

AD-A090 795

AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH
THE RELATIONSHIP BETWEEN PROGRAM EVALUATION RESEARCH AND SELECT--ETC(U)
DEC 79 W J STRICKLAND
AFIT-CI-79-246D

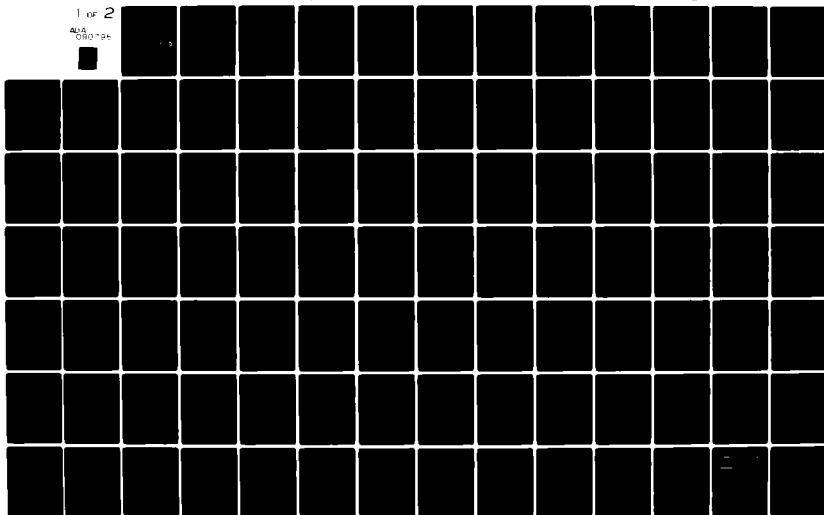
F/6 5/10

UNCLASSIFIED

NL

1 of 2

ALL
ORIGINALS



LEVEL II

①

THE RELATIONSHIP BETWEEN PROGRAM
EVALUATION RESEARCH AND SELECTION SYSTEM
VALIDATION--APPLICATION TO THE
ASSESSMENT CENTER METHOD.

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

AD A090795

DDC FILE COPY.

DTIC
ELECTE
OCT 24 1980
S D E

By
⑩ William James Strickland / B. S., M. A.

11 Dec 77

The Ohio State University

1979

AFI 7-77-2461

Reading Committee:

Approved By

Richard J. Klimoski, Ph. D.
Robert S. Billings, Ph. D.
Edwin T. Cornelius III, Ph. D.

Adviser
Department of Psychology

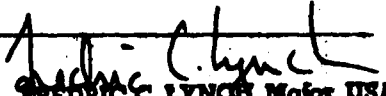
DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

80 10 14 231

017-99

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 79-246D	2. GOVT ACCESSION NO. AD-A090 795	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Relationship Between Program Evaluation Research and Selection System Validation--Application to the Assessment Center Method		5. TYPE OF REPORT & PERIOD COVERED THESIS/DISSERTATION
7. AUTHOR(s) William J. Strickland		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT STUDENT AT: The Ohio State University		8. CONTRACT OR GRANT NUMBER(s)
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/NR WPAFB OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE Dec 79
		13. NUMBER OF PAGES 97
		15. SECURITY CLASS. (of this report) UNCLASS
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES APPROVED FOR PUBLIC RELEASE: IAW AFR 190-17 25 SEP 1980  FREDRIC C. LYNCH, Major, USAF Director of Public Affairs Air Force Institute of Technology (ATC) Wright-Patterson AFB, OH 45433		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) ATTACHED		

80 10 14 231

THE RELATIONSHIP BETWEEN PROGRAM EVALUATION RESEARCH
AND SELECTION SYSTEM VALIDATION -- APPLICATION
TO THE ASSESSMENT CENTER METHOD

William J. Strickland 109p 1979 Thesis (Ph.D.)
The Ohio State University

This field study describes a comparative evaluation approach to determining the effectiveness of a popular personnel program--the assessment center method. In contrast to traditional validity research, the emphasis is on the key roles of multiple (often conflicting) program values or goals, on alternatives to using an assessment center, and on cost-effectiveness. While the data indicate that the specific assessment center under study did as well predicting advancement criteria as comparable centers reported in the literature, it was less effective than other, less costly, alternatives. Further, when it came to less traditional criteria (or goals) such as the prediction of performance effectiveness of managers or enhancing employee development, it was generally inferior to the other "programs" examined. It is proposed that this approach to assessing an assessment center (or any selection system) is in many ways superior to validation programs commonly pursued by industrial/organizational psychologists.

Accession For	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

Copyright by
William James Strickland
1979

ACKNOWLEDGMENTS

It would have been impossible to complete this dissertation without the close cooperation of many people; the brief acknowledgment here cannot adequately express my thanks.

First, Dr. Jack Roose and Mr. Ray Mason provided access to the participating organization, arranged observation periods at the assessment center, and set up data collection procedures within the organization. Drs. Milton Hakel, Edwin Cornelius, and Robert Billings provided necessary guidance on my proposal and draft dissertation. Of course, Dr. Richard Klimoski was intimately involved in the entire project; I will always be grateful that his letters and long distance phone calls made me feel so guilty that I was forced to complete this degree. Major Chuck Durham (USAF, Ph. D.) provided me with his assessment center survey. Finally, my wife, Martha, spent many frustrating evenings trying to read my scribbles, then having to re-type everything when I would change my mind.

VITA

April 12, 1948	Born - Boston, Massachusetts
1970	B. S. United States Air Force Academy, Colorado
1970-1971.	Merston Fellow, The Ohio State University, Columbus, Ohio
1971	M. A., The Ohio State University, Columbus, Ohio
1971-1975.	Behavioral Scientist, United States Air Force Occupational Measurement Center, Lackland AFB, Texas
1975-1977.	Behavioral Scientist, Air Force Institute of Technology
1977	Behavioral Scientist, Headquarters Air Training Command, Randolph AFB, Texas

PUBLICATIONS

Klimoski, R. J., & Strickland, W. J. Assessment centers --
Valid or merely prescient? Personnel Psychology, 1977,
30, 353-361.

THESIS

Predicting selection for graduate study in psychology.

FIELDS OF STUDY

Major Field: Industrial - Organizational Psychology

Studies in Industrial Psychology: individual effectiveness, merit rating, job analysis, training, motivation, assessment centers, job satisfaction. Professors Richard Klimoski, Milton Hakel, Robert Billings, and Edwin Cornelius.

Studies in Organizational Psychology: formal organization theory, organizational effectiveness, leadership. Professor Richard Klimoski.

Studies in Quantitative Psychology: mathematical psychology, inferential statistics, correlational analysis, factor analysis, programming for psychology. Professor Robert Wherry.

Minor Field: Manpower and Industrial Relations

Studies in Industrial Relations: manpower and industrial relations, labor law, administration of interpersonal behavior, collective bargaining. Professor Joseph Yaney.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES.	vii
LIST OF FIGURES	ix
INTRODUCTION.	1
 Chapter	
I. RELATIONSHIP OF EVALUATION AND VALIDATION.	3
Definition of Evaluation Research	3
Definition of Validation.	6
Evaluation Research Approaches.	7
Linkages.	11
II. VALIDITY OF THE ASSESSMENT CENTER	18
Definition of an Assessment Center.	18
Validity of Assessment Centers.	19
III. METHOD.	23
The Assessment Center	23
Organizational Goals.	26
Operational Criteria.	27
Alternative Programs.	30
IV. RESULTS	33
Organizational Goals.	33
Criteria.	39
Assessment Center Performance	41
Alternative Program Performance	53
Comparative Summary	58

V.	DISCUSSION	61
	Validity of the Assessment Center.	61
	Evaluation of the Assessment Center.	63
	Cost-Effectiveness of the Assessment Center.	69
VI.	CONCLUSION	71
	Evaluation of an Assessment Center	71
	Validity of Assessment Centers	72
	Evaluation and Validation.	73
APPENDICES		
A.	Assessment Dimensions.	74
B.	Description of Tests and Exercises	76
C.	Interview Guide.	80
D.	Rating Form.	82
E.	Survey Questionnaire	84
F.	Scale Questions.	89
G.	Supplementary Tables	92
	LIST OF REFERENCES	97

LIST OF TABLES

1.	Validity Studies of Overall Assessment Ratings from Published Sources	20
2.	Characteristics of Assesseees	26
3.	Sources and Priority Listings of Assessment Center Goals	37
4.	Assessment Goals and Criteria.	42
5.	Correlations Between Assessment Center Overall Rating and Criteria for Goal of Predicting Success	43
6.	Salary Comparison of Assesseees with Non-Assesseees.	43
7.	Correlations Between Assessment Center Overall Rating and Criteria for Goal of Identifying Potential.	45
8.	Survey Results for Developmental Goal.	47
9.	Survey Results for Career Paths Goal	48
10.	Survey Results for Feedback Goal	50
11.	Evaluation of Intelligence Testing Goal.	52
12.	Correlations Between Alternative Programs and Criteria for the Goal of Predicting Success.	54
13.	Correlations Between Alternative Programs and Criteria for the Goal of Identifying High Potential	56
14.	Correlations Between Alternative Programs and Intelligence Tests	59
15.	Comparative Validities	60

16.	Correlations Among Criteria for the Goals of Predicting Success and Identifying Potential.	93
17.	Correlations Among Predictors for the Goals of Predicting Success and Identifying Potential.	94
18.	Multiple Regression Summary for the Criterion of Present Grade	95
19.	Multiple Regression Summary for the Criterion of Grade Changes	96

LIST OF FIGURES

1. A model of evaluation research 13
2. Model for evaluation/validation. 14
3. Company rating form. 83

INTRODUCTION

Industrial-organizational psychologists have long been concerned with validation of selection systems; Guion (1976) considers this concern to be "the hallmark of 'traditional' industrial psychology" (p. 779). Methods for establishing the validity of tests have been available since the early 1900's, with statistical techniques still being refined today (although the basic principles bearing on validity have changed very little.) Paralleling this concern with validation has been an evolving multidiscipline of program evaluation, tracing roots back to the mid-nineteenth century, and becoming somewhat formalized during the New Deal in the 1930's (Perloff, Perloff, & Sussna, 1976).

For the most part, industrial-organizational psychologists have not been very involved in what is now known as program evaluation research. Project Head Start, Sesame Street, income maintenance programs, or the introduction of a new teaching method do not seem closely related to interviews, biodata, the CPI, or the TAT. In a recent article outlining opportunities for psychologists in evaluation research, for example, Wortman (1975) sees relevance for experimental, quantitative, social, and clinical psychologists, but never mentions industrial-organizational

psychologists. There is relevance for I-O psychology, however; in some areas, there exist only slight differences in terminology between the fields--in other areas, the differences are more substantial, but the parallels are there.

The present research will more clearly define the relationship between program evaluation research and selection system validation by reviewing some concepts in evaluation and validation, reviewing the evidence of validity surrounding a particular selection technique (the assessment center), and applying an evaluative approach to that selection technique as it exists in a specific field setting. Thus, the results of this research should be applicable in several ways: first, by strengthening the relationship between the two fields; second, by contributing to the validation evidence about assessment centers; finally, by providing information to the appropriate decision-makers within the field-site organization.

I. RELATIONSHIP OF EVALUATION AND VALIDATION

An examination of the relationship between evaluation and validation should naturally begin with definitions of the areas involved and some examples from the literature of each area. Additionally, an examination of some approaches used within evaluation research that are especially relevant to validation would be appropriate. Finally, these evaluation approaches can be linked to selection system validation.

Definition of Evaluation Research

Perloff, Perloff, and Sussna (1976) offered a definition of evaluation research as an empirically oriented research technology to determine "the extent to which a program has achieved one or more of its objectives, the reasons it may not have achieved them, and the relationships among program effects and a variety of input variables and program characteristics" (p. 570). Programs include activities designed to improve the social or economic welfare of an individual or to solve a social or economic problem in the fields of education or mental health. Finally, Perloff et al. include a whole range of techniques "spanning

the gamut from impressionistic procedures, clinical and observational procedures, field or survey methods, and an armamentarium of sophisticated research and quantitative procedures" (p. 570).

Caro (1971) offers several similar definitions taken from various sources. Most of these definitions focus on information about a program's outcomes or judgments about a program's value. For example, Caro notes that Brooks (1965) defines three objectives of evaluation: (1) determining the extent to which a program meets its goals; (2) establishing the relative impact of program variables; and (3) defining the role of the program as opposed to that of external variables. Basically, this definition involves information-seeking activity. Scriven (1967), on the other hand, is more concerned with the judgmental aspects of evaluation. Rather than just providing information about a program's contribution toward goals, the evaluation should include judgments about the relative worth of goals. Glass (1971) goes even further: Since official program goals may be questioned, an evaluation should include an analysis of those goals.

Researchers often distinguish two types of evaluation: formative and summative (terms originated by Scriven, 1967). Essentially, formative evaluation is appropriate for relatively new programs, while summative evaluation is more

concerned with a stable or well-established program. As Caro (1971) concludes, "Formative evaluation is designed to improve a program while it is still fluid; summative evaluation is designed to appraise a product after it is well established" (p. 4).

As noted previously, candidate programs for evaluation research have typically been in the education, mental health, or social/economic welfare areas. Perloff et al. (1976) cite numerous examples of published evaluation research, grouped as evaluations of service programs (mental health and psychiatric services, alcohol and drug rehabilitation programs, halfway houses for ex-offenders, etc.), educational and training programs (evaluating teachers and courses, the benefits of higher education, etc.), and miscellaneous programs (manpower or research utilization).

Deming (1975) provides a simple summary for most of these definitions: "Evaluation is a pronouncement concerning the effectiveness of some treatment or plan that has been tried or put into effect" (p. 53). The emphasis, then, is on causal relationships.

Definition of Validation

Typically, the purpose of a selection system is to predict future behavior. Validity usually refers to the relationship between some aspect of the selection system (a test, for example) and some measure of behavior (a criterion). As Guion (1965) notes, "validity is concerned with how relevant test scores are to something else" (p. 123). Of course, the concept of "a test" includes any aspect of the selection system, and the notion of "a criterion" does not preclude multiple or composite criteria.

Guion (1976) describes ten "tenets of orthodoxy" shared by industrial psychologists up to and during the 1950's, outlining the traditional model of test validation:

1. The purpose is to predict future job performance ...
 2. Predictors and criteria should be selected on the basis of job analysis ...
 3. Measuring instruments must be standardized ...
 4. Tests should be empirically evaluated ...
 5. Validation is situation-specific ...
 6. More than one test should be used ...
 7. But only one criterion should be used ...
 8. Tests are preferred over "non-test" predictors ...
 9. Individual differences should be recognized in evaluating tests ...
 10. Tests are supplements to existing employment processes
- [pp. 783 & 784.]

Many examples of published validity studies are available: Korman (1968) reviewed validity research on the selection of managers--Schuh (1967) reviewed validity research with tenure as the criterion--Guion and Gotier

(1965) reviewed personality measure validity--and others. Additionally, numerous unpublished validity studies have been carried out, ranging from atheoretical application blank validity to conceptually based, complex performance tests.

Validation has been characterized by the search for predictive relationships. As Guion (1976) summarizes, validation is "the empirical testing, where possible and appropriate, of rationally developed hypotheses about individual, situational, or subgroup characteristics which may influence job behavior or its effects" (p. 777). Thus, like evaluation, this search should be emphasizing causal processes, as noted by the Division of Industrial-Organizational Psychology (1975): "Predictor constructs should be chosen for which there is an empirical or logical foundation" (p. 4).

Evaluation Research Approaches

Several major approaches to program evaluation are available. Perloff et al. (1976) discuss five of these approaches, ranging from those primarily concerned with the appropriate methodology for evaluation (clinical and quasi-experimental design approaches) to those more concerned with conceptual issues in evaluation (values-linked, management-oriented, and benefit-cost analytic approaches).

While there are clearly lessons in the clinical, quasi-experimental, and management-oriented approaches to evaluation research, the issues raised in this literature are familiar to industrial-organizational psychologists. More relevant to the present research are the issues involved in the values-linked and benefit-cost analytic approaches to evaluation research.

Values-linked approaches to evaluation place great weight on the values, preferences, and goals of the "consumer" of evaluation research--usually the decision-maker. A major purpose of advocates of this approach is to point out that if an evaluation is to have any effect on a program, the decision-maker's values must be considered. That is, the goals against which the researcher chooses to evaluate a program can range from totally irrelevant to crucial in anyone else's priorities. Advocates of this approach propose a decision-theoretic method for quantifying subjective values or preferences (Perloff et al., 1976).

Some researchers consider a values-orientation to be more than just an approach within evaluation, however; they consider this orientation as defining evaluation research. For example, Suchman (1971) argues that evaluation research differs from basic or nonevaluative research only because value is attached to the dependent variable

in evaluation research. Similarly, Burgoyne and Cooper (1975) write, "the term 'evaluation' implies the valuing of consequences" (p. 55).

Moreover, in many cases, simply having a specific program may be more important or have more value to a decision-maker than the results of that program (Edwards, Guttentag, & Snapper, 1975). Even apart from decision-makers, Messick (1975) notes that the values of evaluators often differ, resulting in different conclusions from the same data (a phenomenon not unknown in industrial-organizational psychology!) He reiterates Hudson's law of selective attention to data: "the greater the ideological relevance of research, the greater the likelihood that the research worker doing it will pay selective attention to the evidence he collects" (p. 964).

Consideration of these values issues by industrial-organizational psychologists would certainly be worthwhile; similarly, the benefit-cost analytic approach to evaluation research is directly relevant to selection system validation. Benefit-cost analysis is designed to answer the question of how to choose among alternative approaches for achieving a set of goals (Rossi, 1972). Perloff et al. (1976) consider benefit-cost analysis as the broadest approach to evaluation research because it "seeks to embrace in one overarching conceptual scheme the benefits

(identified and measured, for example, through some clinical or quasi-experimental approach) of a program as moderated by the program's cost" (p. 574).

Levin (1975) further defines benefit-cost analysis, and identifies some subsets of this concept. Specifically, he considers benefit-cost analysis as the direct comparison of costs and benefits to society of a policy alternative, using some common metric (usually money) for all costs and benefits. A special case of benefit-cost analysis (easier to deal with and, normally, more appropriate) is cost-effectiveness. In this case, costs of each alternative are required in monetary terms, but any convenient metric is used for program outcomes.

Cost-effectiveness analysis (or benefit-cost analysis) must also consider values; that is, these approaches are not mutually exclusive. The distinction exists only in a given approach's research emphasis--defining and quantifying values or identifying and pricing-out costs and benefits. Those emphasizing values must (and do) recognize that lower-cost solutions to problems are often valued more than high-cost solutions; similarly, those emphasizing monetary costs and benefits recognize that values impact the perceived effectiveness of programs.

Linkages

It may appear that issues important in evaluating large-scale social programs--the importance of values and benefit-cost analysis--have little to do with an industrial-organizational psychologist assessing the validity of a predictor in a selection decision for a firm. These same issues are important, however, and this importance has often been recognized. Messick (1975) notes that whenever measurement is attempted, the researcher has made a choice that some things are more important to measure than others--an intrusion of values. Similarly, Guion (1976) agrees that "acceptance of correlation as validity implies acceptance of the correlated criterion as an important concept in vocational success" (p. 786). The very notion of importance implies values. Finally, Cronbach and Glaser (1965) provide an extensive discussion of values--not surprising since they advocate the same type of decision-theoretic methodology underlying the values-linked approach to evaluation.

Benefit-cost analysis also has its place in industrial-organizational psychology. A recognition of the importance of the monetary outcomes of employment decisions goes back more than fifty years: Freyd's 1923 advice was to "consult the cost accountant to find the department in which increased efficiency in selecting employees would bring about

the greatest economic saving to the firm" (p. 218). More recently, there was Brogden and Taylor's (1950) "dollar criterion." Finally, of course, utility theory (Cronbach & Glaser, 1965) is a direct parallel to cost-effectiveness.

Thus, a values orientation and an appreciation of the importance of benefit/cost ratios is as important in selection system validation as in social program evaluation. A model of evaluation research would be useful in isolating components of evaluation--relevant to validation--that are often overlooked in validation. Wortman (1975) provides a model of evaluation research (Figure 1). He views this model as answering the question "whether any coherent picture or description of evaluation research can emerge from this plethora of terms and concepts" (p. 565). The model was designed to outline the procedure for establishing a cause/effect relationship, by tying together the evaluation research processes and indicating for the researcher the relevant organizational components he must be prepared to deal with during each process. Many segments of this model would be recognized by the validity researcher; the relationship between the fields starts to come into focus.

For purposes of more clearly examining the relationship between evaluation and validation, however, the Wortman model should be revised somewhat. Figure 2, which

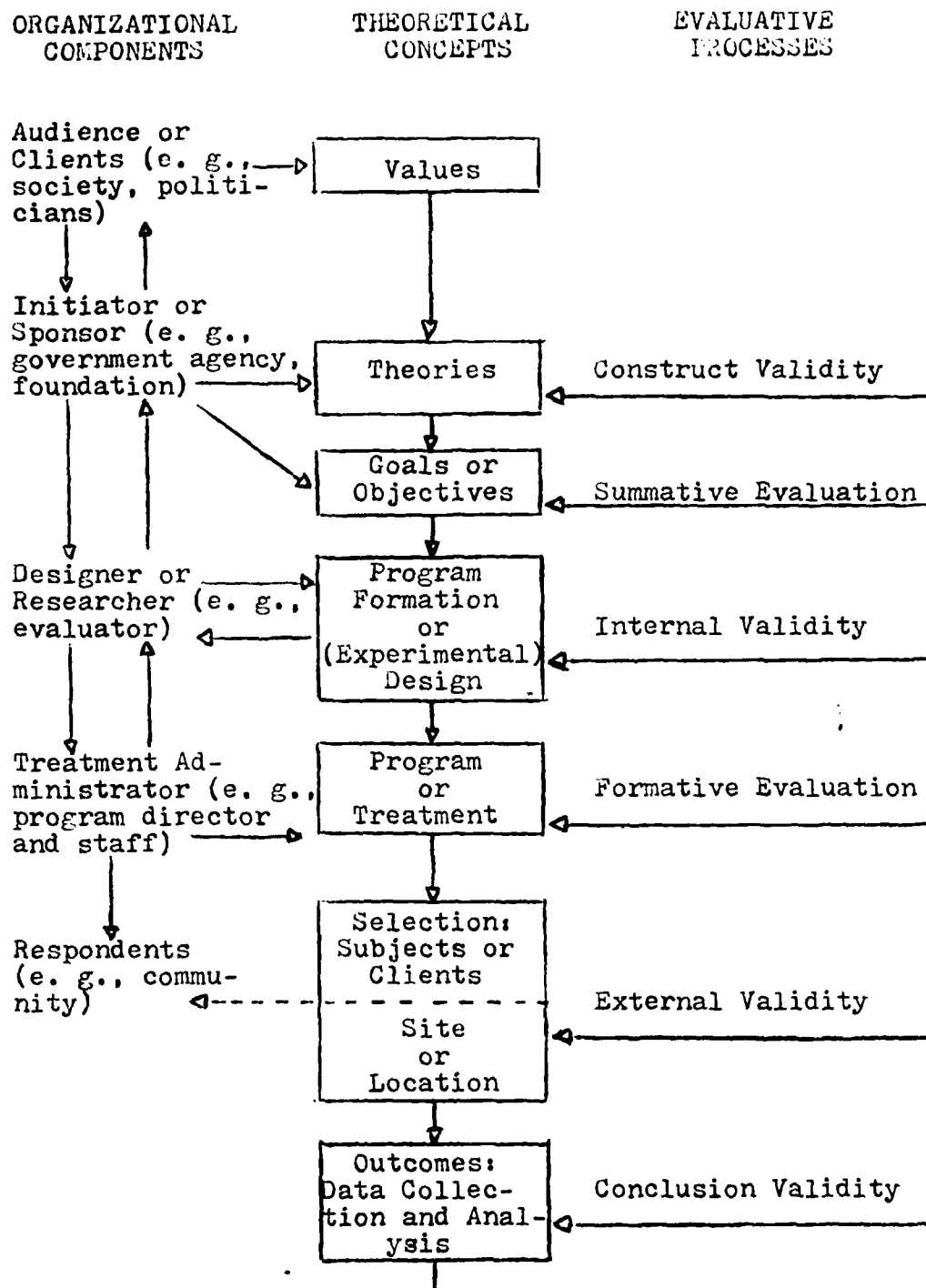


Figure 1. A model of evaluation research.

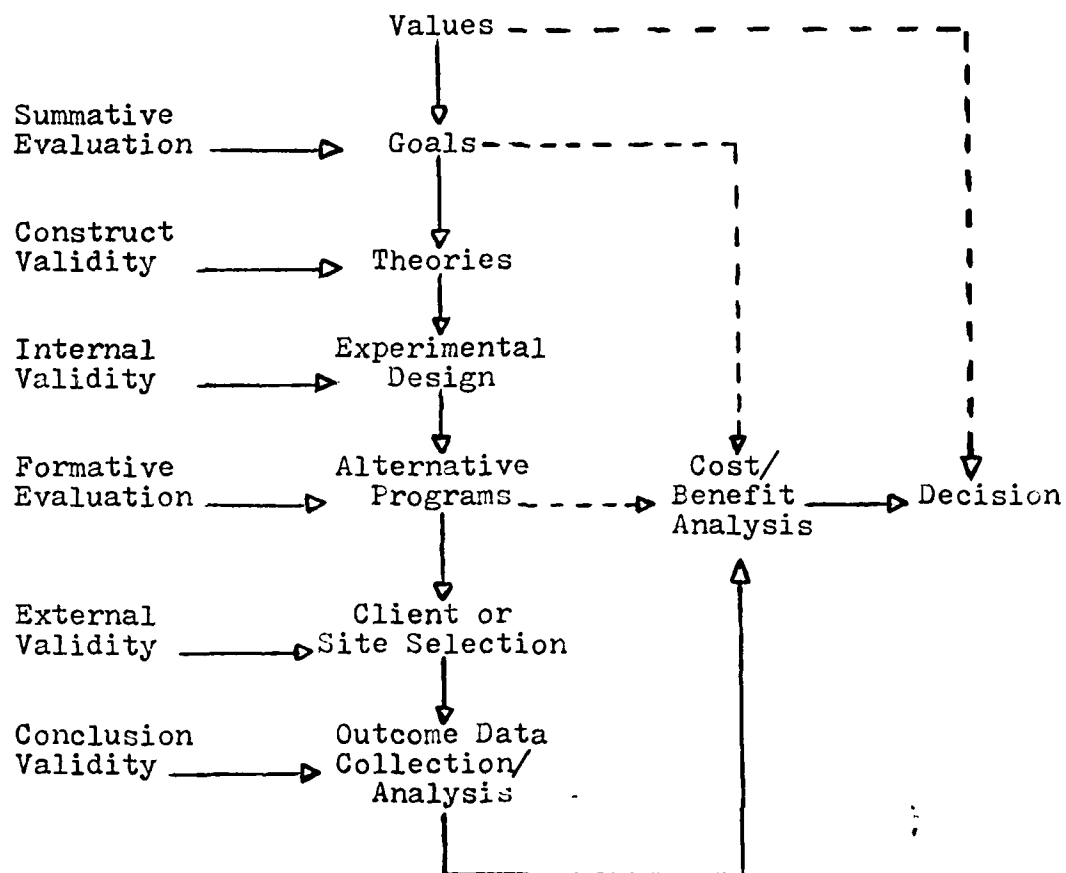


Figure 2. Model for evaluation/validation.

is a model for decision-making, provides a revision to the Wortman model that shows the closeness of evaluation and validation. Specifically, if the values, cost/benefit analysis, and decision components are ignored (and formative and summative evaluation are expressed in other terms), this model would be recognized and accepted as a model for validity research. The addition of cost/benefit analysis and decision components to the Wortman model make that model more complete; that is, rather than stopping when a cause/effect relationship has been demonstrated, the Wortman model ought to provide for consideration of relative costs and the fact that the purpose of the evaluation (or validation) is to aid in decision-making about the program. Similarly, the dashed arrows in Figure 2 are designed to highlight for the researcher (evaluation or validation) that, even though the research has its basis in values and goals, those values and goals will ultimately have direct impact on the cost/benefit analysis and any decisions about the program. That is, researchers must be aware, throughout the process, that program goals and the values behind those goals are critical if any research effort hopes to impact the decision-making process. These additions, then, (to validity research and the Wortman model of evaluation research) result in a model for decision-making, useful in both evaluation or validation,

whether the decision involves a \$20 million Federal education program being evaluated by a sociologist or a \$50 administration of an intelligence test being validated by an industrial-organizational psychologist. The model indicates that the researcher must do more than "just" establish a causal relationship to insure that his efforts will have some impact: He must consider values and costs.

Most of the components of this model are either well-known or have already been discussed. Validity research is usually characterized by an emphasis on conclusion validity--the use and interpretation of appropriate statistical tests. Additionally, the literature on the concept of validity usually points out the importance of construct validity. Researchers are typically aware of the issues involved in internal and external validity, and there is intuitive acceptance of the concept of formative evaluation (e. g., small-scale pilot studies, trying to insure standardized test administration, etc.) However, some components seem to be ignored in validity research: the concept of summative evaluation against values-generated goals; the requirement in summative evaluation to evaluate the goals themselves; the appropriate benefit-cost or cost-effectiveness analysis (including consideration of alternate programs and costs of unanticipated or unwanted outcomes of a program); and, of course, the recognition that there is a

consumer of the research whose values may lead to decisions that the researcher perceives as irrational.

Given that a model for evaluation research has relevance for validity research, the present research will "validate" a selection procedure from an evaluative point of view. A currently popular selection procedure, for which a great deal of validity evidence already exists, is the assessment center concept.

II. VALIDITY OF THE ASSESSMENT CENTER

There has been considerable interest during the past decade in a selection technique generically termed the managerial assessment center. This chapter briefly defines the essential elements of an assessment center and reviews the validity evidence for this technique.

Definition of an Assessment Center

Most reviews of assessment centers trace the origin of the technique in the United States back to the Office of Strategic Services during World War II, then to Bray's work at AT & T. Subsequent work at SOHIO, IBM, Sears, and GE set the stage for rapid growth of the technique (MacKinnon, 1975; Finkle, 1976; Cayer & Kirschner, 1977).

MacKinnon (1975) defines an assessment center as "a method for the psychological evaluation of individuals that involves testing and observing of individuals in a group setting, with a multiplicity of tests and procedures, by a number of staff members" (p. 1). Finkle (1976) concurs that essential elements of an assessment center include group settings, multiple techniques, and multiple assessors, but adds one other requirement: emphasis on situational exercises.

Typically, groups of assesseees are observed by several high-level managers or psychologists as they perform various group and individual tasks. Additional information is gathered in the form of background questionnaires, individual interviews, and tests. The basis for each exercise or test is ideally found in a performance dimension identified through job analysis. All of this information is then integrated by the assessors to arrive at consensus ratings for each assessee on the various dimensions, plus some overall measure (for example, a promotion decision, a prediction of advancement, or a rating of potential.) Often, assessors formulate a developmental plan for each assessee, and, usually, assesseees receive some kind of feedback.

Validity of Assessment Centers

Numerous reviews of the assessment center method and techniques are available (e. g., MacKinnon, 1975; Finkle, 1976), as well as reviews specifically oriented toward the validity evidence on assessment centers (e. g., Huck, 1973). Klimoski and Strickland (1977) cite 17 published validity studies on assessment centers over the last ten years and note that "regardless of the center format used, these results have been impressive, positive, and consistent" (p. 354). Table 1 summarizes these studies (taken from Klimoski & Strickland, 1977, p. 356).

TABLE 1
Validity Studies
of Overall Assessment Ratings from Published Sources

Source	Criteria	Assessors	Company
Bray & Grant, 1966	Management level, salary, and salary progress	Psychologists	AT & T
Campbell & Bray, 1967	Ratings, ranking and number of promotions	Mixed ^a	AT & T
Bray & Campbell, 1968	Special performance review	Managers	AT & T
Wollowick & McNamara, 1969	Increase in responsibility	Managers	IBM
Hinrichs, 1969	Salary standing	Managers	IBM
Carleton, 1970	Ratings, salary progress, and number of promotions	Mixed	SOHIO
Thomson, 1970	Ratings (timing of criterion measures varied)	Mixed	SOHIO
Jaffee, Bender, & Calvert, 1970	Interview with superior	Managers	Union Carbide
Kraut & Scott, 1972	Promotions and demotions	Managers	IBM
McConnell & Parker, 1972	Ratings, but obtained concurrent with assessment center	Managers	Various

TABLE 1 (continued)

Source	Criteria	Assessors	Company
Ginsburg & Silverman, 1972	Ratings, but obtained con- current with assessment center	Managers	Hospital
Thoreson & Jaffee, 1973	Ratings, but obtained con- current with assessment center	Managers	Rohem & Haas
Byham & Wettengel, 1974	Ratings, but obtained con- current with assessment center	Managers	State Govern- ment
Moses & Boehm, 1975	Management level achieved	Managers	AT & T
Mitchel, 1975	Salary growth	--	SOHIO
Worbois, 1975	Ratings, but obtained con- current with assessment center	Managers	--
Huck & Bray, 1976	Rating & ranking	Managers	AT & T

^a"Mixed" includes some combination of managers and psychologists.

Typical validity coefficients for the studies outlined in Table 1 are in the .3 to .5 range. Cohen, Moses, and Byham (1974) report a median correlation of .40 between assessment center ratings of promotion potential and number of promotions above the first level. Based on the evidence, Cayer and Kirschner (1977) conclude that "the assessment center method generally may be a more valid method of management selection and promotion than more traditional methods such as supervisory appraisals or paper-and-pencil testing" (p. 21).

Thus, within industrial-organizational psychology, the assessment center method enjoys a reputation as a valid selection device, in a wide range of organizations. Most of these validity studies, however, have considered only limited criteria; the rationale relating criteria to organizational (or individual) goals is not usually present; there usually is no consideration given to alternative techniques or relative costs; finally, one suspects that, in many firms (especially smaller firms with recently established assessment centers) the method itself is valued more than its results--validity evidence is essentially irrelevant to any decision by anyone. An evaluative approach to establishing the validity of an assessment center would therefore be worthwhile.

III. METHOD

The approach taken was a comparative evaluation of an ongoing assessment center: goal statements were obtained from a variety of organizational sources; appropriate operational criteria for each expressed goal were developed and data obtained on those criteria; alternative programs were hypothesized and data obtained on those programs and on the assessment center; finally, the performance of the assessment center was contrast with the performance of the alternative programs for each expressed goal.

The Assessment Center

The assessment center evaluated in this study is operated by a large, midwestern-based firm on a regular schedule. The center had been in operation for over five years when the study was initiated, and only very limited validity evidence was available.

Procedures. Midlevel company employees were nominated by their supervisors for attendance at the center. Center staff made final selection for attendance after reviewing personnel records of nominees; only those employees of judged high potential were to be assessed. Assesseees

traveled to the company training site (geographically separated from operational company activities) for a 2½ day assessment by high-level managers. Twelve employees were assessed by three managers chosen by the center staff to ensure that none of the assessors and assessees were acquainted.

Generally, assessors received an orientation manual in advance, received ½ day of on-site training while the assessees were taking paper and pencil tests, and received a briefing on each exercise immediately before observing it. The permanent assessment center staff concentrated on running the exercises rather than on observing the performance of assessees. Assessors noted only behaviors in each exercise; dimensions were not rated until after the last exercise. Following the last exercise, assessees returned to their jobs while assessors spent about two days reaching consensus on each assessee.

For the assessor meeting, the managers were joined by the center director (who had been present for all exercises) and by a consulting psychologist (who had been present for none of the exercises). Each assessee was discussed individually: the center director provided biographical information and peer rating results; the psychologist interpreted test scores; the appropriate assessor gave an interview report; and assessors provided behavioral examples from each exercise. Each assessor then individually rated the assessee on thirteen dimensions and an overall potential

rating. (These dimensions are defined in Appendix A.) The center director insured that each rating was discussed until consensus was reached. Finally, developmental and supervisory recommendations were discussed, with consensus also required. Following assessment, the center director wrote a feedback report for each assessee based on the consensus ratings/recommendations. This report was provided to the assessee's supervisor and the manpower planning function of the organization; additionally, policy called for a personal discussion of the report with the assessee by either his supervisor or the center staff.

Exercises. Assesseees completed a biographical data form, participated in a background interview with one of the assessors, and took a number of paper and pencil tests during the assessment center. Additionally, the center consisted of three group exercises and an individual oral presentation. The tests consisted of intelligence and personality measures, including the Watson-Glaser Critical Thinking Appraisal, the Miller Analogies Test, the Test of Non-Verbal Reasoning, the 16 PF, the Leadership Opinion Questionnaire, and several others. Group exercises included a candidate nomination, a case study, and a manufacturing simulation exercise. Appendix B includes brief descriptions of each test and exercise.

Assesseees. Data was available for each of 233 assesseees. These assesseees were mostly male (90%), young (mean

age was 29), and experienced with the company (mean tenure was 7 years). Table 2 more fully describes the assesseees.

Table 2
Characteristics of Assesseees

	Mean	S.D.	Range
Age	29	3.97	23-53
Tenure	6.98	3.88	1-27
Grade	10.18	1.36	7-14

Organizational Goals

A crucial requirement of evaluation research, often ignored in validation research, is to determine the organizational goals for the program under scrutiny. Since this process is so important, multiple potential goal sources were isolated; within each category of goal source, multiple respondents were used.

Assessment center goals were obtained from three categories of sources: individual interviews with decision-makers and those responsible for setting up or administering the center; published company documents; and the assessment center literature.

Four individuals within the organization were interviewed using the interview guide in Appendix C. Those individuals include the company psychologist charged with assessment center research, the consulting psychologist who designed the assessment center and participates in the assessor ratings period, the assessment center director

who is also the company's director of manpower planning), and the company vice-president responsible for the center (who was the approval authority for setting up the center.)

Documentary sources for assessment center goals include the assessor orientation manual and the responses of the company to a survey administered by an outside agency. The orientation manual was developed by the consulting psychologist and the assessment center director (who were both interviewed), and by a psychologist no longer associated with the organization. Thus, the manual most likely represents official company goals for the center at the time it was established. Responses to the survey (which was conducted by a professional association with which the company is affiliated) were formulated by the assessment center director. Therefore, these responses probably represent recent official goals for the center.

Additionally, the general literature in the assessment center area was reviewed for evidence of the official goals of other assessment centers. A discussion of this literature was included in Chapter II.

Operational Criteria

The following criteria were used in the assessment center evaluation: (1) number of grade changes since assessment; (2) rate of grade changes since assessment;

(3) number of salary changes since assessment; (4) types of salary changes since assessment; (5) supervisory rating of performance at least one year after assessment; (6) supervisory rating of potential at least one year after assessment; (7) supervisory rating of promotability at least one year after assessment; (8) number of terminations among assessees; (9) test scores on intelligence measures; (10) grades, grade changes, salary, and salary adjustments for randomly chosen nonassesseees who were eligible for assessment; and (11) a survey of assessee satisfaction, motivation, and development, with normative data from six other organizations. The requirement for each of these types of criteria is discussed in the context of organizational goals for assessment in Chapter IV.

Criteria 1, 3, and 4 were obtained from personnel records maintained at the company's various operating locations. Criterion 2 was computed based on criterion 1 and central assessment center records. Criteria 5, 6, and 7 were obtained from the supervisors of each assessee using the form shown in Appendix D; these criteria were extracted from an established company personnel rating system whose results are not available to employees. Criteria 8 and 10 were obtained from a central organizational data base. Tests listed as criterion 9 were administered during the assessment center. Finally, criterion 11 consisted of a

confidential, mail survey. (A copy of the survey instrument is in Appendix E.)

The survey was mailed to 234 still-employed assesseees. Response rate was 64% (n = 150). Surveys were returned directly to a nonorganizational researcher at a university address--the organization has not had access to survey results identifiable by subject. Survey respondents did not differ significantly from the population of all assesseees on the variables of age, tenure with the organization, sex, or overall assessment center rating. Since the survey was part of a different study (Durham, 1978) concurrent with the present evaluation, some survey scales were not relevant or did not directly address the identified organizational goals for the center; however, within the constraints of the existing questionnaire, useful evaluation data can be obtained.

Since assessment center results (including a global overall rating) were specifically provided to the assessee and his supervisor, and were available to higher management at its request, contamination of many of these operational criteria is a potential problem. However, since the formal assessment report did not routinely follow the assesseees after a change in supervisor (caused by promotion, for example), the contamination is minimized. Additionally, although assessment center staff originally anticipated

that management would check on or at least look at central assessment center records before promoting an individual, the assessment center director indicated that that never happened; his feeling was that assessment center results had no operational impact on promotion decisions. The net result of these considerations is that severe criterion contamination is unlikely, but any contamination present would tend to inflate assessment center validity coefficients.

Alternative Programs

Two alternative programs to the assessment center were considered for cost-effectiveness evaluation purposes. These programs were chosen because they were already in place, not because they are the only or even the best alternatives that could be conceived. However, since cost-effectiveness analysis requires evidence of the relative effectiveness of programs, and since it was not feasible to implement a new program using a predictive validity design, only programs already existing could be considered as alternatives. The two program alternatives were paper and pencil test and the company's existing personnel rating system.

Quantifiable test scores were available on each assessee for nine tests (many containing subscales): (1) the

Miller Analogies Test; (2) the Doppelt Mathematical Reasoning Test; (3) The Test of Non-Verbal Reasoning; (4) the Watson-Glaser Critical Thinking Appraisal; (5) the Leadership Opinion Questionnaire; (6) the 16 PF; (7) a Personal Attitude Inventory; (8) an Analysis of Personal Values; and (9) a Personal Classification Test.

Each of these tests, described in the appendix, was administered during the assessment center. Since assessors had knowledge of test results, these tests may have influenced assessment ratings (although personal observation and anecdotal accounts of the assessor meetings where tests were interpreted makes this possibility seem unlikely.) However, since these test results were neither documented in assessment feedback reports nor communicated to supervisors, criterion contamination for this alternative can be discounted. Additionally, since these tests need not be administered in the context of an assessment center, they do represent a viable program that can be considered as an alternative to the full, on-site assessment center.

As part of the company's personnel rating system, each employee receives a confidential supervisory rating on performance, potential, and promotability. These ratings are only communicated upward, not to the employee, and are in addition to the ratings that the supervisor must discuss with the employee. A copy of the most recent

rating given each assessee before assessment was obtained from company personnel records. (A sample form is in the appendix.)

Many of the criteria used here are clearly contaminated by these performance ratings: the ratings are completed by supervisors and become part of the assessee's permanent record, available to any subsequent supervisor or higher management. However, it is still appropriate to compare assessment center predictions with the predictions that could have been made from other sources available at the same time as the assessment center. That is, these ratings represent a true alternative to the assessment center: if the employee had not been assessed, the most recent rating might have been used for the same purposes as the assessment center rating. An evaluation, therefore, should consider the validity of that rating in addition to considering the validity of the assessment center rating.

IV. RESULTS

A major task of the present effort was to determine appropriate evaluation criteria, based on organizational goals for the assessment center; those goals were operationalized from individual interview responses and published documents. The assessment center's performance against each of these criteria was then evaluated; finally, the performance of the alternative programs was evaluated against several of these criteria.

Organizational Goals

Interviews with four organizational decision-makers yielded some convergence in expressed goals for the assessment center; priority listings among those goals varied, however. Additionally, individual recollections of assessment center goals at its establishment varied from current individual goals. At the highest level interviewed, three goals were expressed, and these goals had not changed since the center was established: (1) The center ought to predict the success of high potential employees--since only those rated as high potential attend the center, the center should be able to confirm or raise doubts about

these individuals; (2) The center ought to be a stressful experience--exposure to a high stress situation serves a developmental function for assessees; (3) The center ought to provide a substitute for intelligence testing that is acceptable to forces external to the company.

The goals of the assessment center director differed substantially from those of the executive responsible for the center. The director felt that the center was initially established with three goals, in order: (1) Early identification of employees with the potential for advancement; (2) Early identification of appropriate career paths; (3) Recommendation of appropriate developmental paths. The director felt that priorities among these goals had changed, but not the goals themselves. His present priorities were development, then identification of potential, and finally identification of career paths.

The organization's consulting psychologist, who designed and implemented the center, expressed the same present and initial goals as the center director, and in the same priorities. He did add one additional goal, however, ranked fourth both initially and at present: The center should provide a quality control check on the company's promotion system.

Finally, the psychologist charged with assessment center research expressed three present goals of the center,

in priority order: (1) Developing employees; (2) Early identification of potential; (3) Prediction of employee success. Since this individual was not associated with the center when it was established, he had no personal initial goals for the center.

Company documents provide a further indication of this organization's goals for its assessment center. The company's orientation manual for assessors lists the following main objectives of the center: (1) To identify, early in an employee's career, those career paths within the company that show the most promise for both the employee and the company; (2) To identify those employees that show high promise for successful performance in more difficult job assignments, and to indicate how rapidly individual employees are likely to develop sufficiently for acceptance of more difficult work; (3) To provide practical recommendations, both to management and the employee, regarding the development activity that will help him to achieve his potential. Additionally, in response to a survey on assessment center use conducted by an agency external to the company and unconnected with the present evaluation, the company replied to a question about center goals as follows: To build self-development plans--internal placement recommendations--determine career path planning.

An additional source within the company for determining "the organization's" goals for the assessment center is the perceptions of assessees. As part of the mail survey of prior assessees, employees responded to the open-end question, "What do you think the major purposes of the assessment center are?" In order of frequency of response, assessees chose: (1) Employee development; (2) Early identification of potential; (3) Assessment of individual abilities; (4) Career path planning; (5) Selection for promotion. These five responses constituted 91% of all responses to the question.

Finally, sources external to the organization can provide useful goals for evaluative purposes. The assessment center literature commonly cites early identification of potential and assessee development as assessment goals (MacKinnon, 1975). More specifically, Alexander (1976) reported a rank-ordering of assessment uses in 65 companies, based on a survey. The five top responses were: (1) Identifying strengths and weaknesses of employees; (2) Making promotional decisions; (3) Developing employees with high managerial potential; (5) Aiding in employee career planning.

Table 3 provides a comparative summary of these various sources of goals for this organization's assessment center.

TABLE 3
Sources and Priority Listings of Assessment Center Goals

Priority	Vice-President Present Initial	Director Present Initial	Consultant Present Initial	Researcher Present Initial
1	Predict Success	Predict Success	Identify Potential	Identify Potential
2	Exposure to Stress	Identify Potential	Identify Potential	Identify Potential
3	Intelligence	Intelligence	Intelligence	Intelligence
4	--	--	--	--
5	--	--	--	--

TABLE 3 continued

Priority	Manual	Survey Response	Assessee	Literature
1	Career Paths	Development	Development	Feedback
2	Identify Potential	Placement Decisions	Identify Potential	Promotion Decisions
3	Development	Career Paths	Feedback	Development Plan
4	--	--	Career Paths	Development Experience
5	--	--	Promotion Decisions	Career Paths

Criteria

Operational criteria were chosen for each of the above listed organizational goals. Four criteria were chosen to represent the goal of predicting success: the organizational level that the assessee has reached; the number of promotions the assessee has had since being assessed; the number of nonroutine salary increases since assessment (excludes length of service, cost of living, or other routine, company-wide increases); and the most recent rating of the assessee's performance (in all cases, at least one year after assessment). Each of these criteria represents some indication of the assessee's "success" in the organization. Additionally, it was possible to generally compare the overall salary level of assessees with their nonassessed contemporaries. While this criterion does not address the differential success of assessees among themselves, it does provide some indication of whether generally "successful" employees are being selected for assessment.

The goal of identifying potential was operationalized by three criteria: the number of promotions the assessee has had since being assessed and the number of nonroutine salary increases since assessment (both of which represent interim managerial judgments that the employee has some potential beyond his present position), and the most recent

rating of the assessee's potential (which represents a direct managerial judgment of potential). Correlations among these criteria are reported in the Appendix.

If an organizational goal for a selection system is that the system be used to guide promotional decisions, there should be a high relationship between the selection system's recommendation and the promotion history of the individual. Therefore, the number of promotions the assessee has had since being assessed is a criterion for this goal. Additionally, evidence about the actual use of the selection system data in the promotion decision process is appropriate in evaluating performance on this goal. Similarly, if the selection system is to provide a check on the promotion process, number of promotions and analysis of selection system operational use are appropriate criteria.

The goals of employee development, career-path planning, feedback to assessees, and exposing assessees to stress were operationalized by a survey of assessees. While several of these goals are amenable to different, more direct operationalizations (for example, independent verification of employee developmental activities, career-path changes, etc.), time and data constraints weighed against using criteria other than the survey scales.

Finally, the goal of providing a substitute for intelligence testing can be evaluated by comparing selection

system outcome with performance on a number of standardized intelligence measures. Table 4 summarizes the mapping of these operational criteria.

Assessment Center Performance

The performance of this company's assessment center can be evaluated against each of the criteria listed in Table 4. From these evaluations, or demonstrations of conclusion validity, a summative evaluation of the center's performance with regard to expressed goals is possible.

Predict success. The correlations between the overall assessment rating and four of the criteria comprising the goal of predicting success are shown in Table 5. The assessment center predicts organizational level attained ("Present Grade") reasonably well; the assessment center rating significantly correlates with number of promotions and nonroutine salary increases since assessment, although the correlations are low; finally, the assessment center overall rating has no relationship to a later supervisory rating of performance. Additionally, as shown in Table 6, assessees have a significantly higher annual salary than a randomly selected sample of nonassessed company employees of comparable grades.

Identify potential. The correlations between the overall assessment rating and the three criteria comprising the goal of identifying high potential employees are shown

TABLE 4
Assessment Goals and Criteria

Goal	Criteria
Predict Success	Organizational Level Attained Number of Promotions Number of Salary Changes Performance Rating Salary Comparison with Non-Assesseees
Identify Potential	Number of Promotions Number of Salary Changes Potential Rating
Promotion Decisions	Number of Promotions
Promotion Check	Number of Promotions
Development	Survey of Assesseees
Career Paths	Survey of Assesseees
Feedback	Survey of Assesseees
Exposure to Stress	Survey of Assesseees
Intelligence Test	Intelligence Tests

TABLE 5

Correlations Between Assessment Center Overall Rating
(Mean = .87, S. D. = .62) and Criteria for the Goal of Pre-
dicting Success

Criteria	N	Mean	S.D.	r	p
Present Grade	205	11.79	1.51	.34	.001
Grade Changes	204	1.34	1.23	.18	.005
Salary Changes	205	5.15	1.96	.15	.01
Performance Rating	205	3.60	.82	-.02	n.s.

TABLE 6

Salary Comparison of Assesseees with Non-Assesseees

Status	N	Mean	S.D.	.F(1,960)
Assessed	223	\$14,762	44.34	60.98, p < .001
Not Assessed	739	12,903	58.18	

in Table 7. Again, the correlations between assessment rating and number of grade and salary changes are low, but significant. Additionally, the assessment rating predicts a future supervisory rating of employee potential reasonably well.

Promotions. The goals of both guiding promotional decisions and acting as a quality control check on the promotional system are evaluated by the criterion of number of promotions since assessment. As reported above, the correlation between assessment center overall rating and number of promotions is .18, which is significant with $p < .005$. Beyond this criterion, however, we can look at how assessment center results were operationally used in the promotion process. Specifically, if the assessment results were actually guiding promotional decisions, someone in the process would be checking assessment results whenever an assessee was being considered for promotion--the assessment center director believed that this rarely happened. Similarly, if the assessment center were providing a quality control check on the promotion process, there would be some feedback to decision-makers on the assessment center performance of those employees who had been promoted--no such routine communication channels had been established.

Development. Eight items from the survey of assessees were chosen to represent the goal of employee development.

TABLE 7

Correlations Between Assessment Center Overall Rating
(Mean = .87, S. D. = .62) and Criteria for the Goal of Identifying Potential

Criteria	N	Mean	S.D.	r	p
Grade changes	204	1.34	1.23	.18	.005
Salary changes	205	5.15	1.96	.15	.01
Potential Rating	204	2.50	.65	.37	.001

Intercorrelations, means, standard deviations, and normative means and standard deviations are shown in Table 8. The items themselves, with a breakdown of percentage responses to each item, are included in Appendix F. A summary of the complete breakdown of responses yields the following:

- (1) Only 14% indicated that they were not provided developmental recommendations;
- (2) Less than 25% indicated that their center experience provided more than moderate help in planning self-development efforts;
- (3) About 41% agreed that the center provided valuable information to aid in self-development;
- (4) Less than 25% indicated that they had actually started a self-development program;
- (5) About 40% agreed that the center had resulted in short-term and long-term efforts to improve weaker skill areas and to develop strengths.

Career paths. Two items from the survey were chosen to represent the goal of identifying appropriate career paths for assessees. These items, numbers 35 and 40, correlated with $r = .48$ ($p < .001$). Means, standard deviations, and normative data are shown in Table 9. The items and percentage breakdowns are included in Appendix F. Responses can be summarized as follows:

- (1) Less than 20% indicated that the recommendations they received would be useful in career

TABLE 8
Survey Results for Developmental Goal (N = 150)

	Item Numbers							Assesseees		Norms (N = 460)	
	34	38	39	43-7	43-8	43-9	43-10	Mean	S.D.	Mean	S.D.
8	.28	.31	-.20	.07	.26	.14	.18	3.40	.89	3.23	.96
34		.71	-.41	.13	.27	.10	.33	2.85	.99	2.94	1.05
38			-.52	.17	.22	.19	.21	3.14	1.02	3.26	1.01
39				-.10	-.23	-.20	-.22	1.76	.43	1.74	.44
43-7					.41	.57	.21	2.25	.71	2.25	.66
43-8						.27	.63	2.24	.78	2.34	.74
43-9							.34	2.27	.67	2.21	.63
43-10							1	2.24	.77	2.32	.72

Note-- $|r| < .13$ not significant ($p > .05$)

planning to either a considerable or a great extent;

- (2) Almost 50% indicated that they would place little or no weight on center results in making changes to career plans.

TABLE 9

Survey Results for Career Paths Goal (N = 150)

Assesseees			Norms (N = 460)	
Item	Mean	S.D.	Mean	S.D.
35	3.13	1.30	2.98	1.40
40	2.60	1.10	2.71	1.07

Feedback. Four items from the survey were chosen to represent the goal of assessing individual abilities or identifying strengths and weaknesses of assesseees. Inter-correlations, means, standard deviations, and normative means and standard deviations are shown in Table 10. The items, and percentage response breakdowns, are included in Appendix F. A response summary includes:

- (1) Less than 5% indicated that they did not receive any feedback on their performance;
- (2) Less than 25% indicated dissatisfaction with their formal feedback sessions;
- (3) Less than 1/3 felt that the center had provided a greater awareness of their own abilities;

- (4) More than 50% felt that the center experience had resulted in a better understanding of their own abilities.

Stress. One survey item was used to represent the goal of the assessment center providing a high stress experience for employees (item number 5). This item had a mean response of 3.46 (S. D. = 1.08), indicating a slightly less stressful experience than some other organizations' centers (mean = 3.31, S. D. = 1.12). Essentially, less than 25% of respondents agreed that stress in the center had affected their performance. The item and percentage responses are included in Appendix F.

While this one item is clearly a deficient criterion for operationalizing the goal of providing a stressful experience for assessees (and it would have been substantially reworded if there had been an opportunity to rewrite the questionnaire), other evidence is available that the assessment process is much less stressful than it could be: specifically, the prebriefing given to assessees before the first assessment exercise is designed to reduce stress. The briefer informs assessees that the process by itself will not "make or break" an employee, that they were all already selected for their high potential, and so forth. Thus, even though no direct measure of the stress induced by the assessment process is available, whatever level of

TABLE 10
Survey Results for Feedback Goal (N = 150)

Item Numbers				Assesseees		Norms (N = 460)	
	31	33	43-1	Mean	S.D.	Mean	S.D.
8	.53	.14	.17	3.40	.89	3.23	.96
31		.23	.34	3.91	1.60	4.07	1.52
33			.45	2.65	1.12	2.73	1.17
43-1			1	2.38	.76	2.45	.72

Note-- $|r| < .13$ not significant ($p > .05$)

stress is present could almost certainly be increased very easily.

Intelligence test. One expressed goal for the center was that it ought to provide a substitute for intelligence testing that is acceptable to forces external to the company. While assessment centers in general have shown some acceptability to external forces (e. g., Equal Employment Opportunity Council or the Federal Court System), a necessary first step toward evaluating the assessment center's performance with respect to this goal is to determine the relationship between assessment center performance and several standardized intelligence measures. Table 11 presents means, standard deviations, correlations, and normative data for four intelligence measures and the assessment center overall rating. The measures used were the Watson-Glaser Critical Thinking Appraisal (W - G), the Miller Analogies Test (MAT), the Test of Non-Verbal Reasoning (N - V), and the Doppelt Mathematical Reasoning Test (DOP).

When the four intelligence measures were submitted to a linear multiple regression procedure with the assessment center overall rating as the dependent variable, $R = .372$, $F(4, 119) = 4.19$, $p < .01$. On cross-validation, $r = .31$ ($N = 82$, $p < .01$). Both the simple and multiple correlations between these intelligence measures and the

TABLE 11
Evaluation of Intelligence Testing Goal

	MAT	N-V	DOP	Assessment	Assesseees		Norms	
					Mean	S.D.	Mean	S.D.
W-G	.54	.50	.48	.33	77.88	8.36	74.4	9.6
MAT		.52	.56	.25	42.62	15.02	51.3	18.9
N-V			.51	.21	24.82	6.63	--	--
DOP				.25	21.99	7.64	22.3	7.4
Assessment				1	.87	.62	--	--

Note-- N = 205, all correlations $p < .001$

assessment center overall rating demonstrate that the assessment center rating is not simply capturing test score performance.

Alternative Program Performance

The performance of two alternative programs, paper and pencil tests and the existing company personnel system, was evaluated against several of the expressed goals of the assessment center in order to determine comparative performance.

Predict success. The performance and potential ratings given to each assessee by his supervisor before assessment can be compared with the same criteria used to evaluate the assessment center's performance; similarly, the predictive ability of each assessee's test scores can be evaluated. Table 12 shows the relevant correlations.

For the criterion of organizational level attained (Present Grade), there are several significant predictors; the best of these predictors is clearly the four-test battery. (Tests include: the 16 PF, C and Q3 scales; the WLW Attitudes, Practical scale; and the Miller Analogies Test.) For the criterion of number of promotions since assessment (Grade Changes), again the four-test battery is the best predictor. (Tests in this battery include: the 16 PF, A, F, and Q3 scales; and the WLW Attitudes,

TABLE 12
Correlations Between Alternative Programs
and Criteria for the Goal of Predicting Success

Criteria	Predictors	N	r(or R)	p
Present Grade	Performance Rating	143	.10	n.s.
	Potential Rating	142	.14	.05
	Best Testa	192	.24	.001
	4 - Test Battery	49	.51	.01
Grade Changes	Performance Rating	143	-.03	n.s.
	Potential Rating	142	.01	n.s.
	Best Testb	204	.16	.01
	4 - Test Battery	49	.58	.005
Salary Changes	Performance Rating	143	.14	.05
	Potential Rating	142	-.09	n.s.
	Best Testc	109	.18	.05
Performance Rating	Performance Rating	143	.39	.001
	Potential Rating	142	.08	n.s.
	Best Testd	205	.16	.01

a)LM Attitudes, Aggression
b)6 PF, H Scale
c)Watson - Glaser, Interpretation Scale
d)Non-Verbal

Reliability scale.) Multiple regression summaries and the correlations among all these predictors are included in Appendix G.

The criterion of number of nonroutine salary increases is best predicted by a test score (the Watson-Glaser Interpretation scale); however, even this predictor results in a low correlation. Finally, the criterion of present supervisory rating of performance is best predicted by the supervisory rating of performance that the employee received immediately before being assessed.

Thus, for this goal, alternative programs can significantly predict each operational criterion; except for the criterion of number of salary changes, the correlations between alternative programs and goal operationalizations are reasonably strong.

Identify potential. Alternative program results for the criteria representing the goal of identifying high potential assessees are shown in Table 13. Again, as indicated above, the best predictor of number of grade changes is a four-test battery, while a single test best predicts the number of salary increases (although the latter correlation is low.) The criterion of present supervisory rating of future potential is best predicted by the supervisory rating of potential given immediately before assessment. Thus, each of the criteria comprising this goal is

TABLE 13
Correlations Between Alternative Programs
and Criteria for the Goal of Identifying High Potential

Criteria	Predictors	N	r(or R)	p
Grade changes	Performance Rating	143	-.03	n.s.
	Potential Rating	142	.01	n.s.
	Best Test ^a	204	.16	.01
	4 - Test Battery	49	.58	.005
Salary Changes	Performance Rating	143	.14	.05
	Potential Rating	142	-.09	n.s.
	Best Test ^b	109	.18	.05
Potential Rating	Performance Rating	143	.11	n.s.
	Potential Rating	142	.51	.001
	Best Test ^c	203	.30	.001

^a16 PF, H Scale

^bWatson - Glaser, Interpretation Scale

^cWatson - Glaser, Total Score

significantly predicted by some alternative program. Correlations among predictors and multiple regression summaries are included in Appendix G.

Promotions. The goals of both guiding promotional decisions and acting as a quality control check on the promotion system are partially evaluated by the criterion of the number of promotions since assessment. As reported above, correlations for the alternative programs range from $-.03$ to $.58$, with the best predictor being a four-test battery. The more important evaluation in this case, however, is the knowledge that this test battery (or any of these test scores) cannot be guiding promotional decisions or acting as a promotion check because decision-makers do not have access to the test results. At the same time, the performance and potential ratings that assessees received immediately before assessment are clearly not related to this operationalization of these goals ($r = -.03$ and $r = .01$, respectively, both not significant.) Thus, while any of these alternatives could be used to fulfill these goals, none of them are presently being used in either of these ways.

Development, career paths, feedback, and stress. No direct evidence is available to evaluate the alternative programs against these goals. It is reasonable to assume that those aspects of the personnel system being considered

here as an alternative program would be very ineffective considering these goals since the ratings are never communicated to the employee. Paper and pencil tests, however, might be useful alternatives; so might other, non-confidential aspects of the present personnel system. Unfortunately, data on the impact of these potential alternatives on these goals is not available.

Intelligence Test. Clearly, the alternative program of paper and pencil tests is not appropriate for evaluating against the goal of providing a substitute to an intelligence test--the intent of the goal was to preclude use of standard tests. However, the personnel system can be evaluated against this goal; Table 14 presents those results.

Obviously, ratings of potential are unrelated to intelligence, while performance ratings are affected by much more than just intelligence.

Comparative Summary

Table 15 extracts from the tables above the comparative validities of the assessment center overall rating and the best alternative predictor for each criterion for each goal.

TABLE 14

Correlations Between Alternative Programs
and Intelligence Tests

Program	N	DOP	TEST		
			MAT	N-V	W-G
Performance Rating	143	-.21**	-.19**	-.04	-.16*
Potential Rating	142	.02	.04	-.04	.13

*p < .05

**p < .01

TABLE 15
Comparative Validities

Goal	Criteria	Assessment Validity	Alternative Predictor	Alternative Validity
Predict Success	Present Grade	.34	4-test battery	.51
	Grade Changes	.18	4-test battery	.58
	Salary Changes	.15	best test	.18
	Performance	-.02	performance rating	.39
Identify Potential	Grade Changes	.18	4-test battery	.58
	Salary Changes	.15	best test	.18
	Potential	.37	potential rating	.51
Promotion Decisions	Grade Changes	.18	4-test battery	.58
Promotion Check	Grade Changes	.18	4-test battery	.58
Intelligence Test	Intelligence Tests	.25a	performance rating	.17a

median correlation

V. DISCUSSION

The overall rating given by the assessment center in this organization has been shown to be a valid predictor of several criteria important to individuals in the organization. Appropriate evaluation requires more than a demonstration of several significant validity coefficients, however. The relationship between criteria and goals, the values behind goals, the comparative performance of other programs, and the relative costs involved must all be considered in the move from a validity study to an appropriate evaluation. Consideration of these factors for this organization's assessment center might lead an evaluation researcher to a decision recommendation that differs from that which a researcher whose goal is to "validate" the assessment center might reach.

Validity of the Assessment Center

This assessment center's overall rating of employees is clearly a valid predictor of all of the quantified individual

criteria used here except the current supervisory rating of job performance. Validity coefficients ranged from .15 for predicting the number of promotion and merit salary increases of assessees to .37 for predicting the current supervisory rating of future potential, all significant beyond the .01 level. (There was, obviously, no relationship between assessment rating and current supervisory rating of job performance.) Each of these validity coefficients is consonant with reported validities of other assessment centers using similar criteria.

The lack of a significant predictive relationship between assessment center overall rating and rated job performance is important of itself. Considered in conjunction with the moderate validity of the assessment center with regard to the typical validity criteria of grade attained ($r = .34$) and later ratings of potential ($r = .37$), this finding reinforces Klimoski & Strickland's (1977) observation that "the distinction between performance and progress is not only conceptually viable but it is important empirically as well" (p. 356). Clearly, if it is performance on later (higher level?) jobs that the validity researcher is really interested in predicting, future validity studies must include some type of performance criteria other than just the assessee's arrival at a higher level job.

A "validation" of this assessment center could reasonably stop at this point. (Reported studies normally do stop here, after noting the many collateral uses of assessment--e. g., development of assessees/assessors, feedback, career path recommendations, etc.) The relationships between assessment center rating and several relevant criteria have been measured, and the distinction between performance and progress has been noted. An evaluative approach, however, reveals that many relevant individuals have goals for the assessment center other than predicting success or identifying potential, and characterize these other goals as more important than the usual criteria for establishing validity. Additionally, an evaluative approach forces consideration of alternatives to the assessment center, even if its validity has been demonstrated.

Evaluation of the Assessment Center

An evaluation of this assessment center must consider the multiple goals that are held for the center. While several sources converge on important goals (employee development, identifying high potential employees, predicting success in the organization, and career path planning), it is clear that the priorities among these goals have changed over the life of the center. More important, however, is the distinction between higher management's

goals for the center and anyone else's goals. Consideration of this distinction (and the values behind the distinction) might explain the continued support of the center by higher management, lacking any evidence that the center was effectively accomplishing anyone else's top-priority goal--developing employees!

An evaluative approach must also lead the researcher to look for evidence concerning the center's developmental aspects--the consultant who established the center, the center director, the organization's research psychologist, company documents, and the assessees themselves all designate employee development as the center's top priority goal. The present evidence does not indicate that this assessment center is fulfilling this goal: while assessees are receiving developmental recommendations, less than half feel that this information is valuable, and less than one-fourth report actually having taken any action based on this information. Similarly, the recommendations received from the center are not perceived to be very useful in career-path planning, nor is the feedback on individual abilities received from the center perceived as particularly valuable.

This evaluation with regard to the development/career-path/feedback goals should be regarded as formative evaluation rather than summative evaluation, however. That is,

even given that the center is not meeting these goals, it may be possible to implement procedural changes that will move toward meeting these goals. Additionally, if survey responses indicated that the center was meeting these goals, it would then become desirable in a summative evaluation to establish the causal linkages between assessment center experience, employee developmental activities, and impact of these activities on the employee and the organization.

Higher management's goals of the assessment center being a stressful experience and a substitute for intelligence testing must also be considered. While measured by only one survey item, preliminary indications are that the assessment center is not a particularly stressful experience (at least, as far as assessees are willing to report.) Again, however, this finding should be considered only as formative evaluation for several reasons. First, assessee orientation procedures are purposely designed to minimize stress. Presumably, if management's goal is communicated to the appropriate staff members, the center could easily be made a more stressful experience. Then, once the assessment center became a stressful experience, management's hypothesized relationship between the assessment center, stressful experiences, developmental activities, and impact on the individual and the organization should be evaluated. Finally, the assessment center rating

correlates reasonably well with standardized intelligence measures (median $r = .25$), but this overall rating clearly cannot be characterized as a substitute for any of these measures. Again, however, knowledge of this goal and the center's low relationship to it could be used in a formative evaluation sense to reorient some of the center's exercises to increase this relationship.

At this point, it is appropriate to note some goals that were not expressed for this assessment center--goals that a researcher might reasonably expect to hear. It is not unusual in evaluation research to be confronted with broad, nonspecific goals for social programs; the validation researcher might expect similar problems in obtaining goal statements from organizational decision-makers. For example, the goal of a selection system from top management's view might be expressed only as "To increase the organization's effectiveness." If the researcher is unable to influence management to be more specific, he is in the unenviable position of having to define that goal in his own way; then, of course, he risks acceptance of the results of his evaluation.

In the present evaluation, goal sources (especially top management) were reasonably specific; even then, however, a mapping from goals to operational criteria was necessary. Given the correlations among the chosen criteria

for the expressed goals (Appendix G, Table 16), there is obviously no clear convergence among operationalizations of the same goal. It is not surprising, however, that a goal that hinges on the concept of organizational success might be multidimensional. As previously noted, and supported by the present research, advancement and performance are clearly conceptually distinct components of "success." The validity researcher must be prepared to confront similar situations where alternative programs would differentially predict various operationalizations of the same goal. In that case, evaluation research provides a conceptual answer in cost-benefit analysis, although there are normally extreme difficulties in quantifying and pricing-out benefits. (In the present research, of course, an alternative to the assessment center provides better prediction than the assessment center for each of the quantified operationalizations for the goals of predicting success or identifying high potential.)

An additional issue that should be raised when comparing evaluation research to validation research is the issue of incremental validity. That is, an accepted procedure in validity research is to consider the increase in validity that results from a proposed selection system when it is used in conjunction with the current selection system; in comparison, alternatives in evaluation research

are usually considered as discrete, mutually exclusive--the proposed program will supplant the current one if it is "better", otherwise the current program will continue to be used.

In the present research, if the personnel rating system were considered to be the current program, our evaluation has concluded that it is superior to the proposed program (the assessment center) in predicting several operational criteria; however, the present evaluation has not considered (not provided for decision-makers) evidence about the predictability of these operational criteria if both the present and proposed programs are considered together. Given the low correlations among alternative predictors (including the assessment center) as shown in Table 17 in Appendix G, substantial increases in prediction would be expected by combining the assessment center with other predictors. Then, decision-makers would have the additional opportunity to consider whether the cost of the assessment center was worth the increased predictability--an alternative that the usual evaluation approach would not raise, but a validation approach might. Again, cost-benefit analysis--an evaluative approach--would provide the basis for decision-making.

Cost-Effectiveness of the Assessment Center

After establishing the assessment center's relationship with each goal, it still remains to consider the center's cost-effectiveness. As shown in the comparative summary in Table 15, the best alternate program for each criterion outperforms the assessment center for each goal except that of providing a substitute intelligence test. Beyond stating the effectiveness of each program, however, relative cost data is required to establish cost-effectiveness.

Precise cost accounting data is not available for this center; however, most commentators on the assessment center method acknowledge that it is a high-cost technique. For example, Cayer and Kirschner (1977) in a survey for the Life Office Management Association, found center development costs ranging up to \$25,000 (mean = \$10,000), and costs per assessee ranging up to \$1,300 (mean = \$375). Informal estimates at this center place the cost per assessee at about \$400. (This cost estimate does not include depreciation on facilities, opportunity costs for assessee time, staff salaries, record-keeping, or other costs necessary for a total cost analysis.) Similarly, no estimates are available for costs of alternative programs in this case; however, since all tests considered as alternatives were administered during the assessment center,

it is reasonable to conclude that a testing program costs less than the current assessment program. At the same time, since the other alternative program considered is the normal personnel system of the company (which will operate with or without the assessment center), its cost is not relevant. Therefore, a logical conclusion is that this assessment center is not cost-effective in meeting the goals of predicting success or identifying potential.

As noted above, no alternative programs were considered for the goals of developing employees, providing career-path planning, providing feedback on abilities, or exposing employees to stress; therefore, comparative cost-effectiveness cannot be established for this center for these goals. However, since evaluation has indicated that the center is not effectively meeting these goals, an appropriate alternative to reorienting the center might be to consider other programs for meeting these goals. Then, if there is a decision to change the center toward these goals, it might be possible to simultaneously implement and evaluate other programs as well.

VI. CONCLUSION

This research set out to operate on three levels:

(1) to provide information to decision-makers in the field-site organization; (2) to contribute to the literature on assessment center validity; and (3) to point out and strengthen the relationship between program evaluation research and selection system validation.

Evaluation of an Assessment Center

An evaluative approach toward establishing the validity of this organization's assessment center has been useful: Center ratings predict advancement about as effectively as previous literature indicates; center ratings do not predict future performance, a criterion widely neglected in previous assessment center literature; the center does not appear to be resulting in hoped-for employee development, useful feedback, or more effective career-path planning; the center experience may not be a high-stress environment; and the center is obviously much more than another intelligence test.

Beyond the multiple validity evidence generated by an evaluative approach, however, such an approach forced

the identification of the multiple goals held for this center--goals that may be inconsistent. It seems reasonable that assessee behaviors will be different in a high-stress center with exercises mostly measuring intelligence whose outcome will be a promotion decision than they will be in a tension-free center with exercises mostly measuring interpersonal skills whose outcome will be a developmental plan. This organization clearly needs to consider these differing orientations to its assessment center.

Finally, an evaluative approach has led to consideration of alternatives, and recognition that some alternatives to this assessment center are probably more cost-effective in attaining many of the organization's goals.

Validity of Assessment Centers

This assessment center would be considered "valid" given the manner of reporting previous assessment center validity research; however, lack of a predictive relationship with performance in this case clearly weakens the argument of generalized validity of assessment centers. At the same time, this finding supports the conceptual distinction between progress and performance and confirms suspicions that previous validity studies have focused on too-narrow a class of criteria.

Also, the limited cost-effectiveness analysis performed here would likely replicate in other organizations: In many previous studies, abilities tests, biographical data, personality tests, and ratings have all demonstrated validities comparable to assessment center validities when using an advancement criterion; since most of these other predictors are less expensive than assessment centers, they would probably be more cost-effective in many organizations.

Evaluation and Validation

The approach taken in the present research is not limited to this organization's assessment center, or to assessment centers in general. The contribution--and applicability--of an evaluative approach to selection system validation has been demonstrated. Examination, consideration, and integration of the conceptual issues in program evaluation into the industrial-organizational psychologist's repertoire will result in more relevant, more useful, and more used validity research.

APPENDIX A: Assessment Dimensions

1. Problem Solving Ability: Ability to define the essential nature of a problem, sort the component parts into their proper relationships, and provide a realistic plan of attack.
2. Efficiency Level: Ability, within a relatively unstructured work situation, to identify essential work and properly plan and use time for efficient achievement.
3. Innovativeness: The extent to which the person tries different modes of attack and introduces improved methods.
4. Flexibility: Inclination, when dealing with peers and subordinates, to adapt and change position to meet the needs of varying requirements and conditions.
5. Reaction to Stress: Ability to perform effectively when faced with difficult situations creating circumstances of stress and pressure.
6. Communication Skills--Oral: Ability to convey ideas in a clear and articulate manner.
7. Communication Skills--Listening: Ability to receive and interpret the signals of others.

8. Interpersonal Impact: Ability to elicit a favorable response from others which should lead to cooperative effort.
9. Risk Acceptance: Willingness to act in face of possible negative results.
10. Reaction to Authority: Ability to work effectively with those in authority over him.
11. Application of Authority: Ability to exercise authority in a responsible and adaptive manner.
12. Job Involvement: Extent to which the individual becomes involved in work and can be counted upon to make sacrifices necessary to get the job done.
13. Career Development Orientation: Extent to which the individual has developed realistic career goals and plans for accomplishment.

APPENDIX B. Descriptions of Tests and Exercises

Tests

Watson-Glaser Critical Thinking Appraisal. An intelligence measure with five subtests: Influence, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments.

16 PF. This test was designed as a measure of the major personality factors identified by factor-analytic work by Cattell. Factors are: (1) Cool, reserved--Warm, easygoing; (2) Dull--Bright; (3) Easily upset--Calm, stable; (4) Not assertive--Dominant; (5) Sober, serious--Happy-go-lucky; (6) Expedient--Conscientious; (7) Shy, timid--Venturesome; (8) Tough-minded--Tender-minded; (9) Trusting--Suspicious; (10) Practical--Imaginative; (11) Forthright--Shrewd; (12) Self-assured--Apprehensive; (13) Conservative--Experimenting; (14) Group-oriented--Self-sufficient; (15) Undisciplined--Self-disciplined; (16) Relaxed--Tense, driven.

Test of Nonverbal Reasoning. A short (50 item) intelligence test. Each item consists of ten figures; each of the first four figures is alike in some way, and

the examinee must choose the two figures from the last six that are like the first four.

Miller Analogies Test. A test developed to measure scholastic aptitude for graduate school. The test consists of 100 incomplete analogies; the examinee must choose the correct expression to complete the analogy.

Doppelt Mathematical Reasoning Test. A measure of the ability to perceive mathematical relationships, designed primarily for selecting students for graduate school.

Leadership Opinion Questionnaire. Designed as a measure of the leadership dimensions of Consideration and Structure.

WLW Personal Attitude Inventory. A self-report, forced-choice format attitude inventory with six subscales: (1) Emotional stability; (2) Friendliness; (3) Aggressiveness; (4) Humility and Insight; (5) Reliability; (6) Supervisory Style.

WLW Analysis of Personal Values. A test designed to provide data for discussion in counseling or personal development efforts. Value scales include: (1) Theoretical; (2) Practical-economic; (3) Social; (4) Personal Power; (5) Aesthetic; and (6) Religious.

WLW Personal Classification Test. A short (ten minutes) intelligence test originally designed for use with business executives. The test consists of thirty-five multiple-choice items.

Exercises

Candidate Nomination Problem. In this task, six assesseees assume the roles of members of a selection committee to provide the vice-president of personnel with a best to least rank order listing of candidates for a personnel manager's job. Each assessee has five minutes in which to present information about a candidate to the other five committee members. Following the presentation of all candidates, the group is asked to arrive at a unanimous rank order list of the candidates. The group is allowed fifty minutes for discussion and consensus on the rank order.

Case Study Problem. On the first day of assessment, assesseees are given a business case study problem for their written solutions. They must decide whether to continue to operate a business or sell it to another, willing firm. The assessment staff then divides assesseees into six-person groups, assuring that both alternative solutions are represented. On the second day of assessment, these groups will be instructed to reach consensus on the issue, in fifty minutes.

Manufacturing Exercise. In this exercise, six assesseees are instructed to operate a company that is engaged in the manufacture and sale of token products. Assesseees

choose roles among themselves (sales, manufacturing, accounting, etc.) Participants have 100 minutes during which to plan their product mix, manufacture, and sell their products. Throughout the exercise, the price of raw materials fluctuates, as do the selling prices of the products. Additionally, various communications are transmitted to the group that introduce minor crises or require some other actions.

Speaking Assignment. Assesseees prepare and deliver a five-minute presentation based upon their selection from a number of business journal articles.

APPENDIX C: Interview Guide

1. How did your assessment center come about? That is, whose idea was it? Where did the idea come from?
2. What specifically were your goals for the assessment center when you decided to implement it? What did you hope the assessment center would accomplish?
3. Could you rank order those goals in terms of their importance to you?
4. Have these goals changed at all since the center has become operational? Have you added or taken away any goals? Have your priorities for these goals changed? That is, would you rank order them today in the same order as when the center was first established?
5. IF ACCEPTABILITY TO EXTERNAL AGENCIES, OR STATUS AMONG PEER ORGANIZATIONS WAS IDENTIFIED AS A GOAL-- Is your assessment center meeting this (these) goal(s)?
6. What would be the effects on _____ if the assessment center were to go out of existence?
7. What would be the effects on individuals responsible for the center if it were to go out of existence?

Would people lose their jobs? Would subunit budgets
be cut? Would people quit?

APPENDIX D. Rating Form

Figure 3 is an example of the firm's confidential rating form.

CONFIDENTIAL
SUMMARY EVALUATION OF POTENTIAL

NAME [REDACTED]

EMPLOYEE NUMBER [REDACTED]

POSITION *Master Adjuster*

DATE OF LAST PERFORMANCE EVALUATION *10-1-76*

SALARY GRADE *11*

MOBILITY ☒ YES ☐ NO

OFFICE/REGION *MEMPHIS*

DEPARTMENT *Claims*

SOCIAL SECURITY NUMBER

D [REDACTED]

(LEAVE BLANK)

MONTH	DAY	YEAR

2	2	0
---	---	---

OVERALL PERFORMANCE ON PRESENT JOB

M ☐ **MINIMUM PROFICIENCY** — A trainee or a person with some time on the job who is making minimal progress.

Q ☐ **QUALIFYING** — Performing beyond minimal level, approaching but not yet at full proficiency.

P ☒ **PROFICIENT** — Performing in a fully acceptable (but not distinguished) way in all aspects of the job.

S ☐ **SUPERIOR** — Performing significantly beyond basic proficiency but short of outstanding performance.

O ☐ **OUTSTANDING** — Providing highest contribution that can be made in this job. Top performance.

CURRENT STATUS OF PROMOTABILITY

P ☐ Recommended now for these positions:

R ☒ Remain on Current job

D ☐ Promotable after further development to these positions:

N ☐ Decision deferred due to newness on job

X ☐ Present placement inappropriate—could be better utilized in:

EXPLANATION OF PROMOTABILITY STATUS

[Faint, mostly illegible handwritten text explaining the status]

ESTIMATE OF LONGER RANGE POTENTIAL

The person's potential for handling higher responsibility is:

9 ☒ Limited 5 ☐ Division Level 3 ☐ Department Level 1 ☐ Officer Level

Supervised by	Date	Concurrence	Date	Personnel Concurrence	Date

APPENDIX E. Survey Questionnaire

This Appendix contains the item stems of the confidential mail survey of assessees. The scales used in the present research (items and response alternatives) are defined in Appendix F.

General Reactions to Assessment Center

1. How accurately do you feel your performance at the Assessment Center reflected the way you perform in "real life" situations?
2. At the time you were assessed, what was your understanding of the purpose of the Assessment Center?
3. In your opinion, what should be the purpose of the Assessment Center?
4. Would your performance in the Center have been different if you had more information about its purposes?
5. My performance during the Assessment Center was impaired by feelings of stress or tension created by the Assessment process.
6. My performance during the Assessment Center was impaired by feelings of stress or tension created by situations at home or on the job.

Evaluation of Feedback

7. What type of feedback did you receive?
8. My feedback included developmental recommendations.

AD-A090 795

AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH F/G 5/10
THE RELATIONSHIP BETWEEN PROGRAM EVALUATION RESEARCH AND SELECT--ETC(U)
DEC 79 W J STRICKLAND
AFIT-CI-79-246D

UNCLASSIFIED

NL

2 of 2

9500000



END
DATE
FILMED
11-80
DTIC

9. Did the feedback information include some of the numerical scores or ratings you received during assessment?
10. How was the feedback presented?
11. If feedback was presented orally, were you allowed to take notes?
12. Were you allowed to keep a written copy of the feedback report?
13. At what time was a formal feedback session held?
14. Who was present during the formal feedback session?
15. Did you like having your supervisor present in the feedback session?
16. Did your supervisor participate in the feedback session?
17. Did your supervisor's presence inhibit your participation in the feedback session?
18. Did you like having individuals other than your supervisor or the person providing feedback present during the session?
19. Did the presence of individuals other than your supervisor or the person presenting feedback inhibit your participation in the session?
20. I feel the following persons should be present for the feedback session.
21. Were you asked to write or fill out a critique of the Assessment Center program?
22. Prior to receiving feedback I was asked to
 1. think about my future plans and goals.
 2. write down my future plans and goals and make them available to the person giving feedback.

23. While receiving feedback I was
1. asked to present my future career plans and goals.
 2. asked questions concerning my future career plans and goals.
 3. asked questions related to my understanding of the information being presented during feedback.
 4. asked to give my opinion of how I performed.
 5. asked to give my opinion of why I performed the way I did.
 6. allowed to take notes.
 7. given a written feedback report to keep.
 8. informed of Center numerical scores or ratings.
24. List three dimensions (e. g., skills, traits, variables, etc.) which were mentioned as strong points in your feedback session.
25. List three dimensions (e. g., skills, traits, variables, etc.) which were mentioned as weak points in your feedback session.
26. Considering the results (e..g., ratings, recommendations, decisions) how well do you think you did in the Assessment Center?
27. Were the evaluations of your performance presented at feedback consistent with your self-evaluations of performance?
28. Were your behaviors in the Assessment Center consistent with what they would have been in real life?
29. How would you rate the overall content of the feedback presented?
30. How would you rate the way in which the feedback information was presented during the session?
31. How satisfied were you with your Assessement Center feedback session?
32. How satisfied are you with the result (e. g., ratings, recommendations, decisions) of your Assessment Center experience?

Value of Assessment Center Experience

33. To what extent do you feel the Assessment Center provided you with a greater awareness of your own abilities?
34. How much help do you believe the Center experience has been or will be in planning your personal self-development?
35. To what extent did you find the developmental recommendations useful in formulating career plans?
36. I feel my Assessment Center experience had a positive effect on me personally (i. e., in terms of self-image, motivation, etc.).
37. I feel my Assessment Center experience had a positive effect on my career (i. e., in terms of promotions, salary increases, etc.).
38. I feel the Assessment Center provided valuable information to aid in my own personal development.
39. I started a self or career development program as a result of my Assessment Center experience.
40. In deciding to make further changes in your career plans, how much weight would you place on the results of your performance in the Assessment Center?
41. I ~~feel~~ the Assessment Center program is a valuable tool for motivating participants to improve in the skills required by their job.
42. I feel my supervisor plays a strong positive role in the successful implementation of my developmental plans and programs.
43. My Assessment Center experience has resulted in the following.
 1. A better understanding of personal abilities.
 2. A better understanding of my potential in the company.
 3. A better understanding of my chances for promotion.
 4. A closer bond with other management or supervisory personnel.

5. A short-term increase in my motivation.
 6. A long-term increase in my motivation.
 7. A short-term effort to improve in my weaker skill areas.
 8. A long-term effort to improve in my weaker skill areas.
 9. A short-term effort to develop my strengths.
 10. A long-term effort to develop my strengths.
 11. A new commitment to my organization.
 12. Other (please specify)
44. Considering everything, how would you rate your overall feelings about your employment situation at the present time?
45. What effect do you feel your Assessment Center performance, as rated by the assessors, has had on your current employment situation?

APPENDIX F. Scale Questions

Goal of Employee Development

8. My feedback included developmental recommendations.
- | | | |
|----|-----------------------------|-----|
| 1. | I did not receive feedback. | 4% |
| 2. | No. | 14% |
| 3. | Uncertain. | 18% |
| 4. | Yes. | 63% |
34. How much help do you believe the Center experience has been or will be in planning your personal self-development?
- | | | |
|----|---------------------|-----|
| 1. | Of no use. | 7% |
| 2. | Little help. | 30% |
| 3. | Moderate help. | 37% |
| 4. | Above average help. | 19% |
| 5. | Extremely helpful. | 5% |
38. I feel the Assessment Center provided valuable information to aid in my own personal development.
- | | | |
|----|----------------------------|-----|
| 1. | Strongly disagree. | 5% |
| 2. | Disagree. | 23% |
| 3. | Neither disagree or agree. | 30% |
| 4. | Agree. | 34% |
| 5. | Strongly agree. | 6% |
39. I started a self or career development program as a result of my Assessment Center experience.
- | | | |
|----|------|-----|
| 1. | Yes. | 24% |
| 2. | No. | 75% |
43. My Assessment Center experience has resulted in the following.
- | | UNCERTAIN | NO | YES |
|---|-----------|-----|-----|
| 7. A short-term effort to improve in my weaker skill areas. | 15% | 44% | 40% |
| 8. A long-term effort to improve in my weaker skill areas. | 21% | 33% | 44% |
| 9. A short-term effort to develop my strengths. | 12% | 47% | 39% |

- | | | | | |
|-----|---|-----|-----|-----|
| 10. | A long-term effort
to develop my
strengths. | 20% | 35% | 44% |
|-----|---|-----|-----|-----|

Goal of Identifying Appropriate Career Paths

- | | | |
|-----|---|-----|
| 35. | To what extent did you find the developmental recommendations useful in formulating career plans? | |
| 1. | No developmental recommendations were given. | 13% |
| 2. | Not at all. | 12% |
| 3. | To a limited extent. | 44% |
| 4. | Uncertain. | 11% |
| 5. | To a considerable extent. | 15% |
| 6. | To a great extent. | 3% |
-
- | | | |
|-----|--|-----|
| 40. | In deciding to make further changes in your career plans, how much weight would you place on the results of your performance in the Assessment Center? | |
| 1. | No weight at all. | 18% |
| 2. | Little weight. | 30% |
| 3. | Moderate weight. | 30% |
| 4. | Above average weight. | 16% |
| 5. | Significant weight. | 4% |

Goal of Providing Feedback to Employees

- | | | |
|----|---|-----|
| 8. | My feedback included developmental recommendations. | |
| 1. | I did not receive feedback. | 4% |
| 2. | No. | 14% |
| 3. | Uncertain. | 18% |
| 4. | Yes. | 63% |
-
- | | | |
|-----|--|-----|
| 31. | How satisfied were you with your Assessment Center feedback session? | |
| 1. | A formal feedback session was not held. | 16% |
| 2. | Very dissatisfied. | 4% |
| 3. | Dissatisfied | 20% |
| 4. | Uncertain. | 12% |
| 5. | Satisfied. | 34% |
| 6. | Very satisfied. | 14% |

33. To what extent do you feel the Assessment Center provided you with a greater awareness of your own abilities?

- | | | |
|----|---------------------------|-----|
| 1. | Not at all. | 13% |
| 2. | To a limited extent. | 43% |
| 3. | Uncertain. | 10% |
| 4. | To a considerable extent. | 31% |
| 5. | To a great extent. | 2% |

43. My Assessment Center experience has resulted in the following.

- | | | |
|----|---|-----|
| 1. | A better understanding of personal abilities. | |
| a. | Uncertain | 16% |
| b. | No | 28% |
| c. | Yes | 54% |

Goal of Being a Stressful Experience

5. My performance during the Assessment Center was impaired by feelings of stress or tension created by the Assessment process.

- | | | |
|----|----------------------------|-----|
| 1. | Strongly agree. | 2% |
| 2. | Agree. | 22% |
| 3. | Neither disagree or agree. | 19% |
| 4. | Disagree. | 39% |
| 5. | Strongly disagree. | 16% |

APPENDIX G. Supplementary Tables

TABLE 16
Correlations Among Criteria for the Goals
of Predicting Success and Identifying Potential

	1.	2.	3.	4.	5.	6.
1. Present Grade	1					
2. Grade Changes		.42	.34	-.02	.22	.34
3. Salary Changes		1	.27	-.10	.11	.18
4. Present Performance Rating			1	.18	.10	.15
5. Present Potential Rating				1	.23	-.02
6. Assessment Center Rating					1	.37
						1

TABLE 17

Correlations Among Predictors for the Goals
of Predicting Success and Identifying Potential

	1.	2.	3.	4.	5.	6.	7.	8.
1. Pre-assessment Performance	1	.22	.09	.00	-.14	-.04	-.16	-.05
2. Pre-assessment Potential		1	.41	.26	.07	-.04	.13	.18
3. WLM Attitudes--Aggression			1	.51	-.06	-.03	.01	.21
4. 16 PF--H				1	-.07	-.15	-.08	.26
5. Watson-Glaser--Interpretation					1	.40	.70	.25
6. Non-Verbal						1	.50	.21
7. Watson-Glaser--Total							1	.33
8. Assessment Center Rating								1

TABLE 18

Multiple Regression Summary
for the Criterion of Present Grade

<u>Variable</u>	<u>B</u>	<u>r</u>	<u>R</u>	<u>R²</u>
16 PF, Q3	-.27	-.30	.30	.09
WLW Values, Practical	.33	.26	.42	.18
Miller Analogies Test	.24	.20	.47	.22
16 PF, C	-.20	-.23	.51	.26

$F(4, 44) = 3.90, p < .01$

TABLE 19
Multiple Regression Summary
for the Criterion of Grade Changes

<u>Variable</u>	<u>B</u>	<u>r</u>	<u>R</u>	<u>R²</u>
16 PF, A	-.12	-.29	.29	.08
16 PF, F	.14	.14	.40	.16
WLW Attitudes, Reliability	.13	.22	.46	.21
16 PF, Q3	-.20	-.17	.58	.33

$F(4, 44) = 5.49, p < .01$

LIST OF REFERENCES

- Alexander, L. D. How organizations use assessment center results. Assessment and Development Newsletter, 1976, 2 (2).
- Bray, D. W., & Campbell, R. J. Selection of Salesmen by means of an assessment center. Journal of Applied Psychology, 1968, 52, 36-41.
- Bray, D. W., & Grant, D. L. The assessment center in the measurement of potential for business management. Psychological Monographs. 1966, 80 (17), Whole #625.
- Brogden, H. E., & Taylor, E. K. The dollar criterion: applying the cost accounting concept to criterion construction. Personnel Psychology, 1950, 3, 133-154.
- Burgoyne, J. G., & Cooper, C. L. Evaluation methodology. Journal of occupational psychology, 1975, 48, 53-62.
- Byham, W. C., & Wettingell, C. Assessment centers for supervisors and managers: an introduction and overview. Public Personnel Management, 1974, 3, 352-364.
- Campbell, R. J., & Bray, D. W. Assessment centers: an aid in management selection. Personnel Administration, 1967, 30 (3), 6-13.
- Carleton, F. O. Relationships between followup evaluations and information developed in a management assessment center. Paper presented at the 78th annual convention of the American Psychological Association, Miami Beach, 1970.
- Caro, F. G. Evaluation research: an overview. In F. G. Caro (Ed.), Readings in evaluation research. New York: Russell Sage Foundation, 1971.
- Cayer, M., & Kirschner, C. Personnel evaluation: the assessment center method. Life Office Management Association, Personnel Administration and Research Division Special Release Number 3, 1977.

- Cohen, B., Moses, J. L., & Byham, W. C. The validity of assessment centers: a literature review. Pittsburgh, Pa.: Development Dimensions Press, 1974.
- Cronbach, L. J., & Glaser, G. D. Psychological tests and personnel decisions. (2nd ed.) Urbana: University of Illinois Press, 1965.
- Deming, W. E. The logic of evaluation. In M. Gutentag & E. L. Struening (Eds.), Handbook of evaluation research. Beverly Hills: Sage Publications, 1975.
- Division of Industrial-organizational Psychology, American Psychological Association. Principles for the validation and use of personnel selection procedures, 1975.
- Edwards, W., Gutentag, M., & Snapper, K. In M. Gutentag & E. L. Struening (Eds.), Handbook of evaluation research. Beverly Hills: Sage Publications, 1975.
- Finkle, R. B. Managerial assessment centers. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand-McNally, 1976.
- Freyd, M. Measurement in vocational selection: an outline of research procedure. Journal of Personnel Research, 1923, 2, 215-249; 268-284; 377-385.
- Ginsburg, L. R., & Silverman, A. The leaders of tomorrow: their identification and development. Personnel Journal, 1972, 51, 662-666.
- Glass, G. The growth of evaluation methodology. AERA curriculum evaluation monograph series, No. 7. Chicago: Rand-McNally, 1971.
- Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.
- Guion, R. M. Recruiting, selection, and job replacement. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand-McNally, 1976.
- Guion, R. M., & Gottier, R. F. Validity of personality measures in personnel selection. Personnel Psychology, 1965, 18, 135-164.

- Hinrichs, J. R. Comparison of "real life" assessments of management potential with situation exercises, paper-and-pencil ability tests, and personality inventories. Journal of Applied Psychology, 1969, 53, 425-432.
- Huck, J. R. Assessment centers: a review of the external and internal validities. Personnel Psychology, 1973, 26, 191-212.
- Huck, J. R., & Bray, D. W. Management assessment center evaluations and subsequent job performance of black and white females. Personnel Psychology, 1976, 29, 13-30.
- Jaffee, C. L., Bender, J., & Calvert, O. L. The assessment center technique: a validation study. Management of Personnel Quarterly, 1970, 2 (3), 9-14.
- Klimoski, R. J., & Strickland, W. J. Assessment centers: valid or merely prescient? Personnel Psychology, 1977, 30, 353-361.
- Korman, A. K. The prediction of managerial performance: a review. Personnel Psychology, 1968, 21, 295-322.
- Kraut, A. I., & Scott, G. J. Validity of an operational management assessment program. Journal of Applied Psychology, 1972, 56, 124-129.
- Levin, H. M. Cost effectiveness analysis in evaluation research. In M. Gutentag & E. L. Struening (Eds.), Handbook of evaluation research. Beverly Hills: Sage Publications, 1975.
- MacKinnon, D. W. An overview of assessment centers. Technical Report No. 1, Center for Creative Leadership, 1975.
- McConnell, J. H., & Parker, T. An assessment center program for multi-organizational use. Training and Development Journal, 1972, 26 (3), 6-14.
- Messick, S. The standard problem: meaning and values in measurement and evaluation. American Psychologist, 1975, 30, 955-966.
- Mitchel, J. O. Assessment center validity: a longitudinal study. Journal of Applied Psychology, 1975, 60, 573-579.

- Moses, J. L., & Boehm, V. R. Relationship of assessment center performance to management progress of women. Journal of Applied Psychology, 1975, 60, 527-529.
- Perloff, R., Perloff, E., & Sussna, E. Program evaluation. In M. R. Rosenzweig & L. W. Porter (Eds.), Annual review of psychology. Palo Alto: Annual Reviews, 1976.
- Rossi, P. H. Testing success and failure in social action. In P. H. Rossi & J. Williams (Eds.), Evaluating social programs. New York: Seminar Press, 1972.
- Schuh, A. J. The predictability of employee tenure: a review of the literature. Personnel Psychology, 1967, 20, 133-152.
- Scriven, M. The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), Perspectives of curriculum evaluation. Chicago: Rand-McNally, 1967.
- Suchman, E. Evaluating social programs. In F. G. Caro (Ed.), Readings in evaluation research. New York: Russell Sage Foundation, 1971.
- Thomson, H. A. Comparison of predictor and criterion judgements of managerial performance using the multi-trait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.
- Thoreson, J. D., & Jaffee, C. L. A unique assessment center application with some unexpected by-products. Human Resource Management, 1973, 12 (1), 3-7.
- Wollowick, H. D., & McNamara, W. J. Relationship of the components of an assessment center to management success. Journal of Applied Psychology, 1969, 53, 348-352.
- Worbois, G. M. Validation of externally developed assessment procedures for identification of supervisory personnel. Personnel Psychology, 1975, 28, 77-91.
- Wortman, P. H. Evaluation research. American Psychologist, 1975, 30, 562-575.