorn, ice cream, candy bars, fruit etc.

Strengths of the Idea

- More food alternatives would be offered to students
- Would make money and support itself
- Could be used during school dances, basketball games, etc.
- Could be staffed by students

Weaknesses of the Idea

- Some students would only eat junk food.
- Snack Bar specialty items are often high priced
- Students may spend money that they should save for more important purposes.

Grant Allocation Exercise

GRANT ALLOCATION IDEA #3

EDUCATIONAL SCHOLARSHIP

You believe that the \$15,000 should be offered as a scholarship for students who plan to continue their education after high school (college, vocational school, trade school, etc.).

The scholarship could be offered on the basis of financial need to students who can not afford education after high school or on the basis of academic achievement.

The whole sum of money, \$15,000, may be offered to one student or may be divided among students. For example, five students may receive \$3,000 or \$3,000 may be given out each year over a period of five years while the rest of the money earns interest in the bank.

Strengths

- Will help students achieve educational or career goals.
- May help someone who cannot afford additional education.
- Reinforces the importance of doing well in school if money is given for academic achievement.
- Goes to a good cause--education.

Weaknesses

- Only a few students will benefit--hard to decide which students should get the money.
- No long term effects--money will be spent in one year.

Grant Allocation Exercise

GRANT ALLOCATION IDEA #4

COMPUTER LAB

You believe that the money should be used to fund a computer lab that is open during school time, afternoon, and some evenings. The lab will contain personal computers that students can use to write papers or for programing. Students could also use educational software programs for specific skills training in Math, Science, Social Studies, or English.

Strengths

- Students will have computer access during off hours--students who do not have a computer at home will benefit greatly.
- All students will have access to these computers.
- Students will learn how to use computers.
- Computers will be used for training in Math, Science, etc. to supplement classroom teaching.
- Long term benefits--the lab will be used over many years.

Weaknesses

- Additional money will be needed in the future to keep educational software programs up-to-date.
- Only two students will be able to use the computers at a time.

Grant Allocation Exercise

GRANT ALLOCATION IDEA #5

PARKING LOT

You believe the \$15,000 should go toward expanding blacktopped parking area. This will allow for additional parking spaces for students during school and for event parking--concerts, sports, plays, etc.

Strengths

- More students will get parking permits and will be able to drive to school.
- Many students stand to benefit.
- Parking will be easier to find during events such as sports, concerts, meetings, etc.

Weaknesses

- Does not have long term benefits since a New High School will open soon.
- Does not eliminate the problem of deciding who gets permits. Not all Students will get permits.

TEAM MANUFACTURING EXERCISE

Instructions to be read to candidates:

You all have been assigned to an assembly team at a manufacturing company. Your team is assigned to assemble industrial fans.

There are folders on the table describing the different jobs that need to be done: 1.) Parts and Supply, 2.) Frame Assembly, 3.) Moving Parts #1, 4.) Moving Parts #2 and 5.) Quality Control. The Final assembly folder explains how to put the final product together from all the different parts.

You should work as a team and may share jobs so do not feel that you must each work on separate jobs.

You may also make any changes in the fans you wish. Remember, this is a new product so the design is somewhat untested.

You will sell each product to me for \$200. If you improve the fan you also should work together to recommend changes in its price. I may also pay you less the \$200 per fan if the fans have any defects. Your teams' goal is to make \$ 2000 which means selling 10 units at \$200.

Any questions?

You have 25 minutes to work together on this task.

Team Manufacturing Exercise

PARTS AND SUPPLY

The following steps are needed to fill the parts and supply role

1. Sort all parts into separate piles
2. Distribute the following parts (for one fan unit) to:

Frame Assembly

7 Green Rods
8 S1 Spools
6 Blue Rods
1 S4 Orange Sleeve

Moving Parts #1

4 Blue Rods
2 Yellow Rods
1 S1 Spool
4 Yellow or Green Fins

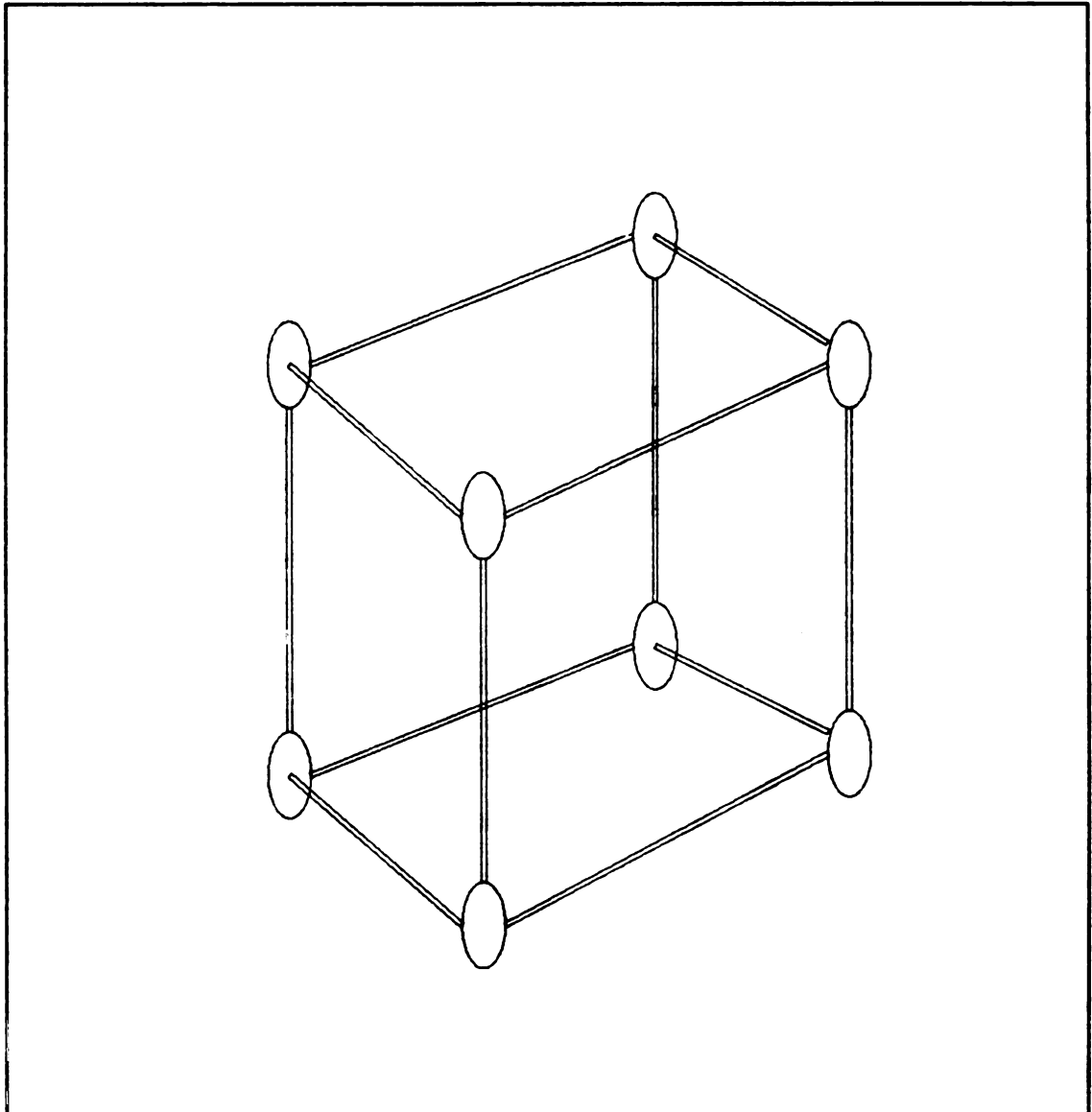
Moving Parts #2

2 S5 Spools
1 S3 Spool
4 Red Rods
1 Orange Rod

Team Manufacturing Exercise

FRAME ASSEMBLY

The role of frame assembly is to assemble frames for the fan according to the attached specifications.

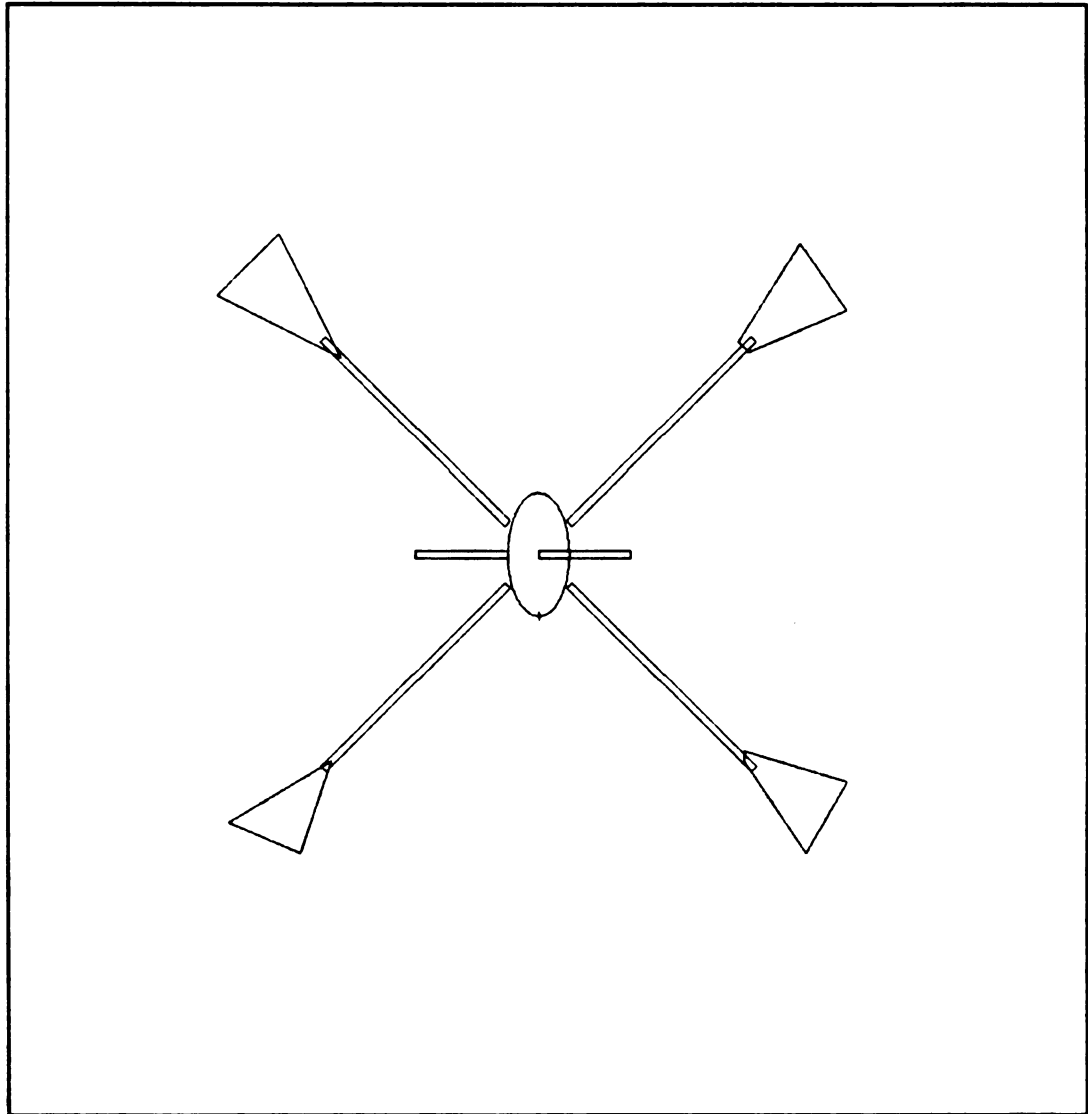


Team Manufacturing Exercise

MOVING PARTS #1

The role of moving parts #1 is to assemble the fan and axle according to the attached diagram.

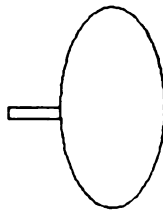
You may use yellow or green fins.



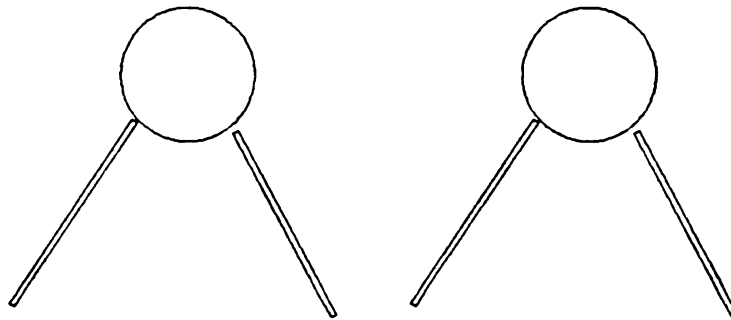
Team Manufacturing Exercise**MOVING PARTS #2**

The role of moving parts #2 is to assemble the Fan Axle Supports and the Speed Control according to the attached diagram.

Control Knob



Fan Axle Supports



Team Manufacturing Exercise

QUALITY CONTROL

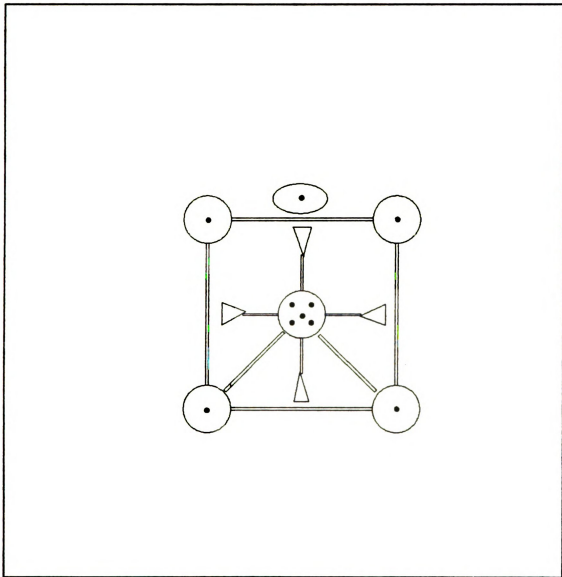
The job of quality control is to run the following tests once the fan is assembled.

1. Ensure that the frame is assembled tightly, making sure that the fan does not wobble in any way.
2. Spin the fan blades so that it rotates freely.
3. Turn the speed control dial around one complete revolution. Then set the dial to off.

Team Manufacturing Exercise

FINAL ASSEMBLY

The role of final assembly is to place the fan into the frame and attach the speed control according to the attached specifications.



CUSTOMER SERVICE EXERCISE

Instructions

This exercise has two parts: 1.) a 12 minute preparation period, and 2.) a meeting with a co-worker that may last up to 15 minutes. You are responsible for making sure you complete the following assignment.

BACKGROUND INFORMATION:

You are an employee who works in the customer service department of a department store. Your job is to review customer service complaints. You worked for this store as a sales clerk before you were assigned to the customer service department.

The attached file includes the results of your investigation of a complaint registered by Mrs. Tilly Brown. The complaint has been registered against a sales clerk, Bert Landy, and Bert's supervisor, Mary Perkins.

In 12 minutes, a co-worker will meet with you to discuss the complaint, and together, you should decide what should be done. Your co-worker has also investigated the complaint and may not have the same information as you. There is a Complaint Action Form in the file which will give you an idea about the kinds of action you may take.

POINTS TO REMEMBER

During the preparation time, before the meeting with your co-worker, you should:

1. read this material thoroughly
2. have an idea about the actions you want to take.

You may take notes on scratch paper, and you may use the notes and the file during the meeting.

During the 15 minute meeting, you and your co-worker, together, should:

1. discuss the problem using file information.
2. complete the Complaint Action Form.

Customer Service Exercise

Mr. Melvin Burns
President, Department Store

Dear Mr. Burns:

I was treated very poorly while purchasing a watch at your store. I asked that a new battery be placed in the watch I purchased. The sales clerk put a new battery in the watch but then he charged me for it. He never told me there would be a charge, and I wouldn't have known it if I hadn't noticed it on the sales receipt. What nerve! I think its only natural that a new watch comes with a new battery.

I really got upset when he said "We aren't responsible for our products." Then even worse, his supervisor came over and told me "old lady, you are out of hand. Leave or I will call security."

I have been shopping at your store for 20 years and deserve better treatment. I expect you to fire the clerk and the supervisor.

Sincerely:

Mrs. Tilly Brown

Customer Service Exercise

MEMORANDUM
(in student's file only)

To: Customer Service Committee

From: Jan Durkheim, Switchboard Operator

Subject: Phone call regarding Mrs. Brown incident

I received a call from an upset customer, Homer Childs, who witnessed the Tilly Brown incident.

Childs said that Mrs. Brown was quite demanding about the new battery and that our clerk was very patient with her. Mrs. Brown had some trouble hearing which made the problem worse.

Mr. Childs was very upset about the behavior of the supervisor. The supervisor did not treat Mrs. Brown with respect and threatened to call security.

Customer Service Exercise

MEMORANDUM
(in role player's file only)

To: Customer Service Committee

From: Stella Bodine, Store Manager

Subject: Tilly Brown

I am aware that you are discussing an incident involving Mrs. Tilly Brown. I thought you should know that this is the fourth incident involving Mrs. Brown in the last three months. She is a chronic complainer and most of her complaints have little basis.

Customer Service Exercise

CUSTOMER SERVICE REPORT FORM

EMPLOYEE FILING REPORT: Mary Perkins, Jewelry Department Supervisor

CUSTOMER'S NAME: Mrs. Tilly Brown

BRIEFLY DESCRIBE THE INCIDENT:

I responded to this incident when I heard Mrs. Brown yelling loudly at my employee, Bert Landy. Mrs. Brown was completely out of hand. She is in the store a lot but never buys anything; she mainly complains. She made a big fuss over a cheap, \$25 watch.

BRIEFLY DESCRIBE ANY ACTIONS YOU HAVE TAKEN TO RESOLVE THE INCIDENT:

I told Mrs. Brown that her behavior was unacceptable and that she must leave the store immediately. I expressed my appreciation to the sales clerk, Bert Landy, who handled Mrs. Brown without getting upset.

Customer Service Exercise

CUSTOMER SERVICE REPORT FORM

EMPLOYEE FILING REPORT: Bert Landy, Sales Clerk
CUSTOMER'S NAME: Mrs. Brown

BRIEFLY DESCRIBE THE INCIDENT:

Mrs. Brown bought a watch and was quite worried that the battery would wear out during her upcoming vacation. She wasn't satisfied until I offered to replace the battery. She became upset when she found the charge on her sales slip. I tried to calm her. Ms. Perkins finally asked Mrs. Brown to leave.

BRIEFLY DESCRIBE ANY ACTIONS YOU HAVE TAKEN TO RESOLVE THE INCIDENT:

I saw Mrs. Brown in the store a few days later (I don't know what happened to her vacation) and apologized.

Customer Service Exercise

COMPLAINT ACTION FORM

1. Which of the following actions do you recommend to take with the customer, Tilly Brown?

- ☐ Customer should receive current purchase at no charge
- ☐ Send customer gift certificate good for next purchase
- ☐ Explain to customer that her behavior has caused difficulty for store.
- ☐ Close customer's account

2. Which of the following actions do you recommend to take with the sales clerk, Bert Landy?

Training

- ☐ 30 minute meeting with customer service trainer to discuss what could have been done more effectively.
- ☐ Attend extensive (one week) customer service training course

Reward for good performance

- ☐ Written commendation in personnel file
- ☐ Recommend for employee of the month

3. Which of the following actions do you recommend to take with the supervisor, Mary Perkins?

Disciplinary Action

- ☐ Suspension from work without pay
- ☐ Written warning in personnel file
- ☐ Dismiss (fire) employee

Training

- ☐ 30 minute meeting with customer service trainer to discuss what could have been done more effectively.
- ☐ Attend extensive (one week) customer service training course

Responses to Customer

- ☐ Employee should apologize to Customer

Reward for good performance

- ☐ Written commendation in personnel file
- ☐ Recommend for employee of the month

TEAM SELECTION EXERCISE

Instructions

This exercise has two parts: 1.) a 12 minute preparation period, and 2.) a meeting with a co-worker that may last up to 15 minutes. You are responsible for making sure you complete the following assignment.

BACKGROUND INFORMATION

You are an employee of a company which makes frozen pizzas. Pizzas are assembled by teams of four people including: a Team Leader and three workers--a Pastry Chef, a Freezing Scientist, and an Ingredients Specialist.

You have just been appointed as a Team Leader and need to select three workers for your team. You will select these workers from the information in the attached file. The qualifications for each job and two candidates for each job are described.

In 12 minutes, you will meet with a co-worker who is also a newly appointed team leader who also must select three workers for a team from the same file. During the meeting, your assignment is to divide up the candidates into teams. You and your co-worker should work together to assign one of the candidates to your team and the other candidate to the other leader's team.

TEAM SELECTION EXERCISE

SUGGESTIONS FOR SELECTING TEAM MEMBERS

1. You will note that each candidate is rated on an Overall Qualification Scale. This is an index of how well qualified the candidate is for the job.
 - 1 = Well below average
 - 2 = Below average
 - 3 = Average
 - 4 = Above average
 - 5 = Well above average
2. Adding qualification ratings of individuals on your team will give you overall rating for the team:
 - 3-5 a poor team
 - 6-9 an acceptable team
 - 10-15 an outstanding team
3. Though you should consider these ratings, you should also consider which candidates will work best with each other.

HERE ARE SOME POINTS TO REMEMBER

Before the meeting, during your preparation time, you should:

1. Read the material thoroughly
2. Have an idea about which candidates you want and why you want them.

You may take notes on scratch paper, and you may use the notes and the file during the meeting.

When your preparation time is up, a co-worker will come into the room to meet with you. You and your co-worker together, should:

1. discuss the candidates' qualifications by job.
2. assign one candidate for each job to each team.

YOU MAY ASK ANY QUESTIONS YOU WANT

Team Selection Exercise

PASTRY CHEF

Pastry Chef mixes dough and makes pizza crusts. Experience is critical for this position.

Bill Sargent

Experience: 15 years in 6 different food companies

Strengths: Self disciplined, hardworking, above average productivity. Has good ideas for making crusts better. Provides excellent advice, but he can make others upset in the way he gives advice.

Weaknesses: Has trouble getting along with others: is impatient with inexperienced workers; tells others how to do their jobs; insists that things must always be done his way. Has never worked for one company longer than 3 years as a result.

Overall Qualification Rating: 5 well Above Average

Sally Shy

Experience: 5 years; all with the current company

Strengths: Productivity is average to above average. Hardworking and will pickup the slack for other team members. Good sense of humor; is well liked by co-workers.

Weaknesses: Quiet--her good ideas can go unnoticed.

Overall Qualification Rating: 4 Above Average

Team Selection Exercise

FREEZING SCIENTIST

Freezing Scientist--packages and freezes the pizza. A food science degree is necessary, and experience is an advantage.

Larry Likeable

Experience: Associate degree (2 years) in Food Science. Only 2 years experience as a freezing specialist, but 15 years experience in food business.

Strengths: Average productivity and quality. Dependable and likeable. Knows the food business.

Weaknesses: Not a leader at all, is easily pushed around by others. Relatively inexperienced as a freezing specialist.

Overall Qualification Rating: 3 Average

Mark Blaze

Experience: Masters Degree (6 years) in Food Science. 8 years as a freezing specialist with various companies, 1 year at the current company.

Strengths: Intelligent--well educated. A good teacher for new employees. Works well with quiet cooperative team members. Productivity is above average, at times.

Weaknesses: Has a history of frequent absences from work. Was fired from last two jobs because he missed too much work, but has been more dependable over the last year. Has trouble getting along with team members who "think they know everything."

Overall Qualification Rating: 2 Below Average

Team Selection Exercise

INGREDIENTS SPECIALISTS

Ingredients Specialist inspects quality of ingredients and places sauce, toppings, and cheese on the crust. Experience is not critical since these skills can be trained.

Madge Beauty

Experience: Owned a beauty salon for 10 years. Had to sell the business because of financial problems last year. Has never worked in food preparation, but has worked as a temporary secretary at the current company for the last 6 months.

Strengths: Dependable and well liked. Comes with high recommendations based on the past 6 months.

Weaknesses: No experience. Productivity is unknown. Needs to work with co-workers who will be patient while she learns the job.

Overall Qualification Rating: 2 Below Average

Paula Potter

Experience: 3 years as Ingredients Specialist at the current company

Strengths: Above average productivity. Highly dependable--has never been absent from work in three years. Has the ability to get along well with a wide range of people.

Weaknesses: Has trouble making ingredients inspection decisions. Has occasionally allowed bad ingredients on pizzas. Needs some one to turn to for advice.

Overall Qualification Rating: 4 Above Average

APPENDIX D
ROLE PLAYER TRAINING MATERIALS

ROLE PLAYER TRAINING SCHEDULE

WEEK 1 Group Training Sessions

- Purpose of Project
- Dimensions and Exercises
- Assessor Role Sets--video modeling

WEEK 2 Video Training

- Work in pairs playing assessor and student roles
- Group viewing of videos and group feedback

WEEK 3 Video Training with Pre-test students

- Role playing with "live" students
- Group viewing of videos and group feedback

CUSTOMER SERVICE EXERCISE

ASSESSOR ROLE SET

Overall Personality

You should begin this role play with a firm view of the problem, namely that Mrs. Brown is a chronic complainer and trouble maker. Add that you might tolerate Mrs. Brown if she spent more money at the store. You should recommend that her account be closed.

You should take a firm stance on your information and remain firm for the first 5 minutes of the role play. You should question the validity of the information from Homer Childs (which only the student has) especially relative to the information from employees.

Take a more conciliatory position, if the student reinforces the importance of the Homer Childs information and gives good reasons such as "the customer is always right."

Specifics

Opening: Engage the student in small talk/idle chatter--i.e. about customer service problems you've had etc. Reminisce about your experience as a clerk and talk about how customers can be a real pain sometimes. Ask student about his/her experiences.

0-5 Minutes: State your position that Mrs. Brown is a chronic trouble maker and that her account should be closed. Reasons for your position include:

- she is a chronic complainer with a history of petty complaints
- she does not spend money--no big loss if account is closed
- a business can't cater to complaints from customers who aren't profitable.
- She can't hear and makes us look bad

If student shows you the phone message from Homer Childs, you should discount the information saying, "I didn't receive that information."

5-15 Minutes: Begin to be willing to compromise and accept phone message from Homer Childs.

Wait to see if the student takes responsibility for completing the form. Take issue with some of the students' recommendations and ask for the students' reasoning behind the recommendations.

Initiate discussion about the form if the student does not.

CUSTOMER SERVICE EXERCISE

ASSESSOR ROLE SET (Cont.)

Key Points to Remember

You think Mrs. Brown's account should be closed because:

- she is a chronic complainer.
- she does not spend money--no big loss if account is closed
- She can't hear and makes us look bad

Your position is based only on the employees information and therefore makes the employees look more favorable.

The truth of the matter is revealed in the Childs memo--Mary Perkins threatened Mrs. Brown. Perkins probably should not be fired; A training course in customer relations and a note in personnel file is a reasonable recommendation for Perkins.

Subjects will probably overlook Bert Landy's error--he charged Mrs. Brown without telling her about the charge. Bert should probably be given some instruction about this as well as a commendation for a job well done.

Make Sure You Do The Following

1. Engage in conversation with student to put student at ease. Talk about your experiences when you were a clerk at this store--ask a few questions about students experience.
2. Be talkative--do not just ask the student questions.
3. Avoid making any suggestions for how to proceed.
4. Mention the problems with Mrs. Brown--four incidents in the last three months--but do not immediately reveal that you have an additional memo.

TEAM SELECTION EXERCISE

ASSESSOR ROLE SET

Overall Personality

During this exercise, you should attempt to get the best team. Be firm about the candidates you want; Add up the ratings on each team so that it is obvious who has the better team.

You should discuss the candidates one job at a time. Encourage student to reread the descriptions if necessary. Encourage the student to go first each time, allowing him/her to state her reasons for selecting a particular candidate.

If student wishes to reject any of these candidates, suggest that it may be possible to get additional candidates but that you don't want to leave the meeting without any recommendations.

Specifics

Opening Engage in some small talk--i.e. you are looking forward to this assignment; you have looked over the candidates and have an idea who you way.

Wait to see if student takes initiative with the task, if nothing happens by 2-3 minutes, say something like "Why don't we discuss one job at a time. Why don't you go first."

0-5 Minutes State your initial preferences. Challenge the students reasoning (or lack of reasoning) for preferring particular candidates.

Try to talk students into taking weaker candidates (Bill, Mark, Madge) by diminishing their weaknesses and emphasizing their strengths.

5-15 Minutes

Begin to compromise, especially if student's reasoning is sound. If you are at a standoff, you may initiate a compromise by saying something like "I'll take Madge if you want Paula"

If the subject has not mentioned that matches and mismatches between specific candidates, mention one of these such as "Madge probably will not work very well with Bill."

TEAM SELECTION EXERCISE

ASSESSOR ROLE SET

Initial Preferences: Sally, Larry, Paula.

Sally--will get along better with others.

If student does not want Bill:

Say Bill has higher rating, better experience, etc.

Larry--gets along well and has a higher rating.

If student does not want Mark:

say Mark has a better education (important for job) and has had better attendance over the last year.

Paula--better rating and hard worker

If student does not want Madge:

Say that Rating is not a true indicator because Madge has not worked. Madge is well liked and may be a "diamond in the rough."

Key Details to Remember:

1. Bill should not be with Mark or Madge
2. Mark could work with Sally and Madge

Ideal solution 1.) Mark, Sally, Madge 2.) Bill, Larry, Paula

Make sure you do the following

1. Engage in conversation with student to put student at ease. Talk about your experiences when you were an ingredients specialist--ask a few questions about students experience.
2. Be talkative--do not just ask the student questions.
3. Avoid making any suggestions for how to proceed. e.g. do not suggest putting a decision on hold and discussing other candidates. do not offer compromise solutions.
4. Take one relatively strong stand. e.g. refuse to take Bill at first. You may compromise after you take a stand.

APPENDIX E
ASSESSOR TRAINING MATERIALS

ASSESSOR TRAINING SCHEDULE

Day One

1. Purpose of project
 - History of employability skills
 - Purpose of ratings--to evaluate program not students
 - Issues of confidentiality--commitment to program
2. Background of Assessment Centers
 - Work sampling approaches
 - History/validity of centers
 - Development applications
3. Dimensions
 - How dimensions were selected
 - Group discussion of definitions
4. Exercises
 - Brief intro of each
 - Cross of dimensions and exercises--dimension/exercise grid
5. Observing behavior
 - Use of checklists
 - Adding own observations to checklists--emphasize behaviors
 - Introduce Grant Allocation Exercise
 - Observe video example-assessors practice using checklists
 - Group discussion of observations and checks

ASSESSOR TRAINING SCHEDULE (CONT.)

6. Rating

- Distribution of Ratings--Normal Curve
- Rater Errors (Demonstrate Each)
 - Halo, Leniency, Central Tendency
- Rate above Grant Allocation Exercise
 - Training in use of rating forms
- Group discussion of observations and ratings
- Observe, rate, and discuss another grant allocation example

7. Introduce Team Manufacturing Exercise

- Observe example, check, rate, discuss

Day Two

8. Introduce Team Selection Exercise

- Observe example, check, rate, discuss

9. Introduce Customer Service Exercise

- Observe example, check, rate, discuss

APPENDIX F
SCORING GUIDELINES

EXCEL RATING FORM

STUDENT# _____ ASSESSOR# _____

EXERCISE _____ TEAM SELECTION (SELECTING PIZZA MAKERS)
 _____ CUSTOMER SERVICE (WATCH COMPLAINT)
 _____ TEAM MANUFACTURING (TINKER TOYS)
 _____ GRANT ALLOCATION (\$15,000)

RATING

	WELL BELOW AVERAGE		TYPICAL		WELL ABOVE AVERAGE
PROBLEM SOLVING	1	2	3	4	5
INTERPERSONAL SKILLS	1	2	3	4	5
INITIATIVE SKILLS	1	2	3	4	5

HOW MUCH SCHOOL-RELATED CONTACT HAVE YOU HAD WITH THIS STUDENT?

_____ None at all.
 _____ Brief, intermittent contact (ex. Like a librarian would have).
 _____ In class or extracurricular activity for 1 quarter/semester.
 _____ In class or extracurricular activity for 1 quarter/semester to 1 year.
 _____ In class or extracurricular activity more than 1 year.

HOW MUCH NON-SCHOOL CONTACT HAVE YOU HAD WITH THIS STUDENT?

_____ None at all.
 _____ Know casually from community (ex. Church).
 _____ Know well from community (ex. Family friend).

OVERALL, HOW WELL DO YOU KNOW THIS STUDENT

_____ Not at all, don't even recognize.
 _____ recognize student, know his/her name.
 _____ know name, know student from reputation.
 _____ know student well from school/community contact.
 _____ know student extremely well from school/community contact.

Grant Allocation
Scoring Guidelines

PROBLEM SOLVING

HIGH

- ___ asks questions to understand others' ideas more fully.
- ___ pulls many different ideas together to form a consensus solution.
- ___ offers decision criteria to help the group make decisions--ex. how many students will benefit, how many years will the money last.
- ___ creates an idea on own rather than using generic ideas.
- ___ lists many benefits of idea.
- ___ gives logical reasons for why others should allocate money to his/her idea.
- ___ uses sound logic in reviewing others' ideas--is quick to find problematic issues in others' ideas.

TYPICAL

- ___ occasionally asks questions of others.
- ___ integrates pieces from a few different ideas together.
- ___ may select a generic idea but improves it somewhat.
- ___ gives a few reasons supporting own idea.

LOW

- ___ does not ask questions--offers very little critique of others ideas.
- ___ idea is not well thought out, seems superficial with few benefits.
- ___ uses a generic idea with little thought or embellishment.
- ___ gives brief statement of idea without giving any supporting logic or benefits.
- ___ throws out opinions without supporting reasons or logic.

Grant Allocation
Scoring Guidelines

INTERPERSONAL SKILLS

HIGH

- _____ frequently states areas of agreement with others.
- _____ frequently praises and supports good points in others' ideas.
- _____ asks questions from a position of genuine interest--does not just ask questions with the purpose of discrediting others' ideas.
- _____ encourages quiet members to share their ideas.
- _____ fosters communication through asking questions and sharing information.
- _____ listens to others' ideas without abruptness or criticism.
- _____ compromises and lends support to others' ideas.
- _____ smiles, seems open and warm toward others.

TYPICAL

- _____ shows sensitivity to the personal impact of ideas--discusses how \$\$ can go to humane purposes.
- _____ offers occasional support praise of others' ideas.
- _____ participates, cooperates, does not offend.
- _____ respects others when they have the floor--does not interrupt.
- _____ maintains eye contact with others.

LOW

- _____ asks questions in a derogatory manner.
- _____ tends to "gang up" in criticizing particular individuals.
- _____ criticizes while offering no praise or support.
- _____ uses humor in a derogatory or sarcastic manner.
- _____ puts down others or their ideas
- _____ dominates conversation with own ideas little listening or tolerance for others' ideas.
- _____ seems unwilling to compromise or accept others' ideas.
- _____ seems non-verbally withdrawn or aloof.

Grant Allocation
Scoring Guidelines

INITIATIVE

HIGH

- _____ is highly active throughout the duration of the exercise.
- _____ pushes for own idea to win.
- _____ repeatedly emphasizes the merits of own idea.
- _____ sticks with own idea even in the face of criticism.
- _____ provides some structure or direction for how the group should proceed.
- _____ frequently restarts or redirects group when group gets stuck.
- _____ focuses on a quality decision rather than "just getting finished"

TYPICAL

- _____ remains involved in discussion from start to finish.
- _____ gives at least one strong push for own idea to win.
- _____ speaks in a relatively direct and assertive manner--some hedging.
- _____ may not set direction for group but does not need prompting to participate.

LOW

- _____ speaks very little, seems uninvolved and non-participative.
- _____ backs off idea quickly, does not give any impression that he/she wants to win.
- _____ goes along with what others decide--offers few ideas on own.
- _____ takes group off task, wastes time, contributes to group stalls.
- _____ does not push for a quality decision, agrees in order to get done quickly.

Team Manufacturing
Scoring Guidelines

PROBLEM SOLVING

HIGH

- _____ asks many questions to promote sharing of information.
- _____ reads plans carefully, gives instructions to others showing that he/she really understands the product features.
- _____ integrates information from many folders to guide in building the product.
- _____ constantly offers ideas for making the product better or for assembling the product faster.
- _____ backs up ideas for improvements or pricing with sound logic.

TYPICAL

- _____ Occasionally asks questions to get information from others.
- _____ reads information from one folder, assembles his/her part correctly.
- _____ occasionally advises others on how to put product together according to plans.
- _____ offers an occasional idea for improving the product.
- _____ supports ideas for improvements or pricing with at least one reason.

LOW

- _____ asks no questions--makes mistakes due to lack of information.
- _____ seems unable to figure out how to build products.
- _____ makes parts incorrectly.
- _____ offers no suggestions for improving the product or for pricing.
- _____ throws out product improvement or pricing ideas without support.

Team Manufacturing
Scoring Guidelines

INTERPERSONAL SKILLS

HIGH

- _____ frequently compliments others for their ideas or work.
- _____ seems warm and friendly toward others, smiles.
- _____ uses humor, engages in small talk.
- _____ promotes communication and coordination with others by talking or asking questions.
- _____ openly seeks others' suggestions and ideas, willing to accept others' ideas through compromise.

TYPICAL

- _____ occasionally praises, supports others; does not offend.
- _____ pays attention, does not interrupt when others speak.
- _____ maintains an open posture toward others.
- _____ smiles, initiates some small talk with others.
- _____ occasionally communicates to coordinate tasks with others.

LOW

- _____ uses humor in a derogatory fashion.
- _____ criticizes others ideas, while offering no constructive suggestions.
- _____ works on a part of the project without telling others what he/she is doing--group effort is uncoordinated as a result.
- _____ seems unwilling to compromise, selfish or bossy.
- _____ completely silent or withdrawn.
- _____ lacks enthusiasm when communicating with others.
- _____ works silently, alone, without coordinating with others.

Team Manufacturing
Scoring Guidelines

INITIATIVE

HIGH

- _____ quickly involves self in task, works faster and harder than others.
- _____ does not tolerate idle time--never rests and gives idle group members work to do.
- _____ sticks with task from beginning to end, even when others' motivation is low.
- states improvement or pricing ideas directly and forcefully until the group accepts.
- _____ states opinion assertively, directly, using "I think," etc.
- _____ suggests ideas for proceeding, gives directions, and sets goals that are accepted by others.
- _____ refers to the time or productivity goals for the group--encourages group to meet goals.

TYPICAL

- _____ works at an even-keeled pace for duration of exercise.
- _____ occasionally directs and gives assignments to others.
- _____ suggests a few ideas for proceeding.
- _____ works responsibly on own--may not extend initiative by setting goals, giving directions to others.

LOW

- _____ does very little on own initiative--needs prompting from others.
- _____ engages in a lot of off task behavior--wasting time etc.
- _____ is easily distracted.
- _____ waits to be told what to do.
- _____ does not initiate any new ideas--follows flow of group.
- _____ no mention of/pays no attention to quality of product.

Customer Service
Scoring Guidelines

PROBLEM SOLVING

HIGH

- _____ seeks additional info. through asking questions.
- _____ questions reference to Mrs. Brown as troublemaker--seeks support.
- _____ highlights conflicting information--ex. mentions Child's memo.
- _____ grasps the complexity of the exercise--ex. recognizes that different perspectives may offer a unique, but truthful views.
- _____ offers creative solutions--ex. suggests reasonable alternatives not on action step list.
- _____ backs up selection of action steps with multiple reasons.
- _____ speculates about potential effects of taking one action vs. another.
- _____ draws on own knowledge or experience as a customer when recommending solutions.
- _____ supports recommendations by stating good business practice/principles--ex. treating customers well is important for business.

TYPICAL

- _____ occasionally asks questions for clarification of other's perspective.
- _____ has grasp of background information--is able to discuss the problem with reasonable fluency.
- _____ has made action steps choices from list prior to exercise.
- _____ supports each action step choice with at least one reason.

LOW

- _____ asks few questions--attempts to solve problem solely on own information.
- _____ asks superficial questions that do not get relevant information.
- _____ does not share unique information he/she possesses.
- _____ does not grasp background information--has trouble articulating the issues involved in the problem.
- _____ agrees with other's view of Mrs. Brown as main cause of problem.
- _____ seems confused by different perspectives on problem--tends to label certain points of view as right or wrong.
- _____ seems ill-prepared to discuss action steps.
- _____ throws out opinions about action steps without substantial backing.

Customer Service
Scoring Guidelines

INTERPERSONAL SKILLS

HIGH

- _____ shows regard for feelings of the people involved in the problem--
ex. sympathetic to Mrs. Brown.
- _____ reinforces other's ideas about problem or action steps.
- _____ seeks and listens to other's point of view.
- _____ does not prejudge whether other's view of Mrs. Brown is right or
wrong.
- _____ enthusiastic, open and warm during small talk.
- _____ uses sense of humor.
- _____ willing to compromise on view of problem or choice of actions as
new information is revealed.
- _____ smiles, maintains open body posture.

TYPICAL

- _____ occasionally says "good idea," etc.
- _____ does not interrupt other
- _____ participates in small talk with ease--shares personal experiences.
- _____ maintains eye contact

LOW

- _____ is quick to criticize or discount other's idea.
- _____ seems uncomfortable with meeting/greeting conversation--reluctant
to speak or share about self.
- _____ interrupts, pays little attention to other's opinions.
- _____ seems unwilling to compromise and unaccepting of other's
perspective.
- _____ is unusually critical of people in scenario--ex. makes derogatory
remarks about Mrs. Brown or supervisor.
- _____ seems withdrawn, non-participative, aloof.

Customer Service
Scoring Guidelines

INITIATIVE

HIGH

- _____ maintains high involvement & activity from beginning to end.
- _____ states opinions about problem or action steps clearly, directly--
ex. "I think you are being too hard on Mrs. Brown".
- _____ sticks to opinion even when challenged by other.
- _____ suggests a method for proceeding in discussion.
- _____ urges starting work on the assignment--interrupts small talk.
- _____ emphasizes the need to make decisions about action steps.
- _____ refers to time remaining in exercise--stresses the need to finish assignment.

TYPICAL

- _____ remains involved until finish of exercise.
- _____ states opinions/preferences with reasonable clarity
- _____ does not give up opinions too quickly--ex. does not back down when first challenged.
- _____ has occasional suggestions for how to proceed in discussion.
- _____ does not impede progress--may not vocalize achievement orientation.

LOW

- _____ seems to lack energy for exercise, activity level is low.
- _____ agrees easily with other--does not challenge what is said.
- _____ backs down from own opinion at first hint of challenge.
- _____ does not state opinions with confidence--hedges or quickly withdraws.
- _____ does not initiate conversation--speaks only when spoken to.
- _____ waits for other to suggest directions--does not act on own
- _____ impedes progress in exercise.
- _____ demonstrates little concern for a quality outcome.

Team Selection
Scoring Guidelines

PROBLEM SOLVING

HIGH

- _____ asks probing questions to discover other's choices and preferences.
- _____ mentions how different candidates can and can't work together
- _____ has problem resolved accurately at the outset of exercise--has teams already composed of individuals that will work together.
- _____ frequently offers suggestions to resolve impasses in the discussion.
- _____ articulates reasons for choices that are detailed and accurate
- _____ offers very convincing arguments why other's should accept his/her solution.
- _____ mentions the long term implications of selecting particular candidates--ex. higher/lower productivity, more/less conflict.

TYPICAL

- _____ asks a few questions to understand other's point of view.
- _____ speaks fluently about strengths and weaknesses of single candidates--may not consistently not pull information together across jobs.
- _____ provides a few reasons to support selections
- _____ provides logical reasoning in support of compromise.

LOW

- _____ asks no questions, does not seek information.
- _____ fails to recognize details regarding which candidates will and will not work together.
- _____ does not show understanding of the complexity of exercise--ex. focuses on getting highest number without any consideration for how team will work as a group.
- _____ states background information incorrectly.
- _____ throws out opinions and choices without logical support.

Team Selection
Scoring Guidelines

INTERPERSONAL SKILLS

HIGH

- ___ mentions the importance of both teams being good.
- ___ invites other's opinion and listens.
- ___ is relaxed and comfortable with meeting/greeting conversation
- ___ compromises in a good natured way--not a sore loser.
- ___ shows warmth by smiling, laughing, using humor.

TYPICAL

- ___ occasionally praises other--ex. "good thinking."
- ___ listens to other without prejudging.
- ___ opens up, shows warmth, and self discloses with invitation.
- ___ not offensive in any way but may not go out of way to make small talk or get to know other.
- ___ willing to compromise in trade for other compromises--ex. "If you take candidate X, I will take candidate Y."
- ___ maintains eye contact

LOW

- ___ is critical of other's choices.
- ___ does not warm up to other--scowling or aloof.
- ___ responds in short, curt, responses.
- ___ does not ask questions or seem interested in other's choices.
- ___ pushes only to get the highest numbers-never willing to compromise.
- ___ no eye contact or facial warm, physically stiff.

Team Selection
Scoring Guidelines

INITIATIVE

HIGH

- _____ sticks to task, continues to work through difficult, silent or awkward periods in the exercise.
- _____ states opinion directly using "I" think statements
- _____ suggests a plan for proceeding (i.e. discuss one at a time)
- _____ shows concern for getting finished
- _____ urges the need to get down to business-interrupts small talk with task consideration

TYPICAL

- _____ remains active from beginning to end of exercise.
- _____ states choices clearly.
- _____ puts forth some debate before conceding to other.
- _____ has a few suggestions for how to complete the assignment--may not consistently exert leadership.

LOW

- _____ demonstrates ambivalence for task, does not stress the need to or work toward completing the assignment.
- _____ -gives up/seems defeated when conversation stalls
- _____ gives ideas in a meek or meager way even when prompted.
- _____ voice does not reflect power, authority, or confidence in any way.
- _____ looks to other to solve the problem and make all suggestions.

LIST OF REFERENCES

LIST OF REFERENCES

- Archambeau, D. J. (1979). Relationship among skill ratings assigned in an assessment center. Journal of Assessment Center Technology, 2, 7-20.
- Balzer, W. K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. Organizational Behavior and Human Decision Processes, 37, 329-347.
- Benson, J. (1981). A redefinition of content validity. Educational and Psychological Measurement, 41, 793-802.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.
- Bernadin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 63, 301-308.
- Bernadin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.
- Bernadin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 62, 64-69.

- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.
- Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 20, 238-252.
- Borman, W. C. (1982). Validity of behavioral assessment for predicting military recruiter performance. Journal of Applied Psychology, 67, 3-9.
- Borman, W. C., & Dunnette, M. D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 60, 561-565.
- Boruch, R. F., Larkin, J. D., Wolins, L., & MacKinney A. C. (1970). Alternative methods of analysis: Multitrait-multimethod matrices. Educational and Psychological Measurement, 30, 833-853.
- Boruch, R. F., & Wolins, L. (1970). A procedure for estimation of trait, method, and error variance attributable to a measure. Educational and Psychological Measurement, 30, 547-574.
- Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. Psychological Monographs, 80, (17, Whole No. 625).
- Burnaska, R. F., Hollmann, T. D. (1974). An empirical comparison of the relative effects of rater response biases on three rating scale formats. Journal of Applied Psychology, 59, 307-312.
- Bycio, P., Alvares, K. M., Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis.

Journal of Applied Psychology, 72, 463-474.

Byham, W. C. (1977). Application of the assessment center method. In J. L. Moses & W. C. Byham (Eds.) Applying the assessment center method. New York: Pergamon Press.

Byham, W. C. (1980, February). Starting an assessment center the correct way. Personnel Administrator, 27-32.

Campbell, J. P., Dunnette, M. D., Arvey, R. D. & Hellervik, L. W. (1973). The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 57, 15-22.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Cardy, R. L. & Kehoe, J. F. (1984). Rater selective attention ability and appraisal effectiveness: The effect of cognitive style on the accuracy of differentiation among ratees. Journal of Applied Psychology, 69, 589-594.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Deutsch, M. (1949). An experimental study of the effects of cooperation and competition upon group process. Human Relations, 10, 307-318.

Dickenson, T. L., & Tice, T. E. (1973). A multitrait-multimethod analysis of scales developed by retranslation. Organizational Behavior and Human Performance, 9, 421-438.

Dreher, G. F. & Sackett, P. R. (1981). Some problems with applying content validity evidence to assessment center procedures. Academy of Management Review, 6, 551-560.

- Fiske, S. T., & Taylor, S. E. (1984). Social Cognition. Reading, Massachusetts: Addison-Wesley.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C. III, & Bentson, C. (1987). Meta-analyses of assessment center validity. Journal of Applied Psychology, [Monograph] 72, 493-511.
- Gaugler, B. B., & Thornton, G. C. III (1989). Number of assessment center dimensions as a determinant of assessor accuracy. Journal of Applied Psychology, 74, 611-618.
- Gorham, W. A. (1978). Federal executive agency guidelines and their impact on the assessment center method. Journal of Assessment Center Technology, 5, 1-7.
- Guilford, J. P. (1954). Psychometric methods. New York: Wiley.
- Guion, R. M. (1987). Changing views for personnel selection research. Personnel Psychology, 40, 199-213.
- Hollenbeck, G. P. (1990, April). The past present and future of assessment centers. In M. H. Newlin (Chair) Simulated performance assessment: Fact or fantasy. Symposium conducted at the Fifth Annual Conference for the Society for Industrial and Organizational Psychology. Miami Beach Florida.
- Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. Journal of Applied Psychology, 64, 502-508.
- Joreskog, K. G., & Sorbom, D. (1986). LISREL: Analysis of linear structural relationships by the method of maximum likelihood, (4th ed.). Chicago: National Educational Resources.

- Kavanagh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 75, 34-39.
- Kenrick, D. T., & Stringfield, D. O. (1980). Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all people. Psychological Review, 87, 88-104.
- Klimoski, R., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. Personnel Psychology, 40, 243-260.
- Landy, F. J., & Farr, J. L. (1980). Performance Rating. Psychological Bulletin, 87, 72-107.
- Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.
- Miller, L. K. & Hamblin, R. L. (1963). Interdependence, differential rewarding, and productivity. American Sociological Review, 25, 768-778.
- Mitchell, T. R. & Silver, W. S. (1990). Individual and group goals when workers are interdependent: Effects on task strategies and performance. Journal of Applied Psychology, 75, 185-193.
- Mulaik, S. A., James, L. R. Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for

structural equation models. Psychological Bulletin, 105, 430-445.

Neidig, R. D., Martin, J. C., & Yates, R. E. (1979). The contribution of exercise skill ratings to final assessment center evaluations. Journal of Assessment Center Technology, 2, 21-23.

Neidig, R.D. & Neidig, P. J. (1984). Multiple assessment center exercises and job relatedness. Journal of Applied Psychology, [Short Note] 69, 182-186.

Norton, S. D. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. Academy of Management Review, 2, 442-453.

Robertson, I., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. Journal of Occupational Psychology, 60, 187-195.

Russell, C. J. (1985). Individual decision processes in an assessment center. Journal of Applied Psychology, 70, 737-746.

Sackett, P. R. (1987). Assessment centers and content validity: Some neglected issues. Personnel Psychology, 40, 13-25.

Sackett, P. R. & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troublesome empirical findings. Journal of Applied Psychology, 67, 401-410.

Sackett, P. R. & Dreher, G. F. (1984). Situation specificity of behavior and assessment center validation strategies: A rejoinder to Neidig and Neidig. Journal of Applied Psychology, 69, 187-190.

- Sackett, P. R. & Hakel, M. D. (1979). Temporal stability and individual differences in using assessment information to form overall ratings. Organizational Behavior and Human Performance, 23, 120-137.
- Sackett, P. R., & Wilson, M. A. (1982). Factors affecting the consensus judgment process in managerial assessment centers. Journal of Applied Psychology, 67, 10-17.
- Schmidt, F. L., & Hunter, J. H. (1980). The future of criterion-related validity. Personnel Psychology, 33, 41-60.
- Schmitt, N. (1977). Interrater agreement in dimensionality and combination of assessment center judgments. Journal of Applied Psychology, 62, 171-176.
- Schmitt, N., Gooding, R. Z., Noe, R.A., & Kirsch, M. (1984) Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.
- Schmitt, N, Schneider, J. R., & Cohen, S. A. (1990). Factors affecting validity of a regionally administered assessment center. Personnel Psychology, 43, 1-12.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.
- Silverman, W. H., Dalessio, A., Woods, S. B., & Johnson, R. L. Jr. (1986). Influence of assessment center methods on assessors' ratings. Personnel Psychology, 39, 565-578.

- Slovic, P., & Lichtenstein, S. (1971). Comparisons of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.
- Smith, P. C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), Handbook of Industrial and Organizational Psychology. Chicago: Rand McNally.
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.
- Society for Industrial and Organizational Psychology Incorporated. (1987) Principles for the Validation and Use of Personnel Selection Procedures. (3rd ed.) College Park, MD.
- Stanley, J. C. (1961). Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. Psychometrika, 26, 205-219.
- Task Force on Assessment Center Standards (1980, February). Standards and ethical considerations for assessment center operations. Personnel Administrator, 35-38.
- Tenopyr, M. L. (1977). Content-construct confusion. Personnel Psychology, 30, 47-54.
- Thornton, G. C., & Byham, W. C. (1982). Assessment centers and managerial performance, New York: Academic Press.
- Tucker, L. R. & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

- Turnage, J. J. & Muchinsky, P. M. (1982). Transsituational variability in human performance within assessment centers. Organizational Behavior and Human Performance, 30, 174-200.
- Uniform Guidelines on Employee Selection Procedures. (1978, March). Federal Register.
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, & criteria. Journal of Applied Psychology, 52, 372-376.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. Applied Psychological Measurement, 9, 1-26.
- Zedeck, S., & Baker, H. T. (1972). Nursing performance as measured by behavioral expectation scales: A multitrait-multirater analysis. Organizational Behavior and Human Performance, 7, 457-466.
- Zedeck, S. (1986). A process analysis of the assessment center method. In B. M. Staw & L. L. Cummings, Research in Organizational Behavior. Greenwich, Connecticut: JAI Press.
- Zedeck, S., & Cascio, W. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67, 752-758.
- Zedeck, S., & Cascio, W. F. (1986). Psychological issues in personnel decisions. Annual Review of Psychology, 35, 461-518.

MICHIGAN STATE UNIV. LIBRARIES



31293008919304