

**A META-ANALYTIC VALIDITY STUDY
OF THE NATIONAL ASSOCIATION OF SECONDARY
SCHOOL PRINCIPALS' ASSESSMENT CENTER PROCESS**

by

Linda D. Pantili

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Educational Administration

APPROVED:


Glen I. Earthman, Chairman


M. David Alexander


David J. Parks


Jim C. Fortune


Claire G. Cole

August, 1992

Blacksburg, Virginia

C.2

LD
5655
V856
1992
P366
C.2

**A META-ANALYTIC VALIDITY STUDY
OF THE NATIONAL ASSOCIATION OF SECONDARY
SCHOOL PRINCIPALS' ASSESSMENT CENTER PROCESS**

by

Linda D. Pantili

Educational Administration

Chairman: Glen I. Earthman

(ABSTRACT)

The purpose of this study was to systematically synthesize and analyze through a meta-analytic method the results of various studies pertaining to the validity of the NASSP Assessment Center process. Thirty-four studies were identified as potential sources of information. The analysis focused primarily on criterion-related validity and the potential of gender, position type and racial bias in the assessment process.

Results of the study indicated criterion-related validity of the NASSP Assessment Center process to be very low. Questions were raised about the ability of the assessment process to discriminate among the twelve skill dimensions. The low discriminant validity in the NASSP process probably reduced the criterion validity reported by the various studies. The results indicated the criterion validity of the process was relatively low across most of the criteria used to substantiate validity. No significance relationships were found between the twelve skill dimensions and the demographic

variables of gender, position type, or race. Although thirty-four studies were identified as sources of data for assessing the validity of the NASSP Assessment Center process, there is a need for more research on both construct and criterion-related validity. Criterion-related validity should be based on the application of skills on the job rather than a global measure of job performance.

Acknowledgments

In the search for outer knowledge I have discovered the integration of all the pieces of my life into an inclusive whole on my path to inner wisdom. For this discovery of my destiny of becoming, integrating physically, mentally, emotionally, and spiritually, and of realizing my true potential, I wish to thank with love and appreciation All, visible and invisible, near and distant for guiding and encouraging me to explore without fear.

To my committee, I will always remember the lessons learned. Each of you played a significant part in the completion of this study. Each of you were individually aware of your significance in this chapter of my life. Dr. M. David Alexander, Dr. Claire Cole, Dr. Jimmy Fortune, Dr. David Parks, and my chairperson, Dr. Glen Earthman, I thank you. To the one who entered this quest with me, and accompanied me in spirit to completion, Dr. Loyd Andrew, your presence was recognized.

Friendship accompanied with technical assistance is a grand combination. I was fortunate to encounter that combination. To Richard A. Flanary, John Williams, David D. Bockes, and Paulette Gardner, I have perpetual gratitude. To the many others who shared this journey with me, your smiles, your words, and your caring was often the exact encouragement I needed to move forward. You validated the idea of one never knowing how greatly you may influence the lives of those around you.

To my husband Bill and my son Billy, who did not always understand, but cared enough to let me accomplish, I love you.

Dedicated to my desire to make a difference in this world.

TABLE OF CONTENTS

| | Page |
|--|------|
| Abstract | ii |
| Acknowledgment | iv |
| Chapter 1 | 1 |
| Introduction | 1 |
| Statement of the Problem | 3 |
| Purpose of the Study | 4 |
| Need for the Study | 4 |
| Scope of the Study | 5 |
| Limitations | 5 |
| Definition of Terms | 6 |
| Organization of the Study | 10 |
| Chapter 2 | 11 |
| Literature Review | 11 |
| History of the Assessment Center Process | 11 |
| Development of the NASSP Assessment Center Process | 16 |
| Assessment Center Studies | 23 |
| Studies Not Included in Analysis | 53 |
| Chapter 3 | 76 |
| Methodology | 76 |
| Introduction | 76 |

| | Page |
|--|------------|
| Selection of Studies | 76 |
| Methodology | 77 |
| Meta-Analysis | 79 |
| Chapter 4 | 83 |
| Results | 83 |
| Section One | 84 |
| Skill Dimensions | 84 |
| Reliability | 88 |
| Skill Dimensions and Overall Placement Recommendation | 88 |
| Interrater Reliability | 89 |
| Section Two | 91 |
| Criterion-Related Validity | 91 |
| Section Three | 104 |
| Possible Bias Related to Participants' Gender, Position Type, and Race | 104 |
| Gender | 104 |
| Position Type | 109 |
| Race | 117 |
| Summary | 122 |
| Chapter 5 | 123 |
| Summary, Conclusions, Implications for Practice, and Recommendations for Further Research | 123 |

| | Page |
|---|-------------|
| Summary | 123 |
| Conclusions | 124 |
| Implications for Practice | 125 |
| Recommendations for Further Research | 126 |
| List of Reviewed Studies | 133 |
| References | 137 |
| Vita | 146 |

LIST OF TABLES

| | Page |
|----------|--|
| Table 1 | Meta-Analytic Formulae Used for Effect Size Estimation 81 |
| Table 2 | Interrater Reliability of the Assessment Center Skill Ratings 92 |
| Table 3 | Criteria Effect Sizes (Z_r) Indicated by Individual Studies, Twelve Skill Dimensions and Overall Recommendation Rating 100 |
| Table 4 | Combined Effect Sizes for Criteria 103 |
| Table 5 | Standardized Effect Size Across Studies (Z_r) for Gender 106 |
| Table 6 | Combined Effect Sizes for Gender 110 |
| Table 7 | Standardized Effect Size Across Studies (Z_r) for Position Type 114 |
| Table 8 | Combined Effect Sizes for Position Type 116 |
| Table 9 | Standardized Effect Size Across Studies (Z_r) for Race 118 |
| Table 10 | Combined Effect Sizes for Race 121 |

LIST OF CHARTS

| | Page |
|--|-------------|
| Chart 1 Description of Studies | 72 |

CHAPTER 1

Introduction

Considerable efforts are being expended to improve selection processes of school principals. The school principal is often referred to as a key actor on the school scene. "As the principal goes, so goes the school," is a common belief of many school leaders.¹ The selection of this individual is a crucial responsibility. To establish objectivity in a here-to-fore highly subjective procedure for principal selection, the National Association of Secondary School Principals (NASSP) developed an assessment process to enhance the identification and selection of potentially successful school principals.

The NASSP Assessment Center Process was developed in 1975 by a team of professionals with the assistance of The American Psychological Association. Initial steps in the development of this program included a detailed job description and job analysis of the principalship to determine the skills necessary for effectiveness. This team, soliciting information from numerous practitioners in education, identified 12 skill dimensions believed necessary for success as a school principal: **problem analysis, judgment, organizational ability, decisiveness, leadership, sensitivity, stress tolerance,**

¹Blumberg, A., & Greenfield, W. (1986). The effective principal: Perspectives on school leadership. Boston: Allyn & Bacon, 239.

oral communication, written communication, range of interests, personal motivation, and educational values.

Participants' performances are assessed on 12 skill dimensions.

Trained assessors observe participants in five simulation exercises indicative of activities school principals encounter on the job, and conduct a personal interview. The simulations include two in-basket exercises, two leaderless group activities, and a fact-finding exercise. As with most assessment procedures, the validity of this performance-based assessment is important to participants, principals, boards of education, universities, and state certification boards. Several research studies and two national validation studies have examined the validity of the NASSP Assessment Center Process.²

Multiple types of criteria used in the research of the NASSP process have made it difficult to determine overall criterion-related validity. Research results vary from study to study. Since the early 1970s, business and industry have researched the validity of the assessment process and

²Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP assessment center. East Lansing, MI: Michigan State University, Department of Psychology.
Schmitt N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP assessment center. East Lansing, MI: Michigan State University, Department of Psychology.

aggregated findings across the numerous studies by using the meta-analytic methodologies.³ Hunter, Schmidt, and Jackson aptly described the logic.

Thus the foundation of science is the cumulation of knowledge from the results of many studies. There are two steps to the cumulation of knowledge: (1) the cumulation of results across studies to establish facts and (2) the formation of theories to place the facts into a coherent and useful form.⁴

Just as immense efforts have been put forth in the testing of validity for assessment centers in business and industry, similar efforts are needed to synthesize and analyze research findings across studies to determine the criterion-related validity of the NASSP Assessment Center Process.

Statement of the Problem

Several studies researched criterion-related validity of the NASSP Assessment Process. Criteria varied from study to study. Research findings also varied relating to the different measures used as criteria. The problem is no summarized results exist to establish criterion-related validity of the NASSP Assessment Process based upon the combined research.

³Gaugler, B., Rosenthal D., Thornton, G., & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 72(3), 493-511.

⁴Hunter, J., Schmidt, F., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.

Purpose of the Study

The purpose of this study is to systematically synthesize the research findings from the various studies pertaining to the criterion-related validity of the NASSP Assessment Center Process and to present these results in a coherent summary. The meta-analytic method is used because of its usefulness in standardizing research findings, combining the findings across studies, and evaluating results based on aggregated data.

Three questions concerning validity of the NASSP Assessment Center Process are addressed. (1) What is the validity of the NASSP Assessment Center Process based upon research findings across various studies? (2) How well do the 12 NASSP skill dimensions assess participants' potential for success as prospective school principals? (3) What does research indicate about bias in the NASSP assessment results related to participants' gender, position type, and race?

Need for the Study

The NASSP Assessment Center Process has experienced increased demand as school districts struggle to improve their selection process for school administrators and to seek appropriate programs for professional growth and development. The increased use of the NASSP assessment center has led educators to question the validity of the process. These

questions have prompted many of the validity studies of the NASSP process.

This meta-analytic study addresses questions about the criterion-related validity of the NASSP Assessment Center Process. Synthesis and analysis of findings from the various studies bring focus to the research results. The aggregated research identifies strengths and weaknesses and may initiate modification and refinement of the NASSP Assessment Center Process.

Scope of the Study

An investigation of the research of the NASSP Assessment Center Process identified 34 studies as possible sources of research findings. Studies included in this meta-analysis must have addressed some aspect of validity. The synthesis and analysis of the findings address the criterion-related validity of the NASSP Assessment Process. To investigate possible bias, the meta-analysis determined effect sizes between participants' assessment scores and participants' gender, position type, and race.

Limitations

This meta-analysis is restricted by limitations of comparable research data across studies. Although 34 studies were identified, many of the results were not reported in a form allowing data to be synthesized and

analyzed across studies. Studies in this analysis were not screened or evaluated based upon design, methodologies, scope, limitations, or output. Measures of criterion-related validity of the NASSP Assessment Process provide the basis for inclusion of the studies. Results of this study apply only to the NASSP Assessment Center Process and not to other assessment procedures.

Definition of Terms

Administrator--person assigned by the superintendent of a school district to carry out the administrative, instructional, and managerial duties necessary for the functioning of a school. The terms administrator, school administrator, principal, and school principal have been used interchangeably throughout the various studies.

Assessment Center--a comprehensive, standardized procedure in which multiple assessment techniques, such as situational exercises and job simulations, are used to evaluate individual employees for various purposes.⁵

NASSP Assessment Center--process developed by the National Association of Secondary School Principals involving behavior observation by trained assessors. Candidates participating in five simulation exercises

⁵Thornton, G., & Byham, W. (1982). Assessment centers and managerial performance. New York: Academic Press, 1.

and a personal interview are assessed on 12 skill dimensions related to the principalship to identify participants' potential for success in school administration.

Assesseees--participants in an NASSP Assessment Center.

Assessors--professional staff members who are trained to observe Assessment Center participants, record observed data relating to the 12 skill dimensions, and evaluate the observed data.

Criteria--dependent variables used to determine the correlation coefficient between the criteria and the NASSP Assessment Center Process. Criteria measures included results from attitude ratings, Administrator Perceiver Interview, Administrative Competency Ratings, Diagnostic Survey for Leadership, Paragraph Completion Test, Purdue Teacher Questionnaire, grades, promotions, and job performance ratings.

Dimensions--behaviors that are specific, observable, and verifiable, and that can be reliably and logically classed together in the NASSP assessment process.⁶

NASSP Skill Dimensions--12 behavior dimensions observed and evaluated during the NASSP Assessment Center Process:

⁶Byham, W. C. (1986). Use of the assessment center method to evaluate teacher competencies. Pittsburgh, PA: Developmental Dimensions International, 5.

1. Problem Analysis--ability to seek out relevant data and analyze complex information to determine the important elements of a problem situation, searching for information with a purpose.

2. Judgment--ability to reach logical conclusions and make high quality decisions based on available information; skill in identifying educational needs and setting priorities, ability to evaluate critically written communications.

3. Organizational Ability--ability to plan, schedule, and control the work of others; skill in using resources in an optimal fashion; ability to deal with a volume of paperwork and heavy demands on one's time.

4. Decisiveness--ability to recognize when a decision is required (disregarding the quality of the decision) and to act quickly.

5. Leadership--ability to get others involved in solving problems; ability to recognize when a group requires direction, to interact with a group effectively and to guide them to the accomplishment of a task.

6. Sensitivity--ability to perceive the needs, concerns, and personal problems of others; skill in resolving conflicts; tact in dealing with persons from different backgrounds; ability to

deal effectively with people concerning emotional issues;
knowing what information to communicate and to whom.

7. Stress Tolerance--ability to perform under pressure
and during opposition; ability to think on one's feet.

8. Oral Communication--ability to make clear oral
presentation of facts or ideas.

9. Written Communication--ability to express ideas
clearly in writing; to write appropriately for different audiences-
students, teachers, parents, et al.

10. Range of Interest--competence to discuss a variety
of subjects-educational, political, current events, economic,
etc.; desire to actively participate in events.

11. Personal Motivation--need to achieve in all activities
attempted; evidence that work is important to personal
satisfaction; ability to be self-policing.

12. Educational Values--possession of a well-reasoned
educational philosophy; receptiveness to new ideas and
change.⁷

Position Type--job assignment of the participants at the time of
assessment. Categories for position types were identified as assistant

⁷Lemley, R., & Hersey, P. (1991, July). Assessment handbook.
Reston, VA: National Association of Secondary School Principals, 11.

principal/non-assistant principal, counselor/non-counselor, quasi-administrator/non-quasi-administrator, and teacher/non-teacher.

Precursor--independent, demographic variables analyzed to determine effects upon participants' scores in the NASSP Assessment Center Process. Gender, position type, and race are the precursors analyzed in this meta-analysis to determine possible bias.

Race--ethnic information was recorded as white and non-white in this analysis. Studies reporting results of race based on other categories were collapsed into the white and non-white categories.

Organization of the Study

This study is arranged in five chapters. Chapter 1 introduces the study with an overview of the National Association of Secondary School Principals' Assessment Center Process. This chapter also includes the problem, purpose, need, scope, limitations, and the organization of the study. Detailed descriptions of the 34 studies and background information about the assessment process in business and education are contained in Chapter 2. Chapters 3 and 4 include the methodology and the results of the meta-analysis respectively. Chapter 5 incorporates the summary, conclusions, implications for practice, and recommendations for further research.

CHAPTER 2

Literature Review

Chapter 2 contains reviews of the literature describing the development of the assessment process in industry and education and describes the 34 research studies related to the NASSP Assessment Center Process. The first section includes the history of the assessment process and information about the research of the assessment process in industry. Section two describes the development of the NASSP Assessment Center Process in education and research studies pertaining to the NASSP Assessment Center Process. A detailed description and the research findings of each study are included in this chapter.

History of the Assessment Center Process

German military psychologists began using the multiple assessment approach in the selection of candidates for officers during the 1920s. The American, British, Australian, and Canadian military refined assessment technology during World War II. Other studies by the Harvard Psychological Clinic, the U. S. Office of Strategic Services, and the Veterans

Administration enhanced the methodology and procedures for personnel assessment.¹

The purpose of the industrial assessment center was to study the cognitive, motivational, and attitudinal characteristics that impact managerial careers.² One of the first industrial applications for management assessment was by American Telephone and Telegraph Company (AT&T). Under the direction of Douglas Bray, the Management Progress Study (MPS) of AT&T, the first longitudinal study of its type, followed the development of 422 managers in the organizational setting for four years.³ A summary of this study by Bray et al. found the assessment process accurately predicted the progress of individuals over the following years documenting promotions and salary progress.⁴

An estimated 2,000 companies and organizations assess over 30,000 participants each year for leadership potential.^{5,6} Byham recorded that 75

¹Thornton, G. C. III, & Byham, W. C. (1982). Assessment centers and managerial performance. New York: Academic Press.

²Bray, D. (1985). Fifty years of assessment centers: A retrospective and prospective view. Journal of Management Development, 4(4), 4-12.

³Bray, D. W., Campbell, B. J., & Grant, D. L. (1974). Formative years in business: A long-term AT&T study of managerial lives. New York: Wiley.

⁴Bray, D. W., Campbell, B. J., & Grant, D. L. (1974). Formative years in business: A long-term AT&T study of managerial lives. New York: Wiley.

⁵Gaugler, B., Rosenthal D., Thornton, G., & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 72(3), 493-511.

percent of assessment centers are used to evaluate individuals for three types of positions: first level foremen in industrial settings, sales managers, and clerical supervisors.⁷

The assessment models for management typically include objective, subjective, and situational exercises, as well as exercises to demonstrate personal motivation, practical intelligence, emotional stability, and interpersonal relationships. Specific assessment techniques include interviews, in-basket exercises, leaderless group discussion, and personal history.⁸

Assessment centers are used to select and promote managers and to evaluate supervisory and managerial potential, their developmental needs, or both.^{9,10} MacKinnon attributed the growth of the assessment process in

⁶Wendel, F., & Sybouts, W. (1988). Assessment center methods in educational administration: Past, present, and future. UCEA Monograph Series. Tempe, AZ: University Council for Educational Administration. (Pub. No. 85287-2211)

⁷Byham, W. C. (1985, July). Assessment centers: A method with proven mettle. Training, 69-70.

⁸Thornton, G., & Byham, W. (1982). Assessment center and managerial performance. New York: Academic Press.

⁹Thornton, G., & Byham, W. (1982). Assessment center and managerial performance. New York: Academic Press.

¹⁰Mitchel, J. (1975). Assessment center validity: A longitudinal study. Journal of Applied Psychology, 60(5), 573-579.

industry to the need to identify effective managers, since hiring an incompetent manager was too costly.¹¹

The literature and research have established that assessment centers are successful in making valid predictions of managerial success for business and industry. Many studies were performed for industry to determine predictive validity of the assessment processes, including those by Cohen, Moses, and Byham,¹² Huck,¹³ and Thornton and Byham.¹⁴

The increasing number of validity studies and the wide range of validity results pertaining to the implementation of assessment centers in business led to the application of the meta-analysis as a means to combine the research results across studies. Cooper and Rosenthal described meta-analysis as a method that "helps draw more accurately conclusions about inconsistent findings in a particular area of research. Statistical procedures

¹¹MacKinnon, D. W. (1975, May). An overview of assessment centers (CCL. Tech. Rep. 1). Berkeley: University of California, Center for Creative Leadership.

¹²Cohen, B. M., Moses, J. L., & Byham, W. C. (1974). The validity of assessment centers: A literature review. Monograph II. Pittsburgh: Development Dimensions Press.

¹³Huck, J. R. (1973). A review of the external and internal validities. Personnel Psychology, *26*(2), 191-212.

¹⁴Thornton, G., & Byham, W. (1982). Assessment center and managerial performance. New York: Academic Press.

replace the traditional literature review, which has been criticized for its 'great information-gate-keeping potential.'¹⁵

Literature reviews are highly influenced by the biases of the reviewer and may neglect large amounts of information provided in the original research reports to the amount of research covered. Statistical techniques of aggregation have been suggested as an alternative to the literature review.¹⁶

Meta-analysis of assessment center research for business and industry has been reported by Schmitt et al.,¹⁷ Hunter and Hunter,¹⁸ Gaugler et al.,¹⁹ Cohen,²⁰ Klimoski and Brinkner.²¹ Neuman, Edwards, and Raju

¹⁵Cooper, H., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. Psychological Bulletin, 87, 442.

¹⁶Gaugler, B., Schmidt, F., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage Publishing, 16.

¹⁷Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

¹⁸Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

¹⁹Gaugler, B., Rosenthal D., Thornton, G., & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 72(3), 493-511.

²⁰Cohen, S. L. (1980, February). The bottom line on assessment center technology: Results of a cost benefit analysis survey. The Personnel Administrator, 25(2), 54.

performed a meta-analysis that examined the effects from 126 studies in industry which employed organizational development interventions to modify satisfaction attitudes toward self, others, job, and organization. Neuman's findings indicated results to be situationally specific and therefore subject to the need for local validation.²² Nathan and Alexander analyzed data from a subset of 2,795 validity coefficients for criteria collected by Pearlman.²³ These results supported the practice of combining validity results across different criterion measures when conducting validity generalization research.²⁴

Development of the NASSP Assessment Center Process

"A major challenge facing school systems throughout the United States is the selection of the most competent people to fill key

²¹Klimoski, R., & Brinkner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. Personnel Psychology, 40, 243-260.

²²Neuman, G., Edwards, J., & Raju, N. (1989). Organizational development interventions: A meta-analysis of their effects on satisfaction and other attitudes. Personnel Psychology, 42(3), 461-491.

²³Nathan, B., & Alexander, R. (1988). A comparison of criteria for test validation: A meta-analytic investigation. Personnel Psychology, 41, 517-535.

²⁴Pearlman, K., Schmidt, F., & Hunter J. (1980). Validity generalization results for tests used to produce job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

administrative positions."²⁵ Clear indications of leadership behaviors and skills are necessary to estimate potential. "If a principal is the single most important person in a school . . . then states, local districts, and higher education need to look at how principals are selected, prepared, and rewarded."²⁶ Studies by Martin and Willower²⁷ and Pitner and Russell²⁸ have documented school principal activities. Skills necessary for the principalship included dealing with task, goals, time, environment, service, finance, faculty, students, and staff. With the identification of the necessary skills, the selection of principals could be decided by effectively assessing leadership potential of individuals based upon behavior characteristics and skills.

²⁵Hersey, P. W. (1987). How NASSP helps identify, develop superior principals. Seminar on Innovative Approaches to Assessment to Teaching Staff in the Education Service. Edinburgh, Scotland: NASSP, Reston, VA, 1.

²⁶Southern Regional Education Board. (1986). Effective school principals. A report to the Southern Regional Education Board by Its Commission for Educational Quality. Atlanta, GA: Author.

²⁷Martin, W. J., & Willower, D. J. (1981). The managerial behavior of high school principals. Educational Administration Quarterly, 17(1), 69-90.

²⁸Pitner, N. J., & Russell, J. S. (1985-86). Structured observation of school administration work activities: Methodological limitations and recommendations for research, part I. Educational Research Quarterly, 10(2), 13-24.

The use of assessment processes in education began in the 1960s. McIntyre²⁹ and Culbertson³⁰ were early researchers of selection criteria for administrative competencies. Earlier research studies by Gallagher,³¹ Merino,³² Reighard,³³ Streitman,³⁴ and Zubay³⁵ looked at the application of the assessment process in the field of education prior to the involvement of National Association of Secondary School Principals (NASSP). Goodlad addressed the importance of principal selection and suggested the use of the assessment center process from the private sector

²⁹McIntyre, K. E. (1966). Selection of educational administrators. A University Council for Educational Administration (UCEA) Position Paper, Columbus, OH.

³⁰Culbertson, J., Farguhar, R. H., Gaynor, A. K., & Shibles, M. R. (1969). Preparing educational leaders for the seventies. Columbus, OH: The University Council for Educational Administration.

³¹Gallagher, M. (1978). The development and pilot operation of an assessment center with implications for the selection of school principals. Unpublished doctoral dissertation, Western Michigan University.

³²Merino, A. (1973). The development of an assessment center for the selection of school administrators. Unpublished doctoral dissertation, University of Massachusetts.

³³Reighard, C. H. (1977). A procedural model for selecting elementary school principals utilizing the assessment center technique. Unpublished doctoral dissertation, Miami University.

³⁴Streitman, W. H. (1975). The use of simulation techniques to identify potentially effective educational administrators. Unpublished doctoral dissertation, Georgia State University.

³⁵Zubay, A. H. (1976). The use of the assessment center to identify potentially effective educational administrators. Unpublished doctoral dissertation, Georgia State University.

in A Place Called School.³⁶ NASSP became involved in assessment in the mid 1970s to assist in the selection of elementary and secondary school administrators.

NASSP, with the assistance of The American Psychological Association, established the first pilot assessment center for education. Initial steps in the development of this program included a detailed job description and job analysis of the principalship to determine the skill dimensions necessary for effectiveness. Following the selection of the 12 skill dimensions, appropriate simulation exercises were developed wherein the skill dimensions could be evaluated. By 1979, NASSP Assessment Centers had been established in nine school districts. NASSP awarded a contract to Michigan State University in 1979 to conduct a national validation study to demonstrate content and predictive validity of the NASSP Assessment Center Process. A second national validation study was completed in 1990 to again assess the validity of the NASSP Assessment Center program.³⁷

³⁶Goodlad, J. I. (1984). A place called school: Prospects for the future. New York: McGraw-Hill.

³⁷Flanary, R. (1991, December). History of the NASSP Assessment Center. Unpublished manuscript, Blacksburg, VA.

The assessment process has "withstood the test of time" from the research perspective.³⁸ Also of major importance is the fact that assessment centers have likewise "withstood the test of time" in regard to legal aspects of the process. "The body of case law on employment testing generally was well developed before any extensive use of Assessment Centers to screen educational personnel."³⁹

In 1980, William C. Byham completed his Review of Legal Cases and Opinions Dealing With Assessment Centers and Content Validity, describing 18 court cases in which assessment centers were cited.⁴⁰ Floyd G. Delon stated, "Comparing the validation of the NASSP Assessment Center with the requirements stipulated in the *Grigg, Davis, and Albemarle Paper Company* decisions, the courts are most likely to hold that states are legally justified in using the process in connection with job placement."⁴¹ Two cases, Edwards v. City of Evanston et al. and Macchiarola et al. v. New York City

³⁸Cohen, S. L. (1980, February). The bottom line on assessment center technology: Results of a cost benefit analysis survey. The Personnel Administrator, 50-56.

³⁹Thomas, T. (1983). The testing of teachers' legal parameters and procedures, school law-update 1982. 189-202.

⁴⁰Byham, W. (1979). Review of legal cases and opinions dealing with assessment centers and content validity. Pittsburgh, PA: Development Dimensions International.

⁴¹Delon, F. G. (1988). Assessment center screening for school administration certification and employment: Possible legal challenges. West Publishing Company, 43 W. Educ. L. Rep. 841.

Board of Examiners et al., are examples of the courts mandating assessment centers for selection of personnel.⁴² The assessment center process has also been beneficial in addressing the Equal Employment Opportunity Commission (EEOC) regulations. Byham reported the process to be "uniquely fair to women and minorities because of its emphasis on actual behavior rather than psychological constructs."⁴³

The NASSP Assessment Center Process has been implemented in numerous school districts with more than 60 centers located throughout the United States, Canada, England, Germany, and Australia. The process has been mandated by law in South Carolina,⁴⁴ Kentucky,⁴⁵ and Missouri.⁴⁶ Maryland and New Jersey have mandated the assessment process through certification requirements. Gousha, Lopresti and Jones in 1984-85 reported that 89 percent of state departments of education, school districts, universities, and colleges were altering or seriously considering changing

⁴²James C. Edwards et al. v. City of Evanston et al. (1975) and Frank J. Macchiarola et al. v. New York City Board of Examiners et al. (1981), Civil Action No. 74C 2686.

⁴³Byham, W. C. (1986). Use of the assessment center method to evaluate teacher competencies. Pittsburgh, PA: Development Dimensions International, 38.

⁴⁴S.C. Code Ann., ch. 24, 59-24-10 (1984).

⁴⁵KY ST s 156.105 (1990).

⁴⁶Mo. Rev. Stat. #168.405 (1985).

their standards, criteria, and selection procedures for school administrators. The assessment process was considered in many of these changes.⁴⁷

Many of the changes involving school administrators were by-products of the effective schools research of Brookover,⁴⁸ Edmonds,⁴⁹ and Sergiovanni.⁵⁰ Research indicated positive relationships exist among successful administrators, organizational effectiveness, and effective schools. Melton and Miller stated, "We see a strong relationship between the leadership of the principal and effective schools."⁵¹ Concern for the mediocrity in schools during the early 1980s was expressed in the National Commission on Excellence in Education, Nation at Risk Report; the Governors' Task Force report, Action for Excellence; and John Goodlad's

⁴⁷Gousha, R., Lopresti, P., & Jones, A. (1987). Administrators join revolution in professional standards. The School Administrator, 44(2), 10-15.

⁴⁸Brookover, W. B., & Lezotte, L. W. (1979, May). Changes in school characteristics coincident with changes in student achievement. East Lansing, MI: Institute for Research on Teaching, Michigan State University.

⁴⁹Edmonds, R. (1979, October). Effective schools for the urban poor. Educational Leadership, 37(1), 15-18, 20-24.

⁵⁰Sergiovanni, T. J. (1987). The principalship: A reflective practice perspective. Boston: Allyn & Bacon.

⁵¹Melton, J., & Miller, R. (1987, February). Assessment project makes major milestone. The School Administrator. Arlington, VA: American Association of School Administrators, 20.

report, A Study of Schooling.⁵² The American public was demanding a more effective and efficient school system.

The NASSP Assessment Center Process was implemented as part of these changes. One way to address the increased demands placed upon school systems was to ensure the validity of the measures used in the selection of school principals. The increased use of the NASSP Assessment Center Process made it appropriate and important to bring focus to the results of the criterion-related validity studies.

Assessment Center Studies

A total of 34 studies were identified as possible sources of validity data related to the NASSP Assessment Center Process. The studies which specifically addressed NASSP Assessment Center validity are described in the beginning of this section. Studies that address other research aspects related to the NASSP Assessment Center Process are described in the latter part of this chapter.

⁵²National Commission on Excellence in Education. (1983). A nation at risk: The imperative for education reform. Washington, DC: U.S. Government Printing Office.

Blanck compared NASSP assessment results of participants for five educator groups in Michigan.⁵³ Do participants, serving in a position as assistant principal or quasi-administrator, have an advantage of scoring higher than classroom teachers when assessed by the assessment center? The research question addressed the possibility of significant differences among mean scores on 12 skill dimensions of the five role groups in Michigan. This study identified skill dimensions which indicated significant differences of performance among five role groups in Michigan.

The 394 participants, elementary teachers (53), middle school teachers (47), high school teachers (38), assistant principals (140), and quasi-administrators (116) were assessed by the Michigan Principal Assessment Center (MPAC) using the NASSP process. Quasi-administrators were defined as counselors, department chairpersons, psychologists, and consultants. Blanck described the relationship of job assignment and assessment performance and extended the research to include content validity of the NAP Assessment Center Process.

Significant differences among all groups were identified in **problem analysis, organizational ability, decisiveness, leadership, oral communication, and overall placement recommendation, $p < .05$** . No significant differences

⁵³Blanck, J. I. (1989). Comparison of five role groups of Michigan participants in NASSP assessment centers. Unpublished doctoral dissertation, Western Michigan University.

were found among the three teacher groups on any of the 12 skill dimensions. The data indicated the three teacher groups to have equal opportunities for success based upon outcome performances from the NASSP Assessment Center Process.

There were significant differences among quasi-administrators and the three teacher groups in the skill dimensions of **organizational ability, problem analysis, leadership, oral communication, and overall placement recommendation**, $p < .05$. The teacher groups scored significantly lower than the quasi-administrator group on these dimensions. Teachers did not score significantly lower than assistant principals. Significant differences were not found among scores of quasi-administrators and assistant principals.

As a means of supporting construct validity of the NASSP Assessment Center Process, Blanck compared Standards of the NASSP Assessment Center with the Standards and Ethical Considerations for Assessment Center Operation as adopted and revised by the International Congress of the Assessment Center. The Standards and Ethical Considerations for Assessment Center Operation serves to monitor assessment activities throughout various centers. Each standard recommended by the International Congress was addressed within the NASSP Assessment Center format.

Blanck also compared the results of her study to the results of the national validation study.⁵⁴ This comparison identified significant differences among the five groups in **problem analysis, organizational ability, leadership, decisiveness, oral communication, range of interests, and personal motivation**. Blanck attributed the differences to the method of grouping the participants. The national study compared two groups, i. e. assistant principal and non-assistant principals, quasi-administrator and non-quasi-administrator. Blanck used five role groups for her study, allowing for increased opportunity to identify more areas of differences and more exact comparisons.

Breed compared the relationship of scores on the Administrator Perceiver Interview (API) and results of the NASSP Assessment Center Process.⁵⁵ Breed calculated concurrent validity based on performances on two procedures used in the selection of school administrators, API and assessment center results. The sample group consisted of 19 educators from the University of Nebraska-Lincoln Assessment Center Project.

⁵⁴Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. Reston, VA: NASSP Assessment Center.

⁵⁵Breed, R. D. (1985). Concurrent validity of two administrator selection procedures. Unpublished doctoral dissertation, University of Nebraska.

Breed compiled means and standard deviations based on the 12 NASSP skill dimensions and overall placement recommendation. The sample group from the project scored slightly lower than the results reported by Schmitt in the national validation study.⁵⁶ Mean scores and standard deviations were also compiled for the API using the identified themes: mission, human resource development, relator, delegator, arranger, catalyzer, audience sensitivity, group enhancer, discriminator, performance orientation, work orientation, and ambiguity tolerance.

Analysis indicated no significant correlations among scores on the API themes and scores from the NASSP Assessment skill dimensions. Correlations among NASSP Assessment Center skill dimensions and the API scores were very near zero. The six areas showing significant correlations of the possible 156 correlations were: **decisiveness/human resource development; leadership/group enhancer; judgment/performance orientation; written communication/delegator; oral communication/discriminator; and oral communication/work orientation, $p < .05$.**

The API theme group enhancer and the NASSP skill dimension **leadership** seem to measure similar information and had the highest significant and positive correlation coefficient. The findings by Breed did not support the concurrent validity of the API themes and the NASSP

⁵⁶Breed, R. D. (1985). Concurrent validity of two administrator selection procedures. Unpublished doctoral dissertation, University of Nebraska.

Assessment Center Process. Results indicated performance on the API provided different information than information gleaned from the NASSP Assessment Center performance.

Dickson considered participants' career aspirations, perceived career opportunities, and professional development plans in relation to their Assessment Center experiences.⁵⁷ Dickson analyzed survey results using the independent variable of gender. Significant differences between female and male performance ratings indicated females typically out-performed males in overall performance recommendations. No significant differences were indicated in overall performance recommendation related to size of the school districts. Factor analysis by Dickson indicated two possible constructs apparent with center results: instrumental and expressive. Dickson cautioned that respondents from this study tend to be those participants who performed well in the center rather than those who had performed poorly.

Douglas and Johnson examined the 12 skill dimensions for potential bias in the NASSP Assessment Process based upon gender, degrees held,

⁵⁷Dickson, K. G. (1987). The Oregon assessment center: Its influence on candidates' career opportunities and professional development. Unpublished doctoral dissertation, University of Oregon.

position type, race, and years of experience of participants.⁵⁸ Participants' scores were compared to determine bias based up gender. Correlations for **problem analysis, organizational ability, leadership, written communication, and overall placement recommendation** were significantly higher for females than for males.

Douglas and Johnson examined possible bias of the NASSP assessment results based upon position type of participants. Significant differences were noted between the scores of teachers and non-teachers in **stress tolerance and educational values**. Teachers' scores were significantly lower than non-teachers' scores. No effects were indicated between degrees held or years of experience for the **overall placement recommendation** and scores on the 12 skill dimensions for participants. Analysis indicated significant differences in participants' scores by race for **decisiveness**. No significant differences were found in the remaining skill areas.

Douglas and Johnson determined the interrater reliability for the 12 skill dimensions and overall placement recommendation. Interrater reliability exceeded an $r = .80$ for nine of the 12 dimensions. **Decisiveness, range of interests, and educational values** were dimensions having r 's $< .70$. Douglas

⁵⁸Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300 879)

and Johnson also examined intercorrelations among the 12 skill dimensions and the overall placement recommendation. The skills having the lowest correlations with the **overall placement recommendation** were **range of interests** and **stress tolerance**. The skills having the highest correlations with **overall placement recommendation** were **judgment, problem analysis, organizational ability, leadership, written communication, and oral communication**.

Analysis indicated **problem analysis, judgment, organizational ability, and leadership** to be the skill dimensions having the highest intercorrelations indicating a lack of discrete measurements for each skill area. Intercorrelations of the 12 skill dimensions were all significant except for **range of interests, educational values, and stress tolerance** with **decisiveness**. These results indicated multicollinearity of the skill dimensions and questioned the potential of the assessment process to measure 12 separate and distinct skills.

Douglas and Johnson analyzed results for the 12 NASSP skill dimensions and overall placement recommendation related to five demographic variables. Assessment scores indicated that gender had the greatest impact. Females scored higher than males in **problem analysis, organizational ability, leadership, written communication** and **overall placement recommendation**. Race differences were significant in the skill dimension of **decisiveness** with non-white participants scoring lower than

white participants. Teachers' scores were significantly lower than quasi-administrators on **stress tolerance** and **educational values**.

Ehinger examined content, construct, and predictive validity of assessment center activities.⁵⁹ Incumbent public school administrators (183) participated in the Skill Review Phase of the Education Management Development Program (EMDP) at The University of Tulsa. Content validity was examined by a group of experienced school administrators serving as recognized and recommended consultants. The consultants determined appropriateness of activities related to a school administrator's responsibilities and provided guidance in establishing performance criteria.

Ehinger's conclusions supported construct validity of the assessment center activities administered during the Skill Review Phase of the EMDP. Ehinger expressed the possibility of a "halo" effect in regard to high ratings in all of the different dimensions. Discriminant validity appeared to be the weakest aspect of the assessment process indicated by multicollinearity of the skill dimensions.

Criterion-related validity was determined by comparing participants' job performance ratings by superintendents to participants' scores from the assessment center. Assessment scores were found to contribute

⁵⁹Ehinger, J. M. (1986). The validation of selected assessment center activities for incumbent school administrators who participated in the University of Tulsa Education Management Development Program. Unpublished doctoral dissertation, University of Tulsa.

significantly to the variance in superintendents' ratings. Validity was examined across superintendents' ratings and assessment results from the 1984 sample and used to predict superintendents' ratings for the 1985 group. Significant correlations were found for predictions of superintendents' ratings of prospective principals based upon assessment results. Ehinger indicated that by using assessment scores, superintendents could predict job performances of principals.

Farmer investigated five areas: (1) the extent to which 53 participants perceived the assessment procedures as acceptable, (2) principals' and teachers' perceptions of participants compared to ratings from the assessment center, (3) the extent to which job-related conditions permit utilization of the 12 skill dimensions, (4) unanticipated results of the assessment center, and (5) order of importance the participants ranked the skill dimensions for the job roles of teacher, assistant principal, and principal.⁶⁰

Teachers did not give overall endorsement. Prior information concerning the center and a fair opportunity in the total NASSP Assessment Center Process were concerns of the teachers. The teachers' opinions often

⁶⁰Farmer, T. N. (1987). An analysis of the effects of participation in the National Association of Secondary School Principals' Assessment Center in Howard County, Maryland. Unpublished doctoral dissertation, University of Maryland.

depended upon whether they received "recommend" or "not recommend" scores to be placed in the pool of potential assistant principals.

Assistant principals who had high assessment scores and were either recommended and promoted or recommended and not promoted agreed with the opportunity to use skills on the job. Assistant principals who had low scores and were not recommended disagreed with the opportunity to use the 12 skills on the job. Both groups of participants agreed with the school districts as to the skills having great importance for administrative jobs, as well as the importance of integrating the assessment approach into the selection and promotion functions.

In addressing construct validity of the assessment process, Farmer indicated that non-assessor trained principals did not give higher scores on the assessment skills than assessor trained principals using an on the job performance rating. Data also indicated that principals' ratings of participants were significantly higher than the NASSP assessment ratings in **problem analysis, judgment, organizational ability, written communication, range of interests, and educational values**. Results indicated that given an extended period of time to observe assistant principals and teachers, principals could present ratings similar to those of the assessment center, supporting construct validity of the process.

Teachers and assistant principals disagreed on the use of assessment results for promotional decisions. Assistant principals tended to be more

positive about the use of the results for promotions. Although both groups wanted the results included in the selection process, they did not want them to receive more weight than job performance and interview results. Teachers indicated low agreement that teachers' jobs provide experiences and training in the 12 skill areas.

Farmer indicated that assistant principals and teachers, whether promoted or not, agreed upon the fairness and objectivity of the NASSP Assessment Center Process. Those participants who were recommended for promotion and those recommended but not promoted endorsed the use of the results as part of the selection process, yet both disagreed on the use of the results as sole criterion for promotional decisions. Those not recommended and not promoted did not advise the use of NASSP Assessment Center results for promotion. Ford researched NASSP Assessment Center participants' attitude ratings toward the assessment process.⁶¹

Ford surveyed attitude ratings regarding the assessment process, overall management of the assessment center, the NASSP skill dimensions, and assessment exercises. Analysis of the participants' ages and participants' attitude ratings found no significant differences in attitude

⁶¹Ford, M. J. (1987). An attitude survey of NASSP principal assessment center candidates. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

ratings toward the assessment process, overall management of the assessment center, and the assessment exercises. There were, however, significant differences in attitude ratings related to age of the participants when comparing the skill dimensions of **judgment** and **overall placement recommendation**. Participants 36 years and older, scoring low in **judgment** and low on placement ratings had a positive attitude rating; whereas, those participants ages 26 to 35 scoring lower on **judgment** and **overall placement recommendation** had a significantly lower attitude rating.

Ford analyzed the participants' attitude ratings toward the assessment process and participants' gender. Ford found significantly higher positive mean attitude ratings for females than for males in the assessment process, assessment center management, and assessment exercises. Overall significant differences were identified between female and male attitude ratings toward all skill dimensions except **range of interests**. No significant differences in attitude ratings were indicated when both females and males received positive recommendations and higher skill dimension ratings. A two-way ANOVA was used to analyze the effect of position type for participants' attitude ratings toward the assessment process. No significant differences were indicated among attitude ratings and participants' position types to the assessment process, assessment center management, and assessment exercises.

Ford also analyzed the effect of race and participants' attitude ratings toward the assessment process. Significant differences in attitude ratings were found between white and non-white participants who scored low in **decisiveness**. Non-white participants scoring low in **decisiveness** had positive attitude ratings toward the assessment process whereas the white participants scoring low had negative attitude ratings. No significant differences were found in the attitude ratings of white and non-white participants for the remaining eleven skill dimensions, assessment center management, or the assessment exercises.

An ANOVA was used by Ford to analyze the effects of regions of the country and attitude ratings of participants toward the assessment process. Five regions were identified: Northeast, Southeast, Northwest, Southwest, and Canada. No significant differences in attitude ratings of participants were identified among NASSP assessment with regard to the regions of the country in which they lived.

There were, however, significant differences among attitude ratings and scores in **written communication, overall placement recommendation,** and region of the country. For **written communication**, the scores of the Western region fluctuated significantly among low, average, and high while attitude ratings remained relatively positive. For Canada and the Eastern regions, attitude ratings increased with an increase in **written communication** scores.

Low placement recommendation scores from the Western region were related to negative attitude ratings. Low placement recommendation scores from the Eastern regions and Canada resulted in neutral attitude ratings. All five regional groups scoring high on the placement recommendation were rated high on attitude ratings. Participants from the Western regions tended to be less positive toward the assessment process. No significant differences were noted in attitude ratings among the overall management of assessment centers and assessment exercises.

Ford analyzed the effect of years of experience and participants' attitude ratings toward the NASSP Assessment Center Process. Ford divided 254 respondents into three groups according to years of experience: group 1 (0-10 years, n = 65); group 2 (11-15 years, n = 87); and group 3 (16 plus years, n = 102). Ford found no significant difference among years of experience and attitude ratings of participants toward the assessment process. Results indicated positive attitude ratings toward the process from all three groups. No significant differences were indicated among years of experience and attitude ratings toward assessment center management and assessment exercises. Significant differences were identified among years of experience of participants and attitude ratings for **decisiveness** and **overall placement recommendation**. Group two with 11-15 years of experience and low scores in **decisiveness** had negative attitude ratings toward the process. Groups one and three with low **decisiveness** scores had neutral attitude

ratings, whereas those with average or positive scores had positive attitude ratings toward the skill dimensions. **Overall placement recommendations** indicated significant differences among attitude ratings and years of experience. Groups one and three had positive correlations among increased **overall placement recommendation** and increased attitude ratings. Participants in group two receiving negative placement recommendations also had negative attitude ratings.

Ford analyzed the dependent variable of attitude ratings of the participants across the 12 skill dimensions and the four types of assessment exercises, (in-basket, leaderless group, fact-finding, and personal interview). Attitude ratings toward the four exercises were positive. Significant interactions occurred among five skill dimensions and **overall placement recommendation** influencing participants' attitude ratings toward the assessment process. The five skill dimensions showing significant interactions with **overall placement recommendation** were **problem analysis, judgment, leadership, written communication, and range of interests**. The analysis indicated the higher the **overall placement recommendation** and any one of the five skill dimensions scores, the more positive the attitude ratings of the participants. Conversely, the lower the scores on the skill dimensions and **overall placement recommendation**, the lower the attitude ratings.

Overall management of the NASSP Assessment Center was viewed by all participants as positive. Responses indicated the importance of the

director's role in making the experience a positive one. Being properly informed prior to attending the center received the lowest ranking. Ford indicated **judgment** and **decisiveness** to be the skill dimensions having the most influence on the difference in attitude ratings of the participants. Age, position type, race, region of the country, and years of experience did not make a difference in terms of overall attitude ratings. Gender analysis indicated females had better attitude ratings toward the process than males. The analysis indicated the higher the **overall placement recommendation** the more positive the attitude rating toward assessment.

Heater validated the NASSP Assessment Process using the Purdue Teacher Opinionaire (PTO), a measure of teacher morale, as a criterion measure.⁶² The study assumed that some factors of teacher morale were influenced by school administrator behaviors as measured by nine of the 12 skill dimensions assessed by the NASSP Assessment Center Process. NASSP skill dimensions were grouped into the three clusters of administrative, interpersonal, and communication skills as suggested by Schmitt to reduce large numbers of intercorrelations.⁶³ Factors of the PTO

⁶²Heater, J. D. (1989). Teacher morale as a criterion of the predictive validity of the National Association of Secondary School Principals' assessment center. Unpublished doctoral dissertation, University of Nebraska.

⁶³Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. Reston, VA: NASSP Assessment Center.

included: teacher rapport with the principal, satisfaction with teaching, rapport among teachers, teacher salary, teacher load, curriculum issues, teacher status, community support, school facilities and services, and community pressure.

Statistical analyses were conducted using three NASSP clusters of nine skill dimensions, PTO scores, and demographic variables: length of teacher service, length of time teacher served under principal, teacher's educational level, and length of time since principal was assessed. The conclusions based on this analysis found no significant correlations to support any type of criterion-related validity among the three cluster groups of nine NASSP assessment dimensions, the Purdue Teacher Opinionaire, and the two demographic variables. Heater determined that teacher morale was not an appropriate criterion measure to validate the NASSP Assessment Center Process.

Holman analyzed the relationship among conceptual level and participant performance in the NASSP Nebraska Assessment Center.⁶⁴ Holman researched the difference in performance on the 12 skill dimensions related to high and low conceptual levels, gender, and occupational levels. The Paragraph Completion Test was the criterion used to measure the

⁶⁴Holman, D. M. (1986). The relationship of participant performance in the Nebraska Assessment Center to participant conceptual level/cognitive complexity. Unpublished doctoral dissertation, University of Nebraska.

conceptual level of 100 participants who had been assessed at the Nebraska Assessment Center. The conceptual theory was defined as processing information, assessing how a person thinks, and structuring relationships that exist among elements of information. Individuals who process greater amounts of information in a more complex fashion were seen as having higher conceptual levels.

The sample respondents (61 percent), included 43 males and 18 females. Analysis of participants' scores included a three-way ANOVA with two conceptual levels, high and low; male and female; and two occupational levels of teacher and administrator. No significant effects were identified related to the Paragraph Completion Test, participants' gender, and nine skill dimensions. Significant effects were identified for **organizational ability**, **decisiveness**, and **written communication** ($p < .05$). High conceptual females performed better than high conceptual males on **organizational ability** and **written communication**, and better than low conceptual males on **organizational ability**. High conceptual females scored significantly better than low conceptual females in **written communication**.

No significant differences were found between high and low performance on the Paragraph Completion Test for males and the 12 dimensions except for **decisiveness**. Low conceptual males scored higher than high conceptual males with all males scoring higher than high conceptual females on **decisiveness**. Holman attributed the difference to the

time for processing the information often required by the highly conceptual individuals. **Decisiveness** was defined as recognizing the need to make a decision, regardless of quality. This was Holman's rationale behind the low performance of high level conceptual individuals in **decisiveness**.

No significant effects were identified among conceptual levels and occupational levels in 11 skill dimensions. Significant differences were indicated among conceptual levels by occupations for **oral communication**. High conceptual level teachers scored significantly higher than low conceptual level teachers in **oral communication**. Low conceptual level administrators scored higher than high conceptual level administrators and higher than high and low conceptual level teachers in **oral communication**.

Jacobson determined the relationship among achieving success as a school administrator, ratings on the skill dimensions from the NASSP Assessment Center, peer and faculty ratings received in Educational Administration 850 (EDAD 850), and grades for 43 educators.⁶⁵ A panel of experts developed an instrument to measure occupational success that included human relations, job performance, goal attainment, job satisfaction, and promotion. Significant correlations with 4 of 11 success elements and three NASSP skill dimensions, **problem analysis**, **organizational ability**, and **leadership** indicated limited predictability ($p < .05$).

⁶⁵Jacobson, B. J. (1986). Predictions of success for school administrators. Unpublished doctoral dissertation, University of Nebraska.

Significant correlations were found among peer ratings from EDAD 850 and the NASSP Assessment Center results in the areas of **problem analysis, judgment, and leadership**. Significant correlations were also indicated among instructor ratings and **problem analysis, organizational ability, and leadership**. Jacobson documented moderate relationships among EDAD 850 grades and assessment scores in **problem analysis, organizational ability, and leadership**.

The NASSP assessment ratings provided some predictability for identifying successful administrators. Yet, the low number of significant relationships from the numerous possibilities would indicate assessment results should not be the only assessment tool used when selecting school administrators. From Jacobson's study the NASSP skill dimensions of **leadership, organizational ability, and problem analysis** had the highest criterion-related validity, (Table 2).

Lester studied construct validity of the NASSP Assessment Center Process using convergent and discriminant validity as indicators.⁶⁶ Lester explained that convergent validity of skill dimensions, measured by different exercises with moderate to strong correlation, would support construct validity. Lester indicated in order to have construct validity, the process

⁶⁶Lester, R. (1991). Construct validity of a National Association of Secondary School Principals' Assessment Center. Unpublished doctoral dissertation, Georgia State University.

must also demonstrate discriminant validity. If discriminant validity is not demonstrated, construct validity in NASSP Assessment Center Process is not supported. Should the ratings of exercises across dimensions show significance, with moderate to strong correlations, discriminant validity is not demonstrated. Lester investigated whether scores of 150 participants represented the exact skill dimensions they were purported to measure or if scores were measuring the assessment exercises used to elicit the skill dimensions. Lester concluded that the NASSP Assessment Center Process was in fact measuring ratings of participants' performances on exercises and not skill dimensions. Participants' ratings indicated congruence within the exercises, rather than with participants' skill levels. Lester's results did not support the construct validity of the NASSP Assessment Center Process.

Micek analyzed relationships among NASSP skill dimensions and conceptual levels of building administrators in Nebraska schools as perceived by teachers they supervise.⁶⁷ Micek concluded that conceptual levels and participants' scores were predictors for administrator effectiveness. The conceptual levels of 55 principals were obtained by analysis of the Paragraph Completion Test (PCT). PCT is a semi-projective test used to measure the ability to think in multi-conceptual terms and measures concrete

⁶⁷Micek, T. J. (1983). Teacher perceptions of performance of the NASSP assessment center behavior dimensions as it relates to the conceptual complexity of principals. Unpublished doctoral dissertation, University of Nebraska.

to abstract thinking. Ratings on the 12 skill dimensions were obtained from 455 teachers using The Administrative Behavior Dimensions Questionnaire developed for this study as data for comparison.

The analysis showed no significant differences among low and high conceptual principals and 11 skill dimensions. Significant differences were found among low and high conceptual principals and **range of interests**. High conceptual level principals had significantly higher ratings on **range of interests** than the low conceptual principals. The findings represented no support for a prediction that principals with higher degrees of conceptual levels were perceived by teachers to perform more effectively than principals with lower conceptual levels using the 12 skill dimensions.

Ogawa and Oxaal compared assessment scores of 139 participants from the Intermountain NASSP Assessment Center Project (IACP) at the University of Utah and results from the national validation study.⁶⁸ Data from the national study were used in a secondary analysis. Analysis from the IACP indicated correlations to be significant among all of the 12 skill dimensions and overall placement recommendation. Ogawa and Oxaal's analysis indicated one administrative factor accounted for 81 percent of the variance. The results from the national study indicated four factors:

⁶⁸Ogawa, R. T., & Oxaal, I. (1986, April). A reexamination of NASSP assessment center ratings: research notes. Paper presented to the Annual Meeting of the American Educational Research Association, San Francisco.

administrative analysis, presentation of self, sensitivity, and written communication.⁶⁹ Ogawa and Oxaal identified multicollinearity among the skill dimensions and questioned the discriminant validity of the process to measure 12 separate and distinct skill dimensions.

Rolland examined the relationships among participants' scores in the Minnesota Principal Assessment Center (MNPAC), participants' demographics, and promotions to principal or assistant principal.⁷⁰ Analysis of 103 participants surveyed revealed the skill dimension of leadership to be most important but not statistically significant. The conclusion indicated no significant relationships among MNPAC participants' scores and their subsequent promotions to principal or assistant principal.

Participants' scores were analyzed to determine the degree of influence of the demographic factors ($p < .05$). Age was negatively related to organizational ability, decisiveness, leadership and overall placement recommendation and positively related to range of interests. Significant differences were identified in problem analysis, organizational ability, range

⁶⁹Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. Unpublished report, Michigan State University, Department of Psychology, East Lansing, MI.

⁷⁰Rolland, L. M. Y. (1990). A study of the Minnesota Principal Assessment Center and its impact on career opportunities for its participants. Unpublished doctoral dissertation, The University of Minnesota.

of interests, and written communication related to participants' gender with females scoring significantly higher than males. Grade point average was positively related to **problem analysis, leadership, oral communication, and overall placement recommendation**. Years of experience and degrees held by participants were positively related to **organizational ability**. Marital status was significantly correlated to **oral communication** and position type to **decisiveness**. Race analysis identified significant differences in **problem analysis, judgment, organizational ability, decisiveness, oral communication, written communication, and overall placement recommendation** with white participants scoring higher than non-white participants.

Schmitt's validation study analyzed criterion-related validity for the NASSP Assessment Process.⁷¹ Areas for consideration included interrater agreement, interdimensional relationships, and data regarding possible demographic difference in skill ratings. Significant relationships were found among skill ratings and job performance ratings. Higher correlations were calculated among job performance ratings by administrators and superiors compared to teachers and support staff ratings. Student ratings of school climate were significantly related to **problem analysis, judgment, written communication, sensitivity, decisiveness, and overall placement**

⁷¹Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

recommendation ($p < .05$). Positive correlations among assessment ratings and performance ratings supported criterion-related validity of the NASSP Assessment Center Process.

Schmitt also analyzed participants' scores related to participants' demographics. Significant differences in participants' scores for gender were indicated in the skill areas of **judgment, educational values, oral communication, and written communication** with female participants receiving higher ratings than male participants ($p < .05$). Schmitt wrote:

These differences are not necessarily an indication of unfairness; if similar performance differences exist, they are simply reflective of job performance. . . Differences between subgroups (such as males vs. females) may or may not be unfair to a particular group. If similar differences are observed in performance ratings, no problem exists. That is the Assessment Center is merely identifying job relevant subgroups differences.⁷²

Position types were analyzed comparing counselors, assistant principals, elementary, middle, and high school teachers. No significant differences were indicated when comparing skill dimension scores for the three levels of

⁷²Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology, 2, 18.

teachers. Non-teaching personnel scored higher than teachers in every dimension except **sensitivity** but **personal motivation** was the only skill having a significant difference ($p < .05$). Counselors compared to non-counselors scored higher in every skill area as well as on overall placement recommendation. Assistant principals scored significantly higher when compared to non-assistant principals in **problem analysis** and **decisiveness**. Higher ratings of non-teaching groups were attributed to greater exposure and to experience with administrative decision making.

Significant differences between participants' scores and race were indicated in **problem analysis, judgment, decisiveness, leadership, written communication, organizational ability, and overall placement recommendation** ($p < .05$). White participants received higher ratings than non-white participants. Analysis of the demographic variables indicated that females scored slightly higher than males, whites scored higher than non-white participants, and non-teachers scored higher than teachers.

A second validation study was completed by Schmitt and Cohen in 1990.⁷³ The second study expanded the sample population and updated the previous information from the first three year validation study, presenting more current and comprehensive results from the NASSP

⁷³Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

Assessment Center Process. High interrater agreement was indicated. No significant interactions were found among the demographic characteristics of the assessors and the participants. Assessors viewed the dimensions as measuring somewhat different skills, but in general, participants performing high on one dimension performed well on other dimensions. Schmitt suggested more effectiveness in differentiating among skill dimension scores of participants by using assessors who are currently serving as principals and who do not know the participants.

Schmitt and Cohen also analyzed gender as a demographic variable and verified that female participants scored significantly higher than male participants on all the skill dimensions and overall placement recommendation. The largest difference in participants' scores related to gender was in **written communication and personal motivation** ($p < .05$).

Position types were analyzed comparing counselors, assistant principals, elementary, middle, and high school teachers. No significant differences were found when comparing the skill dimension scores of the three levels of teachers. Results comparing non-teaching personnel to teachers indicated non-teachers scored higher in 11 dimensions. No differences were indicated in **range of interests** ($p > .05$).

Race was analyzed as an independent variable. White participants received higher scores than non-whites in **problem analysis, judgment, organizational ability, decisiveness, leadership, sensitivity, stress tolerance,**

oral communication, written communication, and overall placement recommendations ($p < .05$).

Criterion-related validity of the NASSP Assessment Center was determined by correlating participants' skill dimension scores with supervisors, teachers, and self performance ratings. Overall, teachers rated whites higher than non-whites. Whites rated themselves lower than non-whites on performance dimensions and overall placement recommendation. The combined supervisor and teacher ratings of job performance, and not self ratings, were considered most valid indicators of participants' job performance by Schmitt and Cohen ($p < .05$).

In summary, the second study indicated non-white participants scored lower than white participants in the 12 skill dimensions and overall placement recommendation. Females performed slightly better than males and assistant principals performed better than non-assistant principals. Schmitt and Cohen indicated moderate relationships among the skill dimensions with r 's ranging from .35 to .65. Analysis indicated two factors accounted for 67 percent of the total variance. The administrative dimensions of **problem analysis, organizational ability, and judgment** accounted for 59 percent of the variance on factor one. Interpersonal, values, and personal motivation dimensions accounted for eight percent of the variance on factor two.

Tryjankowski analyzed results from the Assessment and Development Center of New Jersey using the national validation study as a model.⁷⁴

Tryjankowski analyzed interrater reliability, validated the assessment activities on the state level, and investigated the relationships among participants' demographics (gender and race) and ratings of skill levels. The 65 New Jersey Assessment Center participants included 37 male and 28 female; 28 Black, 7 Hispanic, and 30 white.

Scores indicated significant differences related to participants' gender on **judgment, decisiveness, problem analysis, and overall placement recommendation** ($p < .05$). Female participants scored significantly higher than male participants in these areas. Scores were compared related to participants' race. The only significant difference was in **written communication** with white participants scoring higher than Hispanic participants.

Tryjankowski performed factor analyses and identified two factors. Factor one was administrative and included **problem analysis, judgment, organizational ability, decisiveness, and leadership**. The second factor of interpersonal values included the remaining skill dimensions. Tryjankowski indicated a low to moderate range of correlations among skill dimensions

⁷⁴Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association.

(.23 to .64) and the possibility of a "halo effect" which "suggests that fewer rating dimensions would likely lead to similar overall results."⁷⁵

Studies Not Included in Analysis

Alvy surveyed new principals to compare difficulties of those hired for the principalship from within the school district and those hired outside the school district.⁷⁶ Alvy's study did not find evidence related to the use of the assessment center process in the identification and selection of new principals. Alvy found new principals from within the district to experience more difficulties than new principals hired from outside of the district. Results from questionnaires indicated that those hired from outside the school district had fewer difficulties in the human relations than those familiar with the district, exercised more objectivity and freedom within the new environment, and consequently, made decisions without the pressure of local ties or prejudices. Acceptance of community and school district goals was not a problem for the new school principals in either case. Working with the community to identify program needs was likewise difficult for new principals whether inside or outside the school district. Alvy's findings also

⁷⁵Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association, 26.

⁷⁶Alvy, H. B. (1983). The problem of new principals. Unpublished doctoral dissertation, University of Montana, 163.

indicated that principals were unlikely to function as a change agent in the bureaucracy of school administration.

Bley compared attitudinal survey responses from participants (61) and non-participants (59) of the assessment center from four San Diego School Districts.⁷⁷ He compared group differences between participants' and non-participants' attitudes to identify potential administrators, to diagnose strengths and weaknesses, and to plan staff development. No significant differences were noted among the groups' attitude ratings and the three variables.

Bley compared attitudes of participants and non-participants in regard to the use of assessment center results and gender. The non-participant group indicated no significant difference in attitudes between males and females. Analysis of participants' attitudes indicated significant differences in **problem analysis, decisiveness, and leadership** for males and females. High scoring female participants rated reliability higher in diagnosing strengths and weaknesses of the process. When all 12 skills dimensions were considered, however, no significant differences between male and female participants were identified.

⁷⁷Bley, J. E. (1983). Use of an assessment center for selection assessment center for selecting administrators and diagnosing inservice needs. Unpublished doctoral dissertation, Northern Arizona University.

Britt analyzed the relationship among participants' assessment scores and teachers' ratings of participants (24) using the Administrative Competency Rating Scale (ACRS).⁷⁸ Although not the NASSP skill dimensions, five of the six skills identified by Britt were the same as those in the NASSP process. The skill dimensions used by Britt included individual work characteristics, decision-making style, organizing and planning style, leadership behavior, interpersonal characteristics, and intelligence testing. Very high intercorrelations were indicated among five of the six skill areas. The intelligence test was not intercorrelated with the other five skill dimensions. The high correlation among the remaining five skill dimensions indicated multicollinearity among the dimensions. ACRS involved seven competencies: complements the goals and objectives of the school, guides instruction, allocates the staff to accomplish goals, develops and utilizes materials for goals, in-services to improve instruction, assesses needs of school, and evaluates process and products of instruction.

Comparisons among NASSP Assessment Center scores and ratings of teachers on the ACRS indicated a low relationship. Correlations did not support evidence that the NASSP Assessment Center scores and the ACRS ratings were measuring the same behaviors. Britt noted this low relationship was anticipated since the dimensions of the assessment center and the

⁷⁸Britt, W. C. (1976). Assessment center ratings of school administrators. Unpublished doctoral dissertation, Georgia State University.

ACRS were not designed to measure the same behaviors. These findings suggested that the assessment center did not effectively predict whether a principal possessed the 36 behavioral competencies of the ACRS but did predict effectiveness on the job. Participants were also rated by assistant superintendents using a high-low effective designation as a second form of evaluation. Britt, by comparing assessment center scores of the principals with ratings by their assistant superintendents, supported the conclusion that assessment center results can predict high-effective or low-effective administrators. In summary, Britt determined the criterion-related validity of the assessment center process to be high compared with superordinates' ratings. The assessment center process had low criterion-related validity when comparing evaluation results of the principals using teachers' ratings on the ACRS.

Crumpton addressed the extent to which 83 school superintendents used NASSP Assessment Center results for hiring, the importance of assessment center results relative to other factors used in principal selection, the rating of importance of assessment center results by first-time principals in South Carolina appointed since the mandate in 1984, and similarity of

superintendents' ratings of importance with that of first-time principals.⁷⁹ Crumpton indicated that superintendents and first-time principals did not consider NASSP Assessment Center results as the most important factor in the selection of first-time principals. The 160 first-time principals ranked NASSP Assessment Center results eighth from a list of 25 selection options in regard to importance while superintendents ranked NASSP Assessment Center results ninth. No differences were indicated among superintendents' and first-time principals' perceptions of the importance ranking of the assessment center results. Perception rankings were not related to the demographic variables of gender, race, size, and wealth of school district.

Draper assessed the validity of the assessment center approach for determining developmental needs of educational administrators.⁸⁰

Elementary school principals served as volunteer participants (36) in an assessment center which included 14 dimensions: information acquisition, information processing, problem analysis, information transmission, decision making, implementing decisions, supervision, organizational role

⁷⁹Crumpton, S. L. (1989). The importance of the National Association of Secondary School Principals Assessment Center Report to the appointment of public school principals in South Carolina as perceived by superintendents and first-time principals. Unpublished doctoral dissertation, University of South Carolina.

⁸⁰Draper, J. M. (1979). A comparative study of elementary school principals' on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Unpublished doctoral dissertation, Georgia State University.

performance, interactive action skills, oral communication, administration, written communication, conflict management, and conflict mediation. Six of the 14 skill dimensions are those assessed in the NASSP process.

Superordinate and subordinate ratings were compared to ratings from the assessment center skill dimensions and overall placement recommendation. The Diagnostic Survey for Leadership Improvement (DSLII) was used by the teachers to rate leadership skills of the elementary principals using the dimensions of confidence and trust, communication, control, decision making, and interaction-influence. There was not a significant relationship among the assessment center ratings and the teacher ratings on the DSLII, suggesting that the assessment center did not effectively predict whether an administrator in this sample demonstrated the on-the-job leadership behaviors measured by the DSLII. A correlation of the assessment center dimensions and the DSLII factors produced a low relationship indicating the two procedures do not measure the same behaviors.

Analysis of the DSLII 14 dimensions and the overall placement recommendation indicated high intercorrelations among the dimensions, suggesting that some of the dimensions were measuring the same behaviors. Dimension scores for information transmission (83.6 percent of variance), conflict management (9 percent of variance), and supervision (3.3

percent of variance) accounted for 95.9 percent of the variance among total assessment center scores and indicated interdimensional contamination.

Significant relationships were identified among assessment center ratings and superordinate ratings. When correlated with the 14 skill dimensions, superordinates' ratings indicated the skill of decision making to be the highest. Analysis of results of the assessment center dimension ratings and the effectiveness level of the principals correlated the highest with the dimension of decision making. The seven assessment activities included in-basket exercise, consensus seeking exercise, model building exercise, conflict management exercise, conflict mediation exercise, pilot project exercise, and background interview. The consensus seeking exercise contributed 47.1 percent of the variance to the total assessment center ratings, with model building as second (20 percent). Superordinate ratings correlated the highest with the model building activity. The model building exercise was the best single activity for classifying principals as high or low in job effectiveness. The overall finding of Draper's study indicated positive relationships among assessment center results and effective administrative behaviors reported by superordinates, but not subordinates.

Engel analyzed the content and the method of instructional programs that prepare principals.⁸¹ Engel used the rating exercises developed by NASSP Consortium as described in their 1985 monograph. The Performance-Based Preparation of Principals addressed the extent programs develop generic and specific skills required of principals and the dominant methods of instruction. The results of the research indicated the levels--familiarity, understanding, or application--to which the generic and the specific skills required of principals were emphasized in each class. Generic skills included the 12 skill dimensions in the NASSP assessment process. The specific skills were curriculum and instruction, student activities, support services, staff selection, evaluation and development, community relations, coordination with district and other schools, fiscal management, school plant maintenance, and structures communication.

Five Michigan universities in this study did not develop the generic skills or the specific skills of students on a performance basis to a high degree. The most frequent ratings for the generic skills were at the understanding level. Engel indicated only **written communication** was rated at the application level. The most common rating for the specific skills was the familiarity level.

⁸¹Engel, T. E. (1989). An analysis of the content and methods of instruction at Michigan institutions that prepare principals. Unpublished doctoral dissertation, Western Michigan University.

The most prevalent response to dominant methods of instruction was lecture-discussion. Performance-based elements were secondary and internships were not required in these five Michigan universities. An overall description from this study indicated that while instructors stated the emphasis on skills and a variety of teaching methods, the students reported the opposite to be true. "An especially strong impression which results from the study is that the critical variable is the instructor. . . Only if instructors accept this challenge will principal preparation become more pertinent to the task."⁸²

Friedman investigated the possibility of any systematic differences in assessment center ratings as a function of the race and sex of the ratees or the race of the assessors.⁸³ Non-whites scored significantly lower than whites on nine skill areas. Females scored higher than males, 4.047 compared to 3.732 respectively on a seven point scale in all of the dimensions except adaptability. Scores of females were significantly higher in sensitivity, organization and planning, oral communication, and written

⁸²Engel, T. E. (1989). An analysis of the content and methods of instruction at Michigan institutions that prepare principals. Unpublished doctoral dissertation, Western Michigan University.

⁸³Friedman, M. J. (1980). Differences in assessment center performance as a function of the race and sex of ratees and the race of assessors. Unpublished doctoral dissertation, The University of Tennessee.

communication. Females also scored higher on in-basket and task direction than males, with males scoring higher on leaderless group discussion.

Pokorny, using questionnaires sent to personnel directors in 49 states, reported techniques used in the selection of principals with a sub-question to address gender of recently selected principals.⁸⁴ Pokorny indicated that more males have been hired for administrative positions than females in smaller school districts, 65 percent compared to 35 percent respectively; whereas, the ratio was about 50 to 50 in the larger school districts. Pokorny noted that 27 districts did not respond to the ethnicity question. Of the districts responding, data were collected for American Indian, Asian, Black, Hispanic, and other. Large districts tended to hire more minorities than small districts (20 percent compared to 11.6 percent respectively). The black population represented the largest group of minorities hired equaling about five percent. Large districts hired 15 percent black, while small districts hired six percent black.

A comparison of the four types of selection techniques by district size indicated the interview was used 100 percent of the time by both large and small school districts. In-basket activities were used 25 percent of the time by small districts and 20 percent of the time by large school districts.

⁸⁴Pokorny, N. G. (1985). Techniques used in the selection of public school principals in the United States. Unpublished doctoral dissertation, University of Colorado.

Written tests were used by 12 percent of the small districts and by 8 percent of the larger districts. Assessment center techniques were used least of any of the five techniques with small districts reporting an eight percent usage and large districts only a five percent use of the technique. Assessment results were ranked third, below interviews and past experiences for both large and small school districts as the technique viewed with the highest level of confidence for hiring principals.

Pokorny's findings indicated that most school districts saw a need for training potential administrators. School districts rely heavily on interviews for selection purposes and not selection techniques that are recommended in the literature. Only one percent of the school districts listed credentials or degrees as a selection technique having the most confidence. Recruiting for vacancies was in-district with local and state as the basic advertising market, not a national effort. Only one school district of 146 listed a closer look at transcripts as a basic change in the selection procedures over the last five years. Pokorny reported little has changed in the selection process over the last five years, and that the assessment process is not being widely used for selection purposes.

Scholl's study included questionnaire data from 49 superintendents and 70 principals concerning selection criteria.⁸⁵ Superintendents' and principals' perceptions of appropriate selection criteria for hiring middle school principals, the comparison of the two perceptions, and differences in perception relating to size of school based upon student enrollment were documented.

Preferences indicated 51.2 percent of the applicants with elementary degrees, while 37.2 percent did not have a preference. A masters degree in educational administration was preferred by 89 percent compared to 97.8 percent not regarding a doctorate degree as necessary or preferred for the position. Superintendents responded that 100 percent of the applicants should have taken at least three curriculum courses and 23.9 percent recommended more than six classes in curriculum. Superintendents responded that 80 percent preferred three or more classes regarding the nature and needs of middle school students. Experience was ranked as the most important aspect of preparation by the principals as well as the superintendents.

Scholl indicated very similar responses of superintendents' and principals' perceptions and educational characteristics preferred in selection

⁸⁵Scholl, S. S. (1980). Criteria for selection of Ohio middle school principals with particular attention to the leadership role in curriculum development. Unpublished doctoral dissertation, Miami University.

criteria. Results indicated no significant differences regarding the size of the school district and responses of superintendents or principals. School districts indicated few or no formalized procedures used in the selection of school principals in Ohio with little mention of the assessment process as a viable consideration in the selection process.

Schrup focused on school administrators' perceptions after they returned to work having served as an NASSP assessor.⁸⁶ Eight Indiana school administrators were interviewed and given a post assessor training and post NASSP Assessment Center service self-rating instrument. The research examined differences related to skills application, degrees of reflective thinking, and themes identified for ongoing professional development of administrators. Assessors' ratings of perceived presence of skill application to their jobs included **organizational ability, leadership, and stress tolerance**. Significant differences were noted among post assessor training and post service as an assessor for **judgment, written communication, and range of interests**. From the interview information the greatest number of comments were related to **problem analysis and sensitivity**, perceptions of the process, and increased awareness of self. Professional development themes involved **oral and written communication**

⁸⁶Schrup, M. G. (1989). Role of reflection on administrative job performance following National Association of Secondary School Principals' assessor training and assessment center service. Unpublished doctoral dissertation, Indiana University.

skills, problem analysis, and time management. Responses indicated the training and experience of being an assessor definitely affected the assessors' self examinations of their job performance and skill application.

Storlien reported references, interviews, and administrative experience and not assessment center participation and scores as the most important selection criteria for administrative appointments.⁸⁷ Simulations, tests, academic records, and recommendation letters were the least important selection information. Political interference occurred in both rural and urban school settings. Rural school systems needed to offer more incentives to recruit principals, although they tended to be more traditional and conservative in their selection procedures.

Travetto reported how the school systems use the Maryland Assessment Center results for planning staff development and how participants use results for professional growth and development.⁸⁸ Surveys were returned by 415 participants, 80 percent of the target population. Travetto reported only a small number of school districts used NASSP Assessment Center results for staff development. Participants rated

⁸⁷Storlien, H. T. (1983). Recruitment and selection of secondary principals in Alberta: An analysis of procedures, criteria, and identified problems. Unpublished doctoral dissertation, University of Montana.

⁸⁸Travetto, S. K. (1990). An analysis of school system and participant use of assessment center for professional growth and development. Unpublished doctoral dissertation, University of Maryland.

higher were more apt to use scores for individual professional growth and development. Use of results by participants' varied with usage in individual school systems.

Van Newkirk investigated reasons for the increased growth of NASSP Assessment Center Process as a technique to identify potentially successful public school principals and assistant principals.⁸⁹ The study described the design of the NASSP Assessment Center, the methods of training and selecting assessors, and the validation of the process. Van Newkirk hypothesized that NASSP Assessment Centers had been established because of the objective nature of the process, the expediency of assessment centers in identifying talented administrators, and the predictive elements of the content. Van Newkirk indicated that the assessment center process met the guidelines for EEOC and felt results were significantly related to performance on the job. Results from the first national validation project were included to support the conclusions made in this study.

Voit replicated a Texas study by Witters-Churchill concerning evaluation of principal preparation programs in Michigan and reported similar

⁸⁹Van Newkirk, C. J. (1984). A study of the assessment center as a process for identifying prospective school principals and assistant principals. Unpublished doctoral dissertation, College of William and Mary.

findings.⁹⁰ Categories of development for the programs included not developed, slightly developed, moderately developed, and highly developed. The Michigan principals indicated a moderate development of four of the nine NASSP assessment skill dimensions within the Michigan principal preparation programs: **judgment, leadership, organizational ability, and problem analysis**. Three skills were reported to be slightly developed: **decisiveness, sensitivity, and written communication**. Two skill areas were reported as not being developed: **oral communication and stress tolerance**.

Lecture and discussion were the most frequently used instructional modes and were reported as being moderately effective. Internships were selected as the most ideal instructional mode. Principals in this study suggested that the university principal preparation programs increase the requirement of field-based experiences such as the internship, externship, and cohort opportunities. "Respondents also ask for a practical curriculum delivered by professors having credible, first-hand, educational knowledge and experience."⁹¹ Principals asked for training and development

⁹⁰Voit, L. B. (1989). Perceptive and evaluation of university principal preparation programs by Michigan Public School Principals. Unpublished doctoral dissertation, Western Michigan University.

⁹¹Voit, L. B. (1989). Perceptive and evaluation of university principal preparation programs by Michigan Public School Principals. Unpublished doctoral dissertation, Western Michigan University, abstract.

opportunities in stress management, judgment, problem solving, and leadership.

Walden documented that most NASSP Assessment Center participants plan to improve their skills if they agreed with the center's ratings.⁹² The most positive impetus for professional growth occurred when the school district used the results for staff development and designed follow-up classes. The most negative influence on professional growth occurred when a district used the results as criterion for promotion.

Wendel, Gappa, and Yusten studied the extent assessment center methodology had been incorporated into the administrative preparation programs for institutions with and without membership in the University Council for Educational Administration (UCEA).⁹³ UCEA members (32) and non-members (34) responded to the survey. Assessment centers were located in three-fourths of the states which participated in the survey. Both members and non-members of UCEA serve as headquarters for assessment centers. Two of the responding states had mandated assessment centers for certification or endorsement requirements. Three institutions surveyed

⁹²Walden, D. L. (1985). The effects of the NASSP Assessment Center process on self-improvement of assessees. Unpublished doctoral dissertation, University of Colorado.

⁹³Wendel, F. C., Gappa, L. R., & Yusten, C. R. (1988). Use of the assessment center method in administrator preparation programs. Tempe: Arizona State University, The University Council for Educational Administration.

required students in their administration preparation programs to participate in an assessment center. About 25 percent of both groups use the principles of the assessment center process as key elements in their course work, with 50 percent of their programs being redesigned to incorporate behavioral or skill dimensions. The survey indicated that 11 percent of the universities had held some type of assessment development activities. The differences reported by the UCEA member and non-member universities at the department level were negligible.

Witters-Churchill investigated Texas university administrative preparation programs using the NASSP skills of **judgment, organizational ability, stress tolerance, sensitivity, problem analysis, written communication, oral communication, decisiveness, and leadership.**⁹⁴ **Range of interests, personal motivation, and educational values** were not selected for investigation. Responses indicated that eight of the nine generic skills were moderately developed in their principal preparation programs. **Stress tolerance** was not developed. Internship was described as being the most effective mode of instruction allowing more practical and hands-on learning while lecture and discussion tended to have minimal effectiveness. The internship was selected to be the most ideal instructional mode for eight

⁹⁴Witters-Churchill, L. J. (1988). University preparation of the school administrator: Evaluations by Texas principals. Unpublished doctoral dissertation, Texas A & M University.

of the nine generic skills. Individual and team research were selected to be most effective for the development of the **written communication skill**. Other modes frequently used included group process training, internship, and games and simulations. Results indicated the instructional modes to be moderately effective. Major recommendations for university administrator preparation programs were to improve and/or increase instruction of generic and specific job-related skills, to improve and extend opportunities for experience in the field, and to provide practice-oriented instructors at the universities.

The 34 studies described in this chapter are incorporated in Chart 1. This chart includes a brief description of author, type of study, date, sample size, construct of study, and outcomes of the studies.

Chart 1

Description of Studies

| Author(s) Type of Study Date, Sample Size | Construct of Study | Outcome(s) |
|---|---|---|
| Alvy Dissertation 1983, 70 | Problems of new principals: comparisons of those hired from within and outside of school district | Out of district have better human relations; principals are not change agents |
| Blanck Dissertation 1989, 394 | Comparison of NASSP results for five role groups in Michigan | No significant differences for teacher groups; assistant principals and quasi-administrators score higher than teachers |
| Bley Dissertation 1983, 120 | Attitude ratings comparing NASSP participants and non/participants | No significant difference in attitude ratings |
| Breed Dissertation 1985, 19 | Comparison of NASSP results to Administrator Perceiver Interview | Comparison did not support concurrent validity |
| Britt Dissertation 1976, 24 | Comparison of NASSP results to Administrative Competency Rating Scale | Low correlation did not support criterion-related validity |
| Crumpton Dissertation 1987, 243 | Comparison of superintendents and principals use of NASSP results for selection purposes | NASSP results ranked eighth and ninth from a list of 25 for selection preferences |
| Dickson Dissertation 1987, 201 | Descriptive use of NASSP results for career and professional development | No differences noted for gender or size of school district |

| Author(s) Type of Study Date, Sample Size | Construct of Study | Outcome(s) |
|---|---|---|
| Douglas and Johnson Report 1984, 144 | Comparison of NASSP results by demographic variables | No significant effects based upon demographic variables |
| Draper Dissertation 1979, 36 | Comparison of NASSP results to Diagnostic Survey for Leadership | Low correlation did not support criterion-related validity |
| Ehinger Dissertation 1986, 183 | Compared NASSP results to superintendents' ratings | Significant correlations identified for criterion-related validity |
| Engel Dissertation 1989, 5 | Description NASSP skills in principal preparation programs | Skills not developed to a high degree in university programs |
| Farmer Dissertation 1985, 125 | Description of participation effects in NASSP assessment | Progress to be fair; disagreed on use of results for selection purposes |
| Ford Dissertation 1987, 249 | Compared attitude ratings to NASSP results | Higher NASSP results yielded higher attitude ratings |
| Friedman Dissertation 1980, 256 | Compared assessment results based upon gender and race | Females scored higher than males; whites scored higher than non-whites |
| Heater Dissertation 1989, 152 | Compared NASSP results to Purdue Teacher Questionnaire | No significant correlation to support criterion-related validity |
| Holman Dissertation 1986, 61 | Compared NASSP results to Paragraph Completion Test | Differences related to gender but not to conceptual levels or occupational levels |

| Author(s) Type of Study Date, Sample Size | Construct of Study | Outcome(s) |
|---|---|---|
| Jacobson Dissertation 1986, 43 | Compared NASSP results to evaluation of educational administration class | High criterion-related validity with leadership, organizational ability, and problem analysis |
| Lester Dissertation 1990, 150 | Compared NASSP results to determine construct validity | Construct validity was not supported |
| Micek Dissertation 1983, 455 | Compared NASSP results to Paragraph Completion Test and teacher evaluations | Criterion-related validity was not supported |
| Ogawa\Oxaal Report 1986, 139 | Compared NASSP results to national validation study | Multicollinearity identified |
| Pokorny Dissertation 1985, 141 | Description of selection techniques for principals | Little change reported in techniques for selection |
| Rolland Dissertation 1991, 108 | Compared NASSP results and promotion of participants | Some degree of a relationship with skill dimensions |
| National Validity Schmitt et al. 1982, 39/350 | Compared NASSP results to determine validity of process | Content validity was high, discriminant validity low, criterion-related validity low |
| National Validity Schmitt and Cohen 1990, 2699/2910 | Compared NASSP results to determine validity of process | Content validity high, low discriminant validity, criterion-related validity moderate |
| Scholl Dissertation 1980, 119 | Comparison of selection criteria by principals and superintendents | Formalized selection procedures not apparent, some agreement on criteria between two groups |

| Author(s) Type of Study Date, Sample Size | Construct of Study | Outcome(s) |
|---|--|--|
| Schrup Dissertation 1989, 16 | Description of effects of serving as NASSP assessor | Training and experience affect job performance |
| Storlein Dissertation 1983, 372 | Description of principal selection process | Interview, references, and administrative experience most important |
| Travetto Dissertation 1990, 415 | Description of NASSP results for training and development | Limited use of results indicated |
| Tryjankowski Report 1990, 65 | Compared NASSP results to validate N. J. center | High content validity, low discriminant validity |
| Van Newkirk Dissertation 1984 | Historical description of NASSP Process | Process addressed EEOC requirements and improved selection process |
| Voit Dissertation 1989, 166 | Description of NASSP skills and principal preparation programs by principals | Changes needed to address generic skills and improve instructional modes |
| Walden Dissertation 1985, 12 | Description of NASSP results for self development | Positive if used by school district for development; negative if for promotion |
| Wendel, et al. Report 1988, 225 | Description of NASSP assessment influence in university programs | Negligible differences in universities, members or non-members of UCEA |
| Witters-Churchill Dissertation 1988, 325 | Description of NASSP skills and principal preparation programs by principals | Changes needed to address generic skills and improve instructional mode by internships |

CHAPTER 3

Methodology

Introduction

Chapter 3 is an explanation of the meta-analytic method used for the synthesis and analysis of the research findings related to the NASSP Assessment Center Process. Studies were analyzed to determine criterion-related validity of the NASSP Assessment Process and possible bias in the assessment process related to participants' gender, position type, and race.

Selection of Studies

A literature review utilizing a computer search of the Educational Resources Information Center (ERIC) was employed using key words: *NASSP Assessment Center and principals, school administrators, and administrators*. A search of Dissertation Abstracts, International, NASSP literature, and professional references was conducted to collect studies related to principal assessment. The literature search identified 34 studies, consisting of 29 dissertations, three research reports, and two validity studies. Various facets of principal assessment were researched in 31 studies, two focused on principal selection, and one focused on the problems facing new principals.

Meta-analysis is used to determine the effectiveness of the assessment process to identify participants' potential for success as prospective school principals. No significant relationships were identified between participants' ratings on different performance criteria and participants' NASSP skill dimension scores. Criteria included results from attitude ratings, job performance rating using the Administrator Perceiver Interview, Administrative Competency Ratings, Diagnostic Survey for Leadership, Purdue Teacher Questionnaire, performance in educational administration classes, promotions, and conceptual levels measured by the Paragraph Completion Test. Chapter 2 contains a comprehensive description of each of these studies and their findings.

Methodology

Hyde,¹ Hunter,² and Hedges³ describe meta-analysis as the quantitative accumulation and analysis of descriptive statistics across studies; the application of quantitative methods to the problem of combining

¹Hyde, J., & Linn, M. (1986). The Psychology of Gender, Advances through Meta-analysis. Baltimore: John Hopkins University Press, 3.

²Hunter, J., Schmidt, F., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills: Sage Publications.

³Hedges, L., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press, 2.

evidence from different studies. The number of studies and the different types of criteria used to determine validity of the assessment process supported the selection of meta-analysis to integrate and analyze the NASSP Assessment Center research.

This analysis averaged the effect sizes from studies which had compared participants' NASSP scores and participants' performance on multiple types of criteria. The NASSP scores were measures of the 12 skill dimensions and behavioral responses of the participants determined during the assessment process.⁴ Relationships between participants' NASSP scores and scores on criterion measures are expressed as correlations. Correlations for each of the 12 skill dimensions and the overall placement recommendation indicate the criterion-related validity of the process. The meta-analysis also examined possible bias within the NASSP Assessment Center Process related to gender, position type, and race.

⁴Rosenthal, R., & Rosnow, R. (1975). Primer of methods for the behavioral sciences. New York: John Wiley & Sons.

Meta-Analysis

Rosenthal⁵ and Wolf⁶ developed a method to convert t , F , and r statistics to a standardized Z statistic for the purpose of cross-study comparisons and cross-study averaging. The t test is used to compare two groups to indicate if the differences between group means are large enough to assume that the corresponding population means are different. If the two sample means are far enough apart, the t test will yield a statistically significant difference. This significant difference permits the researcher to conclude that the two populations probably do not have the same mean and suggests the existence of two populations of outcome scores representing the dependent variable.

While t test is designed to compare two groups, the one-way ANOVA and the resulting F statistic is used to compare two or more groups. To determine the correlation of two variables, the Pearson Product-Moment Correlation r is an appropriate parametric technique. The conversion of the t and F statistics to an equivalent Z statistic allows comparisons across studies. Measures used to indicate a relationship between assessment

⁵Rosenthal, R. (1989). Meta-analytic procedures for social research. Applied Social Research Method Series. Newbury Park, CA: Sage Publications.

⁶Wolf, F. M. (1987). Meta-analysis, quantitative methods for research synthesis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-059. Beverly Hills: Sage Publication.

scores and criteria ratings were validity correlation measures of r which range from -1.00 to $+1.00$, with zero indicating no relationship. The higher the value of r , the greater the validity of the results. An r of $.68$ predicts a higher validity than an r of $.23$. The positive or negative r , i.e., $.68$ as compared to $-.68$ has equal magnitude; the negative or positive sign denoting the direction of the relationship between the variables.

Findings from the selected studies were reported as either t , F , or r statistics. The formulae in Table 1 convert t and F statistics to an r . Standardized r or Z_r statistics are then determined to describe combined estimates of the effect sizes. $\sqrt{}$ is the symbol for square root and df is the degree of freedom derived from the sample size.

The studies reported findings with multiple criterion measures. The concept of effect size in this meta-analysis was utilized to describe the relationship between NASSP assessment scores and selected criteria. The results of the findings were transformed to the scale of r . These r scores were converted to standardized effect sizes, Z_r scores, to adjust for size of sample and to normalize distribution, allowing studies to have equal weight, while adjusting for sample sizes. Both weighted and unweighted Z_r scores were computed to identify the influence of sample size upon the mean effect size for the three demographic variables. The Z_r scores permit the combining of effects into average effect sizes across studies. Combined Z_r scores were converted back to correlation coefficients and coefficients of determination

TABLE 1
Meta-Analytic Formulae Used for Effect Size Estimation⁷

| <u>Original Statistic</u> | <u>Conversion to "r"</u> |
|---------------------------|--|
| <i>t</i> | $\text{SQRT}((t^2)/(t^2 + df))$ |
| F | $\text{SQRT}((F)/(F + df \text{ error}))$ where F(1,df error) |
| Pearson r | r |
| <u>Conversion to Z:</u> | $1/2\text{LOG}(\text{base } e)[(1 + r)/(1 - r)]$ |

⁷Rosenthal, R. (1987). Judgment studies: Design, analysis, and meta-analysis. New York: Cambridge University Press.

for interpreting and reporting the relationships between skill dimensions and criteria. There was concern over the presence of two large national studies: Criterion-Related Validity of the NASSP Assessment Center (1982 and 1990) which could have positively skewed the results.⁸ Mean effect sizes were computed with and without the data from these studies, but no substantive differences resulted. Final results were presented with these studies included.

Each study of the NASSP Assessment Center Process singularly presents a small picture of the process. The meta-analysis presents combined evidence of the criterion-related validity of the NASSP Assessment Center Process. Combined results of gender, position type, and race of the participants were also analyzed to determine possible bias within the process.

⁸Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

Schmitt N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

CHAPTER 4

Results

Chapter 4 contains the results of the meta-analysis of the validity related to the NASSP Assessment Center Process. Criterion-related validity represents the relationships among participants' scores in the assessment center and participants' performances on selected criteria. Each skill is characterized as an independent skill dimension when establishing criterion-related validity. Discriminant validity is analyzed in the studies to indicate independence of each of the 12 skill dimensions. This analysis determines the distinguishing ability of the NASSP process to assess 12 separate and distinct skill dimensions. Reliability of the assessment process is a prerequisite to criterion-related validity. Interrater agreement is analyzed to establish support for reliability of the assessment process. Chapter 4 also reports results related to possible bias in the assessment process and participants' gender, position type, and race.

Results are recorded in three sections: interdependency of the 12 skill dimensions indicated by discriminant validity and interrater agreement, results of the criterion-related validity of the NASSP Assessment Process, and results from the analysis of the participants' demographic variables related to bias in the NASSP Assessment results. Effect sizes are calculated across the studies for the demographic variables of gender, position type,

and race. Brief summaries of the studies included in each analysis are presented. More information for each study can be found in Chapter 2.

Section One

Skill Dimensions

Studies reported high intercorrelations among the 12 NASSP skill dimensions. The structure of the NASSP process regarding the discriminating potential to measure 12 separate and distinct skills was questioned. Douglas and Johnson's analysis of the 12 skill dimensions found the highest intercorrelations among **problem analysis, judgment, organizational ability, and leadership** ($p < .05$).¹ Correlation coefficients were not significant for **range of interests, educational values, and stress tolerance with decisiveness** indicating lower intercorrelations among these dimensions ($p < .05$).

Draper's analysis of intercorrelations among the assessment dimensions ranged from .24 to .89, indicating that the dimensions may be

¹Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300 879)

measuring the same underlying constructs.² Although not the exact NASSP skill dimensions, Draper's high correlations among the skill areas supported a common difficulty in assessment processes, which is the discrimination among skill measures.

Ogawa and Oxaal analyzed data from the Intermountain-NASSP Assessment Center Project (IACP) at the University of Utah.³ Ogawa and Oxaal's factor analysis indicated the existence of one factor that accounted for 81 percent of the variance in participants' skill scores. Ogawa and Oxaal's one factor indicated multicollinearity among the skill dimensions and the lack of independence for the skill areas.

Schmitt analyzed interdimensional relationships.⁴ Correlations among individual skill dimensions of **judgment, problem analysis, organizational ability, and leadership** ranged from .35 to .65. The lowest correlations were among the dimensions of **educational values, personal motivation, and range of interests**. Schmitt concluded that assessors viewed the 12 skill

²Draper, J. M. (1979). A comparative study of elementary school principals' on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Unpublished doctoral dissertation, Georgia State University.

³Ogawa, R. T., & Oxaal, I. (1986, April). A reexamination of NASSP assessment center ratings: research notes. Paper presented to the Annual Meeting of the American Educational Research Association, San Francisco.

⁴Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

dimensions as measuring somewhat different skills, but in general, participants performing high on one dimension performed well on other dimensions.

Schmitt and Cohen's analysis indicated four factors.⁵ Factor one included **problem analysis, judgment, and organizational ability** and accounted for 59 percent of the variance in participants' scores. The remaining three factors were interpersonal, values, and personal motivation. Schmitt wrote:

Practically, these results indicate that raters are not distinguishing reliably between candidates' skill levels on all 12 dimensions. . . [F]ewer rating dimensions would likely lead to similar overall results.⁶

Tryjankowski performed factor analyses and identified two factors. Factor one was administrative and included **problem analysis, judgment, organizational ability, decisiveness, and leadership**. The second factor of interpersonal values included the remaining skill dimensions. Tryjankowski

⁵Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

⁶Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology, 12.

indicated a low to moderate range of correlations among the 12 skill dimensions with r 's ranging from .23 to .64.

Britt examined the relationship among six dimensions of an assessment process that included individual work characteristics, decision-making style, organizing and planning style, leadership behavior, interpersonal characteristics, and intelligence.⁷ Britt reported high intercorrelations among five of the six dimensions ranging from an r of .88 to .96. Although the six assessment dimensions were not the specific NASSP skill dimensions, five included similar activities. The intelligence test was not correlated with the other five dimensions. High correlations among the remaining five dimensions indicated multicollinearity.

Schmitt's national validity summary of the skill dimensions indicated high correlations among skill dimensions measured in the NASSP Assessment Center Process. Schmitt noted the reasonableness of these results due to the logical relationship among the dimensions of **problem analysis, judgment, organizational ability, decisiveness, and leadership.** Ogawa and Oxaal indicated multicollinearity among all of the dimensions.

Results indicated a lack of discriminant validity among the skill dimensions. Schmitt et al., Schmitt and Cohen, Draper, Tryjankowski, Britt,

⁷Britt, W. C. (1976). A comparative study of on-the-job ratings to assessment center ratings of school administrators. Unpublished doctoral dissertation, Georgia State University.

Ogawa and Oxaal used factor analysis to identify a range of one to four factors. **Problem analysis, organizational ability, and leadership** were the skills more frequently cited with the highest intercorrelations.

Reliability

Skill Dimensions and Overall Placement Recommendation

The 12 skill dimensions measured by the NASSP Assessment Process have high correlations with the overall placement recommendation. Douglas and Johnson examined the relationship among the 12 skill dimensions and the overall placement recommendation for 96 participants.⁸ The skills having the highest correlations with the overall placement recommendation were **judgment (.91), problem analysis (.91), organizational ability (.93), leadership (.95)**. Ogawa and Oxaal's comparisons indicated correlations to be significant between overall recommendation ($p < .05$).⁹ Schmitt also analyzed relationships between the skill dimensions and overall placement

⁸Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300 879)

⁹Ogawa, R. T., & Oxaal, I. (1986, April). A reexamination of NASSP assessment center ratings: research notes. Paper presented to the Annual Meeting of the American Educational Research Association, San Francisco.

recommendation and determined correlations to range from .51 to .80.¹⁰ Draper's analysis of the relationships between overall assessment rating and five areas similar to the NASSP skill dimensions of **problem analysis, decisiveness, oral communication, written communication** also indicated high correlations ranging from .66 to .89.¹¹ Schmitt surmised the moderate to high correlations between each of the 12 skill dimensions and overall placement recommendation by the assessors indicated internal consistency of the measure and supported reliability of the assessment process.

Interrater Reliability

Multiple activities, multiple observers, and multiple skills allow some measure of reliability to be determined.¹² Studies analyzed interrater reliability of the NASSP Assessment Center Process. Douglas and Johnson indicated significant differences among assessors' scores for **personal**

¹⁰Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

¹¹Draper, J. M. (1979). A comparative study of elementary school principals' on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Unpublished doctoral dissertation, Georgia State University.

¹²Campbell, D. & Fiske, D. (1959, March). Convergent and discriminate validation by the multitrait-multimethod matrix. Psychological Bulletin, **56**, 81-105.

motivation ($r = .22$) and educational values ($r = .25$). No significant differences were indicated for the remaining ten dimensions ($p < .05$).¹³

Schmitt et al. analyzed participants' assessment scores to determine interrater reliability.¹⁴ Schmitt noted high interrater agreement on the 12 skill dimensions ($r > .60$) and overall placement recommendation ($r > .80$). Schmitt and Cohen examined interrater reliability for 2910 participants.¹⁵ The ratings of different assessors were highly correlated on all 12 skill dimensions ($r > .65$) and the overall placement recommendation ($r > .80$). Tryjankowski analyzed interrater reliability for 65 NASSP participants.¹⁶ Tryjankowski concluded ratings of participants by different assessors indicated moderate to high correlations ($r = .35$ to $.83$) between each of the 12 skill dimensions and overall placement recommendation. Research studies in this analysis concluded that the interrater agreement was within a

¹³Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300 879)

¹⁴Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

¹⁵Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

¹⁶Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association.

moderate to high correlation range (Table 2). Moderate to high interrater agreement is indicative of internal consistency and supports reliability and validity of the NASSP Assessment Process.

Section Two

Criterion-Related Validity

Section two consists of the research findings pertaining to the criterion-related validity of the NASSP Assessment Center Process. This analysis incorporated criteria including: Administrator Perceiver Interview, Administrative Competency Rating, Diagnostic Survey for Leadership, Purdue Teacher Questionnaire, attitude ratings, conceptual levels using the Paragraph Completion Test, and various job performance ratings to substantiate criterion-related validity of the assessment process.

Breed compared the results from the Administrator Perceiver Interview (API) to participants' NASSP Assessment scores to determine validity between two procedures used in the selection of school administrators.¹⁷ No significant correlations existed among scores on the API themes and scores from the NASSP skill dimensions. The study indicated scores on the API provided different information than the information from the NASSP

¹⁷Breed, R. D. (1985). Concurrent validity of two administrator selection procedures. Unpublished doctoral dissertation, University of Nebraska.

TABLE 2**Interrater Reliability of the Assessment Center Skill Ratings**

| N = 4 | Douglas/ Johnson | Schmitt, et al. | Schmitt & Cohen | Tryjankowski |
|--------------------------|-----------------------------|----------------------------|--------------------------------|---------------------|
| Problem Analysis | .48 - .84 | .67 - .77 | .71 - .74 | .46 - .69 |
| Judgment | .62 - .78 | .65 - .74 | .70 - .72 | .66 - .83 |
| Organizational Ability | .62 - .89 | .68 - .74 | .75 - .79 | .55 - .72 |
| Decisiveness | .27 - .83 | --- | .77 - .79 | .66 - .83 |
| Leadership | .65 - .88 | .76 - .81 | .73 - .80 | .67 - .83 |
| Sensitivity | .43 - .81 | .60 - .70 | .67 - .70 | .41 - .66 |
| Range of Interests | .34 - .89 | --- | .72 - .76 | .60 - .79 |
| Personal Motivation | .22 - .99 | .62 - .68 | .70 - .75 | .64 - .81 |
| Educational Values | .25 - .99 | .57 - .66 | .67 - .69 | .49 - .76 |
| Stress Tolerance | .50 - .99 | .67 - .73 | .67 - .69 | .49 - .68 |
| Oral Communication | .53 - .95 | .59 - .69 | .68 - .71 | .57 - .81 |
| Written Communication | .46 - .92 | .67 - .76 | .77 - .79 | .64 - .79 |
| Overall Rating | .53 - .97 | .80 - .85 | .82 - .84 | .57 - .73 |

assessment. The findings by Breed did not support criterion-related validity among the API themes and the NASSP assessment.

Britt analyzed the relationship among principals' assessment center ratings and principals' performance ratings by teachers and assistant superintendents.¹⁸ School district assistant superintendents rated principal performance using a high-low effectiveness designation as the criterion measure. Although not the specific NASSP skill dimensions, this assessment included five of the NASSP areas: individual work characteristics, decision-making style, leadership behavior, interpersonal characteristics, organizing, and planning style. Britt's comparison of principals' assessment scores and effectiveness ratings by superordinates supported assessment center scores to accurately predict high-effective and low-effective principals ($p < .05$).

Teachers rated principal performance using the Administrative Competency Rating Scale (ACRS). Comparisons between assessment scores and ratings of teachers on the ACRS did not indicate a relationship. Low correlations did not support assessment ratings and ACRS ratings to be measuring the same behaviors ($r = .21$). The assessment process was determined to have low criterion-related validity when comparing principals' assessment scores and teachers' ratings on the ACRS.

¹⁸Britt, W. C. (1976). Assessment center ratings of school administrators. Unpublished doctoral dissertation, Georgia State University.

Draper compared superordinate and subordinate ratings with assessment scores and overall placement recommendation.¹⁹ Significant correlations were identified between overall assessment ratings and superordinates' performance ratings (r ranging from .66 to .91). Teachers evaluated principals' performance using the Diagnostic Survey for Leadership Improvement (DSLII) with 14 skill dimensions and identified no significant relationships between participants' assessment scores and teachers' ratings. Correlation of the assessment scores and the DSLI produced low coefficients (-.32 to .20) indicating the two procedures did not measure the same behaviors ($p < .05$).

Ehinger compared performance ratings by superintendents for participants and participants' scores from the assessment center.²⁰ Significant correlations were found between predictions of superintendents' ratings of prospective principals and assessment scores. Ehinger indicated by using assessment center scores, superintendents could predict job performance of principals.

¹⁹Draper, J. M. (1979). A comparative study of elementary school principals' on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Unpublished doctoral dissertation, Georgia State University.

²⁰Ehinger, J. M. (1986). The validation of selected assessment center activities for incumbent school administrators who participated in the University of Tulsa Education Management Development Program. Unpublished doctoral dissertation, University of Tulsa.

Farmer compared principals' perceptions of participants to participants' ratings from the assessment center.²¹ Data indicated that the principals' perceptions of participants were significantly higher than the assessment scores on six of the 12 skill dimensions: **problem analysis, judgment, organizational ability, written communication, range of interest, and educational values.** Farmer indicated that given an extended period of time to observe assistant principals and teachers, principals could present ratings similar to those of the assessment center, which supports the criterion-related validity of the process.

Ford compared attitude ratings toward the NASSP Assessment Center and participants' scores across the 12 skill dimensions and the four assessment exercises.²² Significant relationships were identified among participants' attitude ratings toward the NASSP Assessment Process and **problem analysis, judgment, leadership, written communication, range of interests and overall placement recommendation.** The analysis indicated the higher the **overall placement recommendation** and five skill scores, the more positive the participants' attitude ratings. Conversely, the lower the scores

²¹Farmer, T. N. (1987). An analysis of the effects of participation in the National Association of Secondary School Principals' Assessment Center in Howard County, Maryland. Unpublished doctoral dissertation, University of Maryland.

²²Ford, M. J. (1987). An attitude survey of NASSP principal assessment center candidates. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

on overall placement recommendation and five skill dimensions, the lower the attitude ratings.

Heater compared the NASSP assessment scores and the Purdue Teacher Opinionnaire (PTO) to determine criterion-related validity.²³ The study assumed that some factors of teacher morale were influenced by school administrators' behaviors. Heater found no significant correlations to support criterion validity. Heater concluded that teacher morale was not an appropriate criterion measure to validate the NASSP Assessment Center Process.

Micek compared assessment scores of 55 administrators in Nebraska to conceptual level ratings by teachers.²⁴ The conceptual levels of the principals were obtained from ratings on the Paragraph Completion Test (PCT). Analysis indicated no significant differences existed among low and high conceptual principals and 11 of the 12 NASSP skill dimensions. High conceptual level principals had significantly higher scores on range of interests. Micek's findings indicated no support that principals with higher

²³Heater, J. D. (1989). Teacher morale as a criterion of the predictive validity of the National Association of Secondary School Principals' assessment center. Unpublished doctoral dissertation, University of Nebraska.

²⁴Micek, T. J. (1983). Teacher perceptions of performance of the NASSP assessment center behavior dimensions as it relates to the conceptual complexity of principals. Unpublished doctoral dissertation, University of Nebraska.

degrees of conceptual levels were perceived by teachers to perform more effectively than principals with lower conceptual levels using the NASSP skill dimensions as indicators.

Holman compared 100 participants' conceptual levels using the Paragraph Completion Test and participant's scores in the NASSP Nebraska Assessment Center.²⁵ Holman researched the differences in performance on the 12 skill dimensions related to high and low conceptual levels. No significant differences were found between high and low performance on the Paragraph Completion Test and assessment scores.

Jacobson compared the 12 NASSP skill dimension scores to ratings of participants by peers in an educational administration class, instructor ratings, class grades, and a scale for success to determine criterion validity.²⁶ Jacobson found significant relationships between the peer ratings from EDAD 850 class and assessment scores in **problem analysis, judgment, and leadership** ($p < .05$). Significant relationships were also found between instructor ratings and the skills of **problem analysis, organizational ability, and leadership**. Moderate relationships (.35 to .60) were found among EDAD 850 grades and assessment scores in **problem analysis,**

²⁵Holman, D. M. (1986). The relationship of participant performance in the Nebraska Assessment Center to participant conceptual level/cognitive complexity. Unpublished doctoral dissertation, University of Nebraska.

²⁶Jacobson, B. J. (1986). Predictions of success for school administrators. Unpublished doctoral dissertation, University of Nebraska.

organizational ability, and leadership. Significant relationships were found between the scale for success and assessment scores for **problem analysis, organizational ability, and leadership.** The limited number of significant correlations found in four of eleven variables indicated limited predictive validity of the NASSP process to identify successful administrators. From Jacobson's study the NASSP skill dimensions of **leadership, organizational ability, and problem analysis** appeared to have the highest predictive ability in selecting an effective school administrator.

Rolland examined the relationships among participants' scores from the Minnesota Principal Assessment Center (MNPAC), demographic characteristics of participants, and promotion to principal or assistant principal.²⁷ Analysis of the NASSP scores of the 103 participants indicating a promotion status revealed the skill dimension of **leadership** to be most important but not statistically significant. Conclusions indicated no relationship between MNPAC participants' scores and subsequent promotion to principal or assistant principal.

²⁷Rolland, L. M. Y. (1990). A study of the Minnesota Principal Assessment Center and its impact on career opportunities for its participants. Unpublished doctoral dissertation, The University of Minnesota.

Schmitt's validation study determined a significant relationship between the NASSP Assessment skill ratings and job performance ratings.²⁸ Higher correlations were reported among job performance ratings of superiors (.45 to .72) compared to performance ratings by teachers (.49 to .79) and support staff (.38 to .75). Student ratings of school climate were significantly related to **problem analysis, judgment, written communication, sensitivity, decisiveness, and overall performance recommendation**. High correlations between assessment scores and job performance criteria supported the criterion validity of the NASSP Process in predicting potential for success of prospective school administrators.

Schmitt and Cohen correlated performance ratings by supervisors, teachers, and participants with NASSP scores and overall placement recommendation.²⁹ The supervisor and teacher job performance ratings, and not self ratings, had the highest correlations when compared to the participants' assessment center scores.

A summary of the criterion-related validity data included a combining of effect sizes to determine validity of the NASSP Assessment Center

²⁸Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

²⁹Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

TABLE 3
Critere Effect Sizes (Z) Indicated by Individual Studies
Twelve Skill Dimensions and Overall Recommendation Rating

| Study (K = 30) | PA | J | OA | D | L | SE | R | PM | EV | ST | OC | WC | Overall |
|--|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|--------|---------|
| Administrative Competency Rating Survey Performance ¹ | 0.590 | 0.929 | 0.576 | 0.618 | 0.887 | 0.908 | 0.618 | 0.662 | 0.793 | 0.288 | 1.070 | 1.090 | |
| Job Performance ² | 1.422 | | | 1.376 | | | | | | | 0.950 | 1.701 | |
| Job Performance ³ | | | | | | | | | | | | | 0.248 |
| Conceptual Performance ⁴ | 1.059 | 0.201 | 0.184 | 0.081 | 0.209 | 0.070 | 0.226 | 0.106 | 0.153 | 0.173 | 0.073 | 0.076 | |
| Goal Achievement ⁵ | -0.056 | -0.160 | -0.095 | -0.033 | -0.214 | | | -0.181 | | | | -0.294 | |
| Class Performance ⁵ | -0.476 | -0.132 | -0.217 | -0.071 | -0.190 | | | -0.327 | | | | -0.256 | |
| Promotion ⁵ | 0.480 | 0.367 | 0.485 | 0.281 | 0.530 | -0.010 | -0.060 | 0.424 | | | | 0.361 | |
| Peer Relationship ⁵ | -0.405 | -0.201 | -0.120 | -0.087 | -0.247 | 0.030 | -0.010 | -0.413 | | | | -0.240 | |
| Job Satisfaction ⁵ | -0.051 | -0.034 | -0.076 | 0.057 | -0.025 | | | 0.015 | | | | -0.014 | |
| High Conceptual Performance ⁶ | 0.084 | 0.134 | 0.203 | 0.109 | 0.132 | 0.092 | 0.222 | 0.232 | 0.126 | 0.152 | 0.053 | 0.163 | |
| Low Conceptual Performance ⁶ | 0.128 | 0.359 | 0.290 | 0.134 | 0.431 | 0.078 | 0.625 | 0.455 | 0.347 | 0.443 | 0.271 | 0.443 | |
| Application of Skills ⁷ | 0.330 | 0.710 | 0.414 | 0.325 | 0.349 | 0.661 | 0.476 | 0.406 | 0.475 | 0.492 | 0.309 | 0.676 | |
| Reflection ⁷ | 0.727 | 0.281 | | | | 0.407 | | 0.448 | | | | | |
| Development of Goals ⁷ | 0.294 | 0.617 | 0.692 | | 0.612 | 0.302 | 0.619 | 0.516 | 0.403 | | | 0.516 | |
| Community Relations ⁸ | 0.266 | 0.266 | 0.343 | 0.203 | 0.354 | 0.090 | -0.040 | 0.070 | -0.040 | 0.182 | 0.245 | 0.172 | 0.299 |

(table continues)

Table 3 (continued)

| Study (K = 30) | PA | J | OA | D | L | SE | R | PM | EV | ST | OC | WC | Overall |
|--|-------|--------|-------|-------|-------|--------|--------|--------|--------|-------|-------|--------|---------|
| Coordination ^a | 0.224 | 0.182 | 0.266 | 0.172 | 0.299 | -0.050 | -0.080 | -0.050 | -0.080 | 0.050 | 0.234 | 0.090 | 0.321 |
| Curriculum Program ^a | 0.131 | 0.050 | 0.321 | 0.224 | 0.234 | 0.030 | -0.010 | 0.172 | 0.020 | 0.172 | 0.141 | 0.203 | 0.310 |
| Curricular Objectives ^a | 0.224 | 0.141 | 0.354 | 0.213 | 0.321 | -0.010 | -0.060 | 0.070 | 0.030 | 0.131 | 0.182 | 0.213 | 0.321 |
| Development ^a | 0.110 | 0.010 | 0.100 | 0.010 | 0.234 | -0.100 | 0.100 | 0.060 | -0.010 | 0.060 | 0.121 | 0.131 | 0.172 |
| Goal Direction ^a | 0.060 | -0.020 | 0.234 | 0.110 | 0.192 | -0.030 | -0.020 | 0.010 | -0.010 | 0.100 | 0.141 | 0.131 | 0.182 |
| Evaluation ^a | 0.131 | 0.010 | 0.224 | 0.070 | 0.224 | -0.100 | 0.020 | 0.090 | 0.020 | 0.141 | 0.224 | 0.030 | 0.266 |
| Fiscal Responsibility ^a | 0.365 | 0.245 | 0.161 | 0.182 | 0.266 | 0.010 | -0.060 | -0.020 | -0.030 | 0.080 | 0.151 | 0.040 | 0.310 |
| Interpersonal Skills ^a | 0.121 | 0.050 | 0.245 | 0.131 | 0.245 | 0.050 | -0.070 | 0.040 | 0.020 | 0.040 | 0.110 | -0.010 | 0.141 |
| Maintenance ^a | 0.192 | 0.192 | 0.321 | 0.224 | 0.310 | 0.121 | -0.165 | -0.050 | -0.070 | 0.100 | 0.245 | 0.090 | 0.299 |
| Parents ^a | 0.358 | 0.172 | 0.400 | 0.192 | 0.321 | 0.030 | 0.060 | -0.090 | 0.040 | 0.234 | 0.234 | 0.121 | 0.245 |
| Participation ^a | 0.080 | 0.010 | 0.182 | 0.040 | 0.224 | -0.040 | -0.080 | 0.010 | 0.020 | 0.030 | 0.121 | 0.020 | 0.213 |
| Structured Communications ^a | 0.192 | 0.100 | 0.266 | 0.131 | 0.266 | -0.060 | -0.070 | 0.020 | -0.172 | 0.100 | 0.213 | 0.030 | 0.182 |
| Supervision ^a | 0.010 | -0.010 | 0.141 | 0.010 | 0.234 | -0.110 | 0.040 | -0.010 | -0.070 | 0.020 | 0.020 | -0.080 | 0.110 |
| Support Services ^a | 0.245 | 0.234 | 0.277 | 0.182 | 0.299 | 0.131 | -0.060 | 0.100 | 0.161 | 0.151 | 0.224 | 0.172 | 0.354 |
| Overall Performance ^a | 0.213 | 0.210 | 0.424 | 0.161 | 0.365 | 0.080 | 0.010 | 0.040 | 0.080 | 0.151 | 0.234 | 0.192 | 0.255 |

PA = Problem Analysis; J = Judgment; OA = Organizational Ability; D = Decisiveness; L = Leadership; SE = Sensitivity; R = Range of Interests; PM = Personal Motivation; EV = Education Values; ST = Stress Tolerance; OC = Oral Communication; WC = Written Communication.

¹Breed, ²Drepper, ³Ehinger, ⁴Holman, ⁵Jacobson, ⁶Micek, ⁷Schrupt, ⁸Schmitt.

Process. Table 3 represents the effect sizes across criteria by the 12 NASSP skill dimensions and overall placement recommendation. Effect sizes in Table 3 are recorded as standardized Z_r scores allowing comparisons across studies. Results of combined Z_r scores are included in Table 4. The results indicated low relationships (.010 to .240) exist between the NASSP skill dimensions and the various criteria used in the studies to evaluate principals' performance. The overall placement recommendation accounted for a little more than five percent ($r^2 = .0534$) of the variance in the performance criteria.

This meta-analysis indicated low correlations among assessment scores and the various criteria used in the studies. The performance criteria did not effectively measure the types of skills assessed in the NASSP Assessment Process. Several of the skill dimensions appeared to have near zero correlation with performance criteria (e.g., **range of interests, sensitivity, educational values, and personal motivation**). The effect sizes were especially low for the skill dimensions of **sensitivity, range of interest, educational values, and personal motivation** (Table 4). The skill dimensions having the largest effects were **leadership, oral communication, and organizational ability**. These skill dimensions best predict participants' potential for success as prospective school principals. Performance measures developed by Schmitt et al. yielded the highest correlations with the 12 skill dimensions. Schmitt's analysis yielded Z_r 's from .30 to .80

TABLE 4

Combined Effect Sizes For Criteria

| K = 30 | Empirical Estimates from Studies | | | Population Estimate * | |
|------------------------|----------------------------------|---------------------------|---------------------|-----------------------|------------------------------------|
| | Unweighted Mean Z_r | Unweighted Variance Z_r | Weighted Mean Z_r | Unweighted r | Coefficient of Determination r^2 |
| Problem Analysis | 0.218 | 0.126 | 0.129 | 0.215 | 0.0462 |
| Judgment | 0.185 | 0.071 | 0.081 | 0.183 | 0.0335 |
| Organizational Ability | 0.237 | 0.040 | 0.162 | 0.233 | 0.0543 |
| Decisiveness | 0.180 | 0.072 | 0.098 | 0.178 | 0.0317 |
| Leadership | 0.245 | 0.054 | 0.166 | 0.240 | 0.0576 |
| Sensitivity | 0.107 | 0.057 | 0.012 | 0.106 | 0.0112 |
| Range of Interests | 0.100 | 0.058 | -0.002 | 0.010 | 0.0001 |
| Personal Motivation | 0.103 | 0.055 | 0.034 | 0.103 | 0.0106 |
| Educational Values | 0.096 | 0.046 | 0.011 | 0.096 | 0.0092 |
| Stress Tolerance | 0.150 | 0.015 | 0.072 | 0.148 | 0.0219 |
| Oral Communication | 0.242 | 0.063 | 0.116 | 0.237 | 0.0562 |
| Written Communication | 0.199 | 0.155 | 0.079 | 0.197 | 0.0388 |
| Overall Rating | 0.235 | 0.008 | 0.142 | 0.231 | 0.0534 |

* Tabulated by Snedecor & Cochran, 1980, Statistical Methods.

(Table 3). However, none of the NASSP skill dimensions, nor the overall placement recommendation, shared more than a few percent of total variance between the performance criteria and the NASSP Assessment Center Process. Overall, this meta-analysis suggests the NASSP Assessment Center Process has relatively low criterion-related validity given the multiple criteria used to evaluate job performance.

Section Three

Possible Bias Related to Participants' Gender, Position Type, and Race

Participants' assessment scores were compared to determine differences related to possible bias in the NASSP Assessment Center Process. Meta-analysis of the data identified effect sizes for each skill dimension and overall placement recommendation across the studies for the demographic variables of gender, position type, and race.

Gender

Gender was considered an important demographic variable and was included in fifteen studies. Only seven studies reported results conducive for inclusion in the meta-analysis. Participants' scores and gender were investigated to determine possible bias within the NASSP Assessment Process.

Dickson considered participants' career aspirations, perceived career opportunities, and professional development plans related to their Assessment Center experience.³⁰ Dickson analyzed survey results using the independent variable of gender. Significant differences between female and male performance ratings were indicated in **problem analysis, organizational ability, personal motivation, oral communication, and written communication** ($p < .05$). Females typically out-performed males in these five dimensions (Table 5).

Douglas and Johnson examined the 12 skill dimensions for potential bias in the NASSP Assessment Process based upon gender (Table 5).³¹ Correlations for **problem analysis, organizational ability, leadership, written communication, and overall placement recommendation** were significantly higher for females than for males ($p < .05$).

Holman examined the relationship among performance in the Nebraska Assessment Center, conceptual levels, and gender of participants.³²

³⁰Dickson, K. G. (1987). The Oregon assessment center: Its influence on candidates' career opportunities and professional development. Unpublished doctoral dissertation, University of Oregon.

³¹Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300 879)

³²Holman, D. M. (1986). The relationship of participant performance in the Nebraska Assessment Center to participant conceptual level/cognitive complexity. Unpublished doctoral dissertation, University of Nebraska.

TABLE 5

Standardized Effect Size Across Studies (Z_i) for Gender

| K = 7 | Dickson | Douglas | Holman | Rolland | Schmitt ₁ | Schmitt ₂ | Tryjankowski |
|------------------------|---------|---------|--------|---------|----------------------|----------------------|--------------|
| Problem Analysis | 0.813 | 0.003 | 0.100 | 0.804 | 0.124 | 0.185 | 0.232 |
| Judgment | 0.872 | 0.025 | 0.121 | 0.861 | 0.198 | 0.125 | 0.313 |
| Organizational Ability | 0.772 | 0.0003 | 0.202 | 0.786 | 0.092 | 0.182 | 0.227 |
| Decisiveness | | 0.005 | 0.070 | 0.974 | 0.027 | 0.118 | 0.318 |
| Leadership | 0.852 | -0.050 | 0.110 | 0.964 | 0.007 | 0.096 | 0.170 |
| Sensitivity | 0.818 | 0.010 | 0.092 | 0.836 | 0.036 | 0.0997 | 0.113 |
| Range of Interests | | 0.040 | 0.081 | 0.792 | 0.110 | 0.0995 | 0.132 |
| Personal Motivation | 0.798 | 0.040 | 0.061 | 0.912 | 0.134 | 0.160 | 0.176 |
| Education Values | 0.696 | 0.040 | 0.013 | 0.912 | 0.231 | 0.157 | 0.043 |
| Stress Tolerance | 0.685 | | 0.054 | 0.939 | 0.024 | 0.079 | 0.050 |
| Oral Communication | 0.903 | 0.060 | 0.161 | 0.850 | 0.200 | 0.170 | 0.135 |
| Written Communication | 0.746 | 0.006 | 0.234 | 0.517 | 0.424 | 0.379 | 0.242 |
| Overall Rating | | 0.0001 | | 0.792 | 0.098 | 0.171 | 0.282 |
| Sample Size | 344 | 96 | 64 | 176 | 423 | 2831 | 65 |

Level of $p < .05$ is indicated by shaded areas.

Analysis indicated no significant relationships among nine skill dimensions and gender or conceptual levels ($p > .05$). Significant relationships were identified for **organizational ability, decisiveness, and written communication**. High conceptual level females scored significantly higher on these three dimensions than did high and low conceptual level males, and low conceptual level females.

Rolland examined the relationships among participants' scores in the Minnesota Principal Assessment Center (MNPAC) and the demographic variable of participants' gender to determine effects (Table 5).³³ Significant differences were identified in **problem analysis, organizational ability, range of interests, and written communication** related to participants' gender ($p < .05$). Females scored significantly higher than males in these skill areas.

Schmitt analyzed participants' skill dimension scores to determine differences related to gender.³⁴ Significant effects were indicated in **judgment, educational values, oral communication, and written**

³³Rolland, L. M. Y. (1990). A study of the Minnesota Principal Assessment Center and its impact on career opportunities for its participants. Unpublished doctoral dissertation, The University of Minnesota.

³⁴Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

communication ($p < .05$). Female participants received higher ratings than male participants on these skill dimensions.

Schmitt and Cohen concluded females scored significantly higher than males on all skill dimensions and overall placement recommendation.³⁵ The largest differences were in **written communication** and **personal motivation** with female participants scoring higher than male participants.

Tryjankowski investigated the relationships between participants' gender and assessment scores.³⁶ Analysis of effects of participants' gender on scores indicated females received significantly higher ratings on two dimensions: **decisiveness** and **overall placement recommendation**.

Table 5 indicates the standardized Z scores to allow the comparison across studies for gender and each skill dimension and overall placement recommendation. Skill dimensions having significant differences within studies are designated by shaded areas. Effect sizes for gender and the 12 NASSP skill dimensions were combined across seven studies. The sample size distribution was positively skewed due to the presence of the two large national validation studies. Mean effect sizes were computed with and

³⁵Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

³⁶Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association.

without the data from these studies, but no substantive differences resulted. Final results of the meta-analysis for gender presented in Table 6 include the two national validation studies.

Overall, the effect sizes combined across studies suggested that gender does not play a significant role in assessment center scores on any dimension, although individual studies suggested significant statistical differences between female and male scores. **Written communication** ($r = .280$), **judgment** ($r = .232$), and **oral communication** ($r = .228$) were the dimensions with the greatest differences between males and females ($p < .05$). Comparison of substantive differences, however, corroborated the meta-analysis findings of small or trivial gender effects. Although females generally score higher than males on all 12 skill dimensions, differences when scores were combined across studies were not substantive or great enough to indicate the presence of bias based upon gender in the NASSP Assessment Center Process.

Position Type

Five studies recorded NASSP Assessment Center scores and participants' position types. Comparisons of results across studies examined possible bias of the NASSP Assessment Center Process related to position type.

TABLE 6

Combined Effect Sizes For Gender

| K = 8 | Empirical Estimates from Studies | | | Population Estimates * | |
|------------------------|----------------------------------|---------------------------|---------------------|------------------------|------------------------------------|
| | Unweighted Mean Z_r | Unweighted Variance Z_r | Weighted Mean Z_r | Unweighted r | Coefficient of Determination r^2 |
| Problem Analysis | 0.321 | 0.098 | 0.264 | 0.312 | 0.097 |
| Judgment | 0.357 | 0.108 | 0.240 | 0.341 | 0.116 |
| Organizational Ability | 0.321 | 0.087 | 0.255 | 0.312 | 0.097 |
| Decisiveness | 0.215 | 0.105 | 0.151 | 0.211 | 0.045 |
| Leadership | 0.312 | 0.143 | 0.202 | 0.302 | 0.091 |
| Sensitivity | 0.285 | 0.117 | 0.196 | 0.277 | 0.077 |
| Range of Interests | 0.178 | 0.064 | 0.133 | 0.174 | 0.030 |
| Personal Motivation | 0.324 | 0.114 | 0.252 | 0.314 | 0.099 |
| Educational Values | 0.297 | 0.109 | 0.249 | 0.287 | 0.082 |
| Stress Tolerance | 0.260 | 0.125 | 0.174 | 0.254 | 0.065 |
| Oral Communication | 0.352 | 0.110 | 0.272 | 0.338 | 0.114 |
| Written Communication | 0.362 | 0.047 | 0.408 | 0.347 | 0.120 |
| Overall Rating | 0.191 | 0.069 | 0.182 | 0.187 | 0.035 |

* Tabulated by Snedecor & Cochran, 1980, Statistical Methods.

Blanck determined mean scores on the 12 skill dimensions for five role groups in Michigan: elementary, middle, and high school teachers; assistant principals; and quasi-administrators.³⁷ No significant differences were found among the three teacher groups on any of the 12 skill dimensions. Significant differences were identified among assistant principals and the three groups of teachers for the skill dimension of **decisiveness**. Other differences were among quasi-administrators and teacher groups in **organizational ability, problem analysis, leadership, oral communication, and overall placement recommendation**. There were no significant differences among the groups for the remaining skill dimensions.

Farmer addressed participants' perceptions of assessment by asking teachers, assistant principals, and principals to rank the importance of the skill dimensions to their jobs.³⁸ Teachers' opinions often depended upon whether they received a "recommend" or "not recommend" score to be placed in the pool of potential assistant principals. Teachers and assistant principals disagreed on the use of assessment results for promotional decisions; assistant principals tended to be more positive in using scores for

³⁷Blanck, J. I. (1989). Comparison of five role groups of Michigan participants in NASSP assessment centers. Unpublished doctoral dissertation, Western Michigan University.

³⁸Farmer, T. N. (1987). An analysis of the effects of participation in the National Association of Secondary School Principals' Assessment Center in Howard County, Maryland. Unpublished doctoral dissertation, University of Maryland.

promotion. Although both groups wanted the inclusion of the results in the selection process, they did not want scores to receive more weight than job performance or interview results.

Ford analyzed the effect of position on attitude toward the assessment process.³⁹ There was no significant relationships between participants' attitude ratings and position. Ford's conclusions did not support the presence of bias based on position of participants in the NASSP Assessment Center Process.

Holman studied the relationships among conceptual level, participant performance in the Nebraska Assessment Center, and occupational level.⁴⁰ Eleven skill dimensions were found not to be related to either conceptual or occupational level. Significant relationships were identified among conceptual level, occupational level, and the skill dimension of oral communication.

³⁹Ford, M. J. (1987). An attitude survey of NASSP principal assessment center candidates. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

⁴⁰Holman, D. M. (1986). The relationship of participant performance in the Nebraska Assessment Center to participant conceptual level/cognitive complexity. Unpublished doctoral dissertation, University of Nebraska.

Schmitt's first validation study analyzed significant effects for participants' scores and position type.⁴¹ Elementary, middle, and high school teachers, counselors, and assistant principals' scores were compared. No significant relationship was indicated when comparing the assessment scores of the three levels of teachers. Results indicated higher scores for non-teachers in every dimension except **sensitivity**, although **personal motivation** was the only area where a significant relationship was indicated. Counselors compared to non-counselors scored higher in every skill area and overall placement recommendation. Assistant principals scored significantly higher when compared to non-assistant principals in **problem analysis** and **decisiveness**.

Schmitt and Cohen analyzed position type comparing elementary, middle, and high school teachers, counselors, and assistant principals.⁴² There were no significant differences among the scores on the 12 skill dimensions and the three levels of teachers. Non-teaching personnel scored significantly higher than teachers in every dimension except **range of interests**.

⁴¹Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

⁴²Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

TABLE 7
Standardized Effect Size Across Studies (Z_i) for Position Type

| K = 8 | Blanck | Farmer | Holman | Schmitt 1 C/NC | Schmitt 1 AP/NAP | Schmitt 1 T/NT | Schmitt 2 AP/NAP | Schmitt 2 T/NT |
|------------------------|--------|--------|--------|-------------------|---------------------|-------------------|---------------------|-------------------|
| Problem Analysis | 0.235 | 0.503 | 0.143 | 0.188 | 0.177 | 0.012 | 0.039 | 0.070 |
| Judgment | 0.155 | 0.646 | 0.123 | 0.212 | 0.116 | 0.082 | 0.108 | 0.117 |
| Organizational Ability | 0.168 | 0.417 | 0.176 | 0.209 | 0.129 | 0.102 | 0.078 | 0.092 |
| Decisiveness | 0.218 | 0.191 | 0.194 | 0.111 | 0.132 | 0.021 | 0.132 | 0.081 |
| Leadership | 0.185 | 0.182 | 0.139 | 0.162 | 0.061 | 0.112 | 0.134 | 0.129 |
| Sensitivity | 0.133 | 0.161 | 0.056 | 0.138 | 0.055 | 0.062 | 0.054 | 0.079 |
| Range of Interests | 0.151 | 0.209 | 0.095 | 0.109 | 0.060 | 0.043 | 0.094 | 0.058 |
| Personal Motivation | 0.081 | 0.160 | 0.158 | 0.155 | 0.039 | 0.137 | 0.120 | 0.108 |
| Education Values | 0.105 | 0.326 | 0.044 | 0.181 | 0.055 | 0.125 | 0.059 | 0.088 |
| Stress Tolerance | 0.112 | 0.181 | 0.126 | 0.086 | 0.061 | 0.127 | 0.095 | 0.070 |
| Oral Communication | 0.198 | 0.126 | 0.189 | 0.119 | 0.009 | 0.073 | 0.141 | 0.113 |
| Written Communication | 0.096 | 0.317 | 0.233 | 0.155 | 0.032 | 0.077 | 0.012 | 0.070 |
| Overall Rating | 0.200 | | | 0.181 | 0.106 | 0.098 | 0.167 | 0.185 |
| Sample Size | 394 | 172 | 64 | 265 | 313 | 313 | 2831 | 2831 |

C/NC = Counselor/Non Counselor; AP/NAP = Assistant Principal/Non Assistant Principal; T/NT = Teacher/Non Teacher

Level of $p < .05$ is indicated by shaded areas.

Table 7 indicates the standardized Z scores to allow the comparison across individual studies for each skill dimension and overall placement recommendation related to position type. Skill dimensions identified to have significant differences are indicated by shaded areas. Table 7 depicts the differences as represented by the effect sizes from study to study and represents data used to determine combined effect sizes in Table 8.

Effect sizes combined across five studies for position types are recorded in Table 8. These effect size estimates must be considered less reliable than the gender and race variable estimates since the combined effect sizes for position type failed the test for homogeneity. The various groupings of participants' positions made the lines of distinction for position types less exact. Yet, comparing administrative and non-administrative or teacher and non-teacher allows some indication of the relationship among participants' scores and position types.

Although individual studies described differences in participants' scores relating to the 12 skill dimensions and position types, the meta-analysis did not support those differences. Less than two and one half percent ($r^2 = .0246$) of the variance in overall placement recommendations can be attributed to the participants' position types. The skill dimension having the greatest effect size for position type was **judgment** ($r^2 = .0392$). **Organization ability** ($r^2 = .030$) and **problem analysis** ($r^2 = .0299$) ranked second and third for largest effect sizes. Differences in assessment scores

TABLE 8

Combined Effect Sizes For Position Type

| K = 5 | Empirical Estimates from Studies | | | Population Estimates* | |
|------------------------|----------------------------------|---------------------------|---------------------|-----------------------|------------------------------------|
| | Unweighted Mean Z_r | Unweighted Variance Z_r | Weighted Mean Z_r | Unweighted r | Coefficient of Determination r^2 |
| Problem Analysis | 0.171 | 0.021 | 0.128 | 0.173 | 0.0299 |
| Judgment | 0.195 | 0.031 | 0.117 | 0.198 | 0.0392 |
| Organizational Ability | 0.172 | 0.010 | 0.074 | 0.174 | 0.0303 |
| Decisiveness | 0.135 | 0.003 | 0.129 | 0.136 | 0.0185 |
| Leadership | 0.138 | 0.001 | 0.083 | 0.139 | 0.0193 |
| Sensitivity | 0.092 | 0.002 | 0.066 | 0.092 | 0.0085 |
| Range of Interests | 0.102 | 0.003 | 0.122 | 0.102 | 0.0104 |
| Personal Motivation | 0.120 | 0.002 | 0.111 | 0.121 | 0.0146 |
| Educational Values | 0.123 | 0.008 | 0.099 | 0.124 | 0.0154 |
| Stress Tolerance | 0.107 | 0.001 | 0.096 | 0.107 | 0.0114 |
| Oral Communication | 0.121 | 0.003 | 0.089 | 0.122 | 0.0149 |
| Written Communication | 0.124 | 0.009 | 0.137 | 0.125 | 0.0156 |
| Overall Rating | 0.156 | 0.002 | 0.051 | 0.157 | 0.0246 |

* Tabulated by Snedecor & Cochran, 1980, Statistical Methods.

related to participants' position types were very small (Table 8). The differences when scores were combined across studies were not substantive or great enough to indicate the presence of bias in the NASSP Assessment Center Process based upon participants' position types.

Race

Throughout the implementation of the assessment process questions have been raised concerning the NASSP Assessment Center Process and the effects of ethnic backgrounds upon participants' scores. This meta-analysis examined possible bias of the NASSP Assessment Center Process based upon race. Information describing race of participants was reported in seventeen studies. Only four studies recorded data in a format conducive for use in meta-analysis.

Rolland examined the relationships between participants' scores in the Minnesota Principal Assessment Center (MNPAC) and participants' race (Table 9).⁴³ Race analysis identified significant differences in **problem analysis, judgment, organizational ability, decisiveness, oral communication, written communication, and overall placement recommendation** between

⁴³Rolland, L. M. Y. (1990). A study of the Minnesota Principal Assessment Center and its impact on career opportunities for its participants. Unpublished doctoral dissertation, The University of Minnesota.

TABLE 9**Standardized Effect Size Across Studies (Z_j) for Race**

| K=4 | Rolland | Schmitt 1 | Schmitt 2 | Tryjankowski |
|------------------------|---------|-----------|-----------|--------------|
| Problem Analysis | 0.703 | 0.390 | 0.397 | 0.147 |
| Judgment | 0.688 | 0.460 | 0.371 | 0.101 |
| Organizational Ability | 0.698 | 0.525 | 0.413 | 0.243 |
| Decisiveness | 0.804 | 0.445 | 0.243 | 0.141 |
| Leadership | 0.878 | 0.072 | 0.300 | 0.143 |
| Sensitivity | 0.901 | 0.011 | 0.127 | 0.054 |
| Range of Interests | 0.876 | 0.119 | 0.181 | 0.198 |
| Personal Motivation | 0.875 | 0.155 | 0.096 | 0.209 |
| Education Values | 0.877 | 0.411 | 0.038 | 0.207 |
| Stress Tolerance | 0.960 | 0.571 | 0.219 | 0.138 |
| Oral Communication | 0.703 | 0.067 | 0.283 | 0.132 |
| Written Communication | 0.768 | 0.202 | 0.279 | 0.312 |
| Overall Rating | 0.812 | 0.520 | 0.232 | 0.114 |
| Sample Size | 175 | 313 | 2831 | 65 |

Level of $p < .05$ is indicated by shaded areas.

white and non-white participants ($p < .05$). White participants scored higher than non-white participants.

Schmitt's validation study of the NASSP Assessment Center Process indicated significant relations among participants' race and the skills of **problem analysis, judgment, decisiveness, leadership, written communication, organizational ability, and overall placement recommendation.**⁴⁴ White participants received higher ratings than non-whites participants in these skill dimensions. Schmitt wrote that differences among groups may or may not be unfair to a particular group.

If similar differences are observed in performance ratings, no problem exists. That is, the assessment center is merely identifying job relevant subgroup differences.⁴⁵

Schmitt's second validation study also analyzed participants' scores related to race.⁴⁶ White candidates received significantly higher ratings than non-whites in nine skill dimensions (Table 9).

⁴⁴Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

⁴⁵Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology, 18.

⁴⁶Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. East Lansing, MI: Michigan State University, Department of Psychology.

Tryjankowski investigated the relationship among participants' skill ratings for white and non-white participants.⁴⁷ The analysis indicated significant differences between scores of white and non-white participants in **written communication**. Significant differences were not identified in remaining 11 skill dimensions.

Table 9 indicates standardized Z_r scores to allow comparisons across studies for race and skill dimensions. Levels of significance are designated by shaded areas. Effect sizes were combined across four studies for race. The variable was defined by categories of white and non-white. Table 10 records combined effect sizes across studies reported as Z_r scores for the 12 skill dimensions and overall placement recommendation.

Differences attributed to race are represented in Table 10. **Written communication** accounted for 20.3 percent of the differences in participants' scores based upon race. **Organizational ability** constituted 20.1 percent and **problem analysis** represented 20.8 percent of the differences in participants' scores. The overall placement recommendations indicated the r equivalent to be .395, which is considered moderate. Overall race accounted for 15.6 percent of the differences in white and non-white participants' scores.

⁴⁷Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association.

TABLE 10

Combined Effect Sizes For Race

| K = 4 | Empirical Estimates from Studies | | | Population Estimates * | |
|------------------------|----------------------------------|---------------------------|---------------------|------------------------|------------------------------------|
| | Unweighted Mean Z_r | Unweighted Variance Z_r | Weighted Mean Z_r | Unweighted r | Coefficient of Determination r^2 |
| Problem Analysis | 0.410 | 0.039 | 0.415 | 0.456 | 0.208 |
| Judgment | 0.405 | 0.044 | 0.398 | 0.422 | 0.178 |
| Organizational Ability | 0.482 | 0.029 | 0.446 | 0.448 | 0.201 |
| Decisiveness | 0.429 | 0.067 | 0.308 | 0.403 | 0.162 |
| Leadership | 0.442 | 0.075 | 0.354 | 0.416 | 0.173 |
| Sensitivity | 0.288 | 0.126 | 0.177 | 0.281 | 0.079 |
| Range of Interests | 0.330 | 0.102 | 0.221 | 0.319 | 0.102 |
| Personal Motivation | 0.346 | 0.096 | 0.164 | 0.332 | 0.110 |
| Educational Values | 0.283 | 0.123 | 0.010 | 0.277 | 0.077 |
| Stress Tolerance | 0.359 | 0.122 | 0.264 | 0.343 | 0.118 |
| Oral Communication | 0.318 | 0.053 | 0.301 | 0.306 | 0.094 |
| Written Communication | 0.488 | 0.032 | 0.331 | 0.450 | 0.203 |
| Overall Rating | 0.419 | 0.073 | 0.298 | 0.395 | 0.156 |

* Tabulated by Snedecor & Cochran, 1980, Statistical Methods.

These results were borne out by several of the individual studies in which non-whites scored significantly lower than whites.

Summary

Chapter 4 contains the results of the meta-analysis regarding the criterion-related validity of the NASSP Assessment Center Process and the results of effects related to bias within the process. Results from the studies indicated NASSP process to be reliable. Weaknesses appeared in discriminant validity with several studies addressing multicollinearity of the 12 skill dimensions. Findings combined across studies to determine criterion-related validity of the NASSP process found overall effect sizes to be low (Table 4). The presence of bias within the NASSP Assessment Center Process related to gender and position type was not indicated. The results identified race to have moderate effect sizes related to participants' scores (.277 to .456).

CHAPTER 5

Summary, Conclusions, Implications for Practice, and Recommendations for Further Research

The NASSP Assessment Center Process is designed to help identify participants' potential for success as prospective school principals. The purpose of this study was to synthesize the research findings pertaining to criterion-related validity of the NASSP Assessment Center Process and to present these results in a coherent summary.

Chapter 5 is divided into four sections: the summary of the research questions and the results of the meta-analysis, the results of the criterion-related validity analysis, the implications for practice and application of the research findings, and the recommendations for further research to broaden the usefulness of NASSP assessment results.

Summary

The purpose of this study was to systematically synthesize and analyze, through a meta-analytic method, the research findings pertaining to criterion-related validity of the NASSP Assessment Center Process. Three research questions concerning the criterion-related validity of the NASSP Assessment Center Process were addressed. (1) What is the validity of the NASSP Assessment Center Process based upon research findings across the

various studies? (2) How well do the 12 NASSP skill dimensions assess participants' potential for success as prospective school principals? (3) What does research indicate about bias in the NASSP assessment results related to participants' gender, position type, and race?

Analysis of the research indicated the following responses to the research questions. (1) The meta-analysis indicated a low criterion-related validity among the NASSP ratings and the multiple performance criteria in the various studies (Table 3). (2) The skill dimensions having the greatest effect sizes were **written communication, leadership, and organizational ability**. (3) Although several demographic variables were described in the studies, the only data sufficiently reported to allow analysis of effect sizes were for gender, position type, and race. There were no differences in the dimensions or the overall placement recommendation based on participants' gender, position type, or race.

Conclusions

(1) The meta-analysis found the NASSP Assessment Center Process to have low criterion validity related to the performance criteria. The studies used different criteria to measure validity of the NASSP Process, e.g., attitude ratings, conceptual levels, job performance ratings from various personnel, and teacher morale. Participants' assessment scores and ratings on multiple criteria to measure job performance produced low correlations.

Inappropriate criteria used to assess validity may cause low correlation coefficients.

(2) The 12 skill dimensions in the assessment process are highly intercorrelated. High intercorrelations among the skill dimensions indicated multicollinearity and the lack of discriminant validity in the process to distinguish and assess participants' skill levels on 12 separate and distinct dimensions.

(3) The meta-analysis determined no significant differences in participants' assessment scores related to their gender, position type, and race. Although women generally score higher than men, assistant principals and quasi-administrators score higher than teachers, and whites score higher than non-whites, the differences are not substantive.

Implications for Practice

The analysis of the skill dimensions found the skills of **sensitivity, range of interests, personal motivation, educational values, and stress tolerance** to have very low effect sizes across the studies. To report these skill dimensions as descriptive information and not include them as numerical values would increase the effect sizes for the remaining skill dimensions since multicollinearity of the 12 skill dimensions reduce the effect sizes for each dimension. Eliminating the five low effect size skill dimensions in the numerical scores would present a more precise report of the effect sizes for

the remaining administrative skill dimensions giving more accurate administrative profiles of the participants.

The increased interest in the NASSP Assessment Center Process warrants a need for a designated "clearinghouse" for the maintenance of research studies related to the NASSP Assessment Center. This clearinghouse could allow referencing of prior research and assist in making future research more comprehensive. Instituting such a center or assigning the responsibility would seem to be to NASSP's advantage.

Recommendations for Further Research

The following recommendations are submitted to expand the understanding of the assessment process, to allow insight into areas needing refinement, and to enhance the usefulness of assessment results. Implementing these measures could increase opportunities to more accurately identify potentially successful principals, serve to refine principal preparation programs, and enhance professional development.

(1) The reporting of one overall placement recommendation score to NASSP headquarters is insufficient to allow extensive research. Individual scores for each skill dimension should be recorded to enable more analysis. An expanded data base at national headquarters would allow and encourage more research.

The Skill Descriptors and Overall Performance Rating Descriptors used in the assessor training program include thirteen precise descriptors of participant scores.¹ Reporting scores to national in this format is recommended. For example: the participant received the scores of five 4's and one 3 from the assessors, the descriptor would read: "The participant demonstrated a high degree of skill." Or, if the scores were two 2's and four 1's: "The participant demonstrated no skill." Results reported in this manner would allow more specific analysis.

(2) There is a need to identify the specific number of participants having received an **overall placement recommendation** score rather than combining groups of scores: 0 to 1.99; 2.00 to 2.99; and 3.00 to 5.00. Establishing an expanded data base for the NASSP assessment results would allow further research of the total NASSP Assessment Center population.

(3) Meta-analysis of the studies suggested that improvements are needed in the data available for validity research. Additional data reported in a format that allows the combining of research findings across studies is recommended. Reporting of results using *t*, *F*, and *r* statistics would increase accessibility.

¹Flanary, R. A. (1990). Writing the final assessment report. Reston, VA: NASSP Assessment Center Training Materials, NASSP.

(4) High correlation coefficients among the 12 skill dimensions indicated skill dimensions are overlapping. Schmitt, Tryjankowski, Ogawa and Oxaal identified multicollinearity among the 12 skill dimensions. Analysis of the 12 skill dimensions across the studies found the effect sizes for all performance criteria to be low (r ranging from .010 to .237). Criteria effect sizes were relatively small since the effect of one skill dimension may be canceling the effect size of others.

Factor analysis performed at each center may support the existence of multicollinearity among the skills and support the need for more specific indicators to assess the 12 skills individually. Differences in factor analysis occurring from one center to another would influence validity of the process. Differences in the number of factors should not be reflected from center to center when factor analyses are performed. "Unofficial" weighting of various skills could explain these differences. The high performance of a few skill areas leading to an over-rating of all skill dimensions is referred to as "halo" effect which could also result in unofficial weighting of particular skill dimensions. The low criterion-related validity also indicates a need to go back and examine the construct validity of the process. Although not a topic in this research project, there were questions regarding construct validity of the NASSP Assessment Process. Research in the area related to the construct validity is needed to establish construct validity for the process.

(5) Analysis of the current skill dimensions to identify component factors would contribute to a more accurate reporting of NASSP assessment results. Identifying these factors would allow a more effective description of participants' results related to how the skill dimensions work together. Factor analysis of the 12 skill dimensions from the studies suggest one to four performance categories. The grouping of the 12 skills into categories suggested by Schmitt could increase effect sizes and address multicollinearity between the skills.² Schmitt suggested the dominant factor to be administrative and include **problem analysis, organizational ability, and judgment**. **Interpersonal, values, and personal motivation** dimensions would be the secondary factor. **Oral communication, sensitivity, and stress tolerance** were high on both factors. Improved categorization by factor analysis would focus on the administrative performance and more appropriately identify participants' strengths and recommended areas for improvement. Participants' scores reported by categories could increase overall validity of the process by not attempting to assess 12 separate skill dimensions.

(6) Relatively small effect sizes between the NASSP skill dimensions and performance criteria from the various studies suggest a need for more

²Schmitt N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. Unpublished report, East Lansing, MI: Michigan State University, Department of Psychology, 12.

appropriate means to evaluate the presence of the 12 skill dimensions on the job. Criterion-related validity is based upon appropriateness of the criterion measure. Conceptual levels, attitude ratings, and teacher morale fail to appropriately measure principals' job performance related to the NASSP skill dimensions.

Studies comparing participants' NASSP results with actual job performance criteria evaluated by superordinates and subordinates provided higher validity correlations than other comparisons. Schmitt's criterion measures more appropriately measured the application of the skill dimensions to the job. Instruments to better evaluate skill application with performance are necessary to more accurately determine criterion-related validity. There is a need to develop a model to assess specific skill dimensions demonstrated on the job rather than criteria that evaluate overall job performance.

(7) Pre-and post survey could document participants' reactions to the process and support validity. Opinions expressed in the survey would serve as a criterion measure, showing how well participants feel they will and have performed in the exercises.

(8) A longitudinal study to determine success of participants receiving low scores and not recommended compared to those participants receiving high scores and recommended is suggested. Appointments to an administrative position would be the criterion.

(9) Effects of demographic variables should be more comprehensively researched. Years of experience and age were not reported sufficiently to allow analysis. Other demographics to be considered for study include comparisons of assessment scores and performance in university administrative programs, urban and rural scores, and comparisons based on socio-economic factors.

In 1988, Peters and Bagenstos recorded that 74 percent of the states were mandating or planning to mandate principal evaluations.³

Acknowledging the importance of the principalship has increased use of the NASSP Assessment Process for a more comprehensive and objective selection procedure. NASSP established the principal assessment center to provide school districts with an "independent and objective assessment of candidates' readiness to assume administrative responsibilities."⁴

The selection of principals is important and the NASSP process has improved the selection process over the "good ole boys" method of selection.⁵ The increased use of the NASSP process for selection purposes

³Peters, S., & Bagenstos, N. (1988, March). State-mandated principal evaluation: A report on current practice. (Report No. EA 292 889). New Orleans, LA: American Educational Research Association, 12.

⁴National Association of Secondary School Principals. (1988) NASSP Assessment Handbook. Reston, VA, 9.

⁵Baltzell, D., & Dentler, R. (1983b). Selecting principals for American's schools: A source book for educators. Washington, DC: National Institute of Education.

increased the need for a more focused and comprehensive look at the research related to the predictive validity of the process. The purpose of this study was to systematically synthesize and analyze the results of the research findings pertaining to the criterion-related validity of the NASSP Assessment Center Process. This study is a small step to address this void and to confront this need. Perhaps the information gleaned from this meta-analysis will assist in the identification of strengths and weaknesses, serve to initiate refinement and modification of the process, and enhance the NASSP Assessment Center Process to better predict participants' potential for success as prospective school administrators.

As the 21st century approaches, educators are implementing changes to better equip our children for the future. The results of this study challenge NASSP to refine the assessment process which would improve the selection of potentially successful principals and address principals' developmental needs. Further research on the predictive validity of the process, however, is necessary. This is an exciting time for the National Association of Secondary School Principals and others in the field of education to take the initiative and accept the challenges for improvement.

List of Reviewed Studies

Alvy, H. B. (1983). The problem of new principals. Unpublished doctoral dissertation, University of Montana.

Blanck, J. I. (1989). Comparison of five groups of Michigan participants in NASSP assessment centers. Unpublished doctoral dissertation, Western Michigan University.

Bley, J. E. (1983). Use of an assessment center for selection assessment center for selecting administrators and diagnosing inservice needs. Unpublished doctoral dissertation, Northern Arizona University.

Breed, R. D. (1985). Concurrent validity of two administrator selection procedures. Unpublished doctoral dissertation, University of Nebraska.

Britt, W. C. (1976). A comparative study of on-the-job ratings to assessment center ratings of school administrators. Unpublished doctoral dissertation, Georgia State University.

Crumpton, S. L. (1989). The importance of the National Association of Secondary School Principals Assessment Center Report to the appointment of public school principals in South Carolina as perceived by superintendents and first-time principals. Unpublished doctoral dissertation, University of South Carolina.

Dickson, K. G. (1987). The Oregon Assessment Center: Its influence on candidates' career opportunities and professional development. Unpublished doctoral dissertation, University of Oregon.

Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300879)

Draper, J. M. (1979). A comparative study of elementary school principals' on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Unpublished doctoral dissertation, Georgia State University.

Ehinger, J. M. (1986). The validation of selected assessment center activities for incumbent school administrators who participated in the

University of Tulsa education management development program.
Unpublished doctoral dissertation, University of Tulsa.

Engel, T. E. (1989). An analysis of the content and methods of instruction at Michigan institutions that prepare principals. Unpublished doctoral dissertation, Western Michigan University.

Farmer, N. T. (1985). An analysis of the effects of participation in the National Association of Secondary School Principals' Assessment Center in Howard County, Maryland. Unpublished doctoral dissertation, University of Maryland.

Ford, M. J. (1987). An attitude survey of NASSP Principal Assessment Center candidates. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Friedman, M. J. (1980). Differences in assessment center performance as a function of the race and sex of ratees and the race of assessors. Unpublished doctoral dissertation, The University of Tennessee.

Heater, J. D. (1989). Teacher morale as a criterion of the predictive validity of the National Association of Secondary School Principals Assessment Center. Unpublished doctoral dissertation, University of Nebraska.

Holman, D. M. (1986). The relationship of participants' performance in the Nebraska Assessment Center to participant conceptual level/cognitive complexity. Unpublished doctoral dissertation, University of Nebraska.

Jacobson, B. J. (1986). Predictors of success for school administrators. Unpublished doctoral dissertation, University of Nebraska.

Lester, R. A. (1991). Construct validity of a National Association of Secondary School Principals' Assessment Center. Unpublished doctoral dissertation, Georgia State University.

Micek, T. J. (1983). Teacher perceptions of performance of the NASSP assessment center behavior dimensions as it relates to the conceptual complexity of principals. Unpublished doctoral dissertation, University of Nebraska.

Ogawa, R. T., & Oxaal, I. (1986, April). A reexamination of NASSP assessment center ratings: research notes. Paper presented at the Annual Meetings of the American Educational Research Association, San Francisco.

Pokorny, N. G. (1985). Techniques used in the selection of public school principals in the United States. Unpublished doctoral dissertation, University of Colorado.

Rolland, L. M. Y. (1990). A study of the Minnesota principal assessment center and its impact on career opportunities for its participants. Unpublished doctoral dissertation, The University of Minnesota.

Schrup, M. G. (1989). Role of reflection on administrative job performance following National Association of Secondary School Principals' assessor training and assessment center service. Unpublished doctoral dissertation, Indiana University.

Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP Assessment Center. Unpublished report, East Lansing, MI: Michigan State University, Department of Psychology.

Schmitt N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP Assessment Center. Unpublished report, East Lansing, MI: Michigan State University, Department of Psychology.

Scholl, S. S. (1980). Criteria for selection of Ohio middle school principals with particular attention to the leadership role in curriculum development. Unpublished doctoral dissertation, Miami University.

Storlien, H. T. (1983). Recruitment and selection of secondary principals in Alberta: an analysis of procedures, criteria, and identified problems. Unpublished doctoral dissertation, University of Montana.

Travetto, S. K. (1990). An analysis of school system and participant use of assessment center for professional growth and development. Unpublished doctoral dissertation, University of Maryland.

Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association.

Van Newkirk, C. J. (1984). A study of the assessment center as a process for identifying prospective school principals and assistant principals. Unpublished doctoral dissertation, College of William and Mary.

Voit, L. B. (1989). Perceptive and evaluation of university principal preparation programs by Michigan Public School Principals. Unpublished doctoral dissertation, Western Michigan University.

Walden, D. L. (1985). The effects of the NASSP Assessment Center Process on self-improvement of assessees. Unpublished doctoral dissertation, University of Colorado.

Wendel, F. C., Gappa, L. R., & Yusten, C. R. Use of the assessment center method in administrator preparation programs. Tempe: Arizona State University, The University Council for Educational Administration.

Witters-Churchill, L. J. (1988). University preparation of the school administrator: evaluations by Texas principals. Unpublished doctoral dissertation, Texas A & M University.

REFERENCES

- Alvy, H. B. (1983). The problem of new principals. Unpublished doctoral dissertation, University of Montana, 163.
- Baltzell, D., & Dentler, R. (1983b). Selecting principals for American's schools: A source book for educators. Washington, DC: National Institute of Education.
- Blanck, J. I. (1989). Comparison of five role groups of Michigan participants in NASSP assessment centers. Unpublished doctoral dissertation, Western Michigan University.
- Bley, J. E. (1983). Use of an assessment center for selection assessment center for selecting administrators and diagnosing inservice needs. Unpublished doctoral dissertation, Northern Arizona University.
- Blumberg, A., & Greenfield, W. (1986). The effective principal: Perspectives on school leadership. Boston: Allyn & Bacon, 239.
- Bray, D. (1985). Fifty years of assessment centers: A retrospective and prospective view. Journal of Management Development, 4(4), 4-12.
- Bray, D. W., Campbell, B. J., & Grant, D. L. (1974). Formative years in business: A long-term AT&T study of managerial lives. New York: Wiley.
- Breed, R. D. (1985). Concurrent validity of two administrator selection procedures. Unpublished doctoral dissertation, University of Nebraska.
- Britt, W. C. (1976). A comparative study of on-the-job ratings to assessment center ratings of school administrators. Unpublished doctoral dissertation, Georgia State University.
- Britt, W. C. (1976). Assessment center ratings of school administrators. Unpublished doctoral dissertation, Georgia State University.
- Brookover, W. B., & Lezotte, L. W. (1979, May). Changes in school characteristics coincident with changes in student achievement. East Lansing, MI: Institute for Research on Teaching, Michigan State University.

- Byham, W. (1979). Review of legal cases and opinions dealing with assessment centers and content validity. Pittsburgh, PA: Development Dimensions International.
- Byham, W. C. (1985, July). Assessment centers: A method with proven mettle. Training, 69-70.
- Byham, W. C. (1986). Use of the assessment center method to evaluate teacher competencies. Pittsburgh, PA: Developmental Dimensions International.
- Campbell, D., & Fiske, D. (1959, March). Convergent and discriminate validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Cohen, B. M., Moses, J. L., & Byham, W. C. (1974). The validity of assessment centers: A literature review. Monograph II. Pittsburgh, PA: Development Dimensions Press.
- Cohen, S. L. (1980, February). The bottom line on assessment center technology: Results of a cost benefit analysis survey. The Personnel Administrator, 25(2), 54.
- Cooper, H., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. Psychological Bulletin, 87, 442.
- Crumpton, S. L. (1989). The importance of the National Association of Secondary School Principals Assessment Center Report to the appointment of public school principals in South Carolina as perceived by superintendents and first-time principals. Unpublished doctoral dissertation, University of South Carolina.
- Culbertson, J., Farguhar, R. H., Gaynor, A. K., & Shibles, M. R. (1969). Preparing educational leaders for the seventies. Columbus, OH: The University Council for Educational Administration.
- Delon, F. G. (1988). Assessment center screening for school administration certification and employment: Possible legal challenges. West Publishing Company, 43 W. Educ. L. Rep. 841.

- Dickson, K. G. (1987). The Oregon assessment center: Its influence on candidates' career opportunities and professional development. Unpublished doctoral dissertation, University of Oregon.
- Douglas, J. R., & Johnson, M. C. (1984). Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300 879)
- Draper, J. M. (1979). A comparative study of elementary school principals' on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Unpublished doctoral dissertation, Georgia State University.
- Edmonds, R. (1979, October). Effective schools for the urban poor. Educational Leadership, 37(1), 15-18, 20-24.
- Ehinger, J. M. (1986). The validation of selected assessment center activities for incumbent school administrators who participated in the University of Tulsa Education Management Development Program. Unpublished doctoral dissertation, University of Tulsa.
- Engel, T. E. (1989). An analysis of the content and methods of instruction at Michigan institutions that prepare principals. Unpublished doctoral dissertation, Western Michigan University.
- Farmer, T. N. (1987). An analysis of the effects of participation in the National Association of Secondary School Principals' Assessment Center in Howard County, Maryland. Unpublished doctoral dissertation, University of Maryland.
- Flanary, R. A. (1990). Writing the final assessment report. Reston, VA: NASSP Assessment Center Training Materials, NASSP.
- Flanary, R. (1991, December). History of the NASSP Assessment Center. Unpublished manuscript, Blacksburg, VA.
- Ford, M. J. (1987). An attitude survey of NASSP principal assessment center candidates. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

- Friedman, M. J. (1980). Differences in assessment center performance as a function of the race and sex of ratees and the race of assessors. Unpublished doctoral dissertation, The University of Tennessee.
- Gallagher, M. (1978). The development and pilot operation of an assessment center with implications for the selection of school principals. Unpublished doctoral dissertation, Western Michigan University.
- Gaugler, B., Rosenthal, D., Thornton, G., & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 72(3), 493-511.
- Gaugler, B., Schmidt, F., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage Publishing, 16.
- Goodlad, J. I. (1984). A place called school: Prospects for the future. New York: McGraw-Hill.
- Gousha, R., Lopresti, P., & Jones, A. (1987). Administrators join revolution in professional standards. The School Administrator, 44(2), 10-15.
- Heater, J. D. (1989). Teacher morale as a criterion of the predictive validity of the National Association of Secondary School Principals' assessment center. Unpublished doctoral dissertation, University of Nebraska.
- Hedges, L., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press, 2.
- Hersey, P. W. (1987). How NASSP helps identify, develop superior principals. Seminar on Innovative Approaches to Assessment to Teaching Staff in the Education Service. Edinburgh, Scotland: NASSP, Reston, VA, 1.
- Holman, D. M. (1986). The relationship of participant performance in the Nebraska Assessment Center to participant conceptual level/cognitive complexity. Unpublished doctoral dissertation, University of Nebraska.
- Huck, J. R. (1973). A review of the external and internal validities. Personnel Psychology, 26(2), 191-212.

- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.
- Hunter, J., Schmidt, F., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.
- Hyde, J., & Linn, M. (1986). The psychology of gender, advances through meta-analysis. Baltimore: John Hopkins University Press, 3.
- Jacobson, B. J. (1986). Predictions of success for school administrators. Unpublished doctoral dissertation, University of Nebraska.
- James C. Edwards et al. v. City of Evanston et al. (1975) and Frank J. Macchiarola et al. v. New York City Board of Examiners et al. (1981), Civil Action No. 74C 2686.
- Klimoski, R., & Brinker, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. Personnel Psychology, 40, 243-260.
- KY ST s 156.105 (1990).
- Lemley, R., & Hersey, P. (1991, July). Assessment handbook. Reston, VA: National Association of Secondary School Principals, 11.
- Lester, R. (1991). Construct validity of a National Association of Secondary School Principals' Assessment Center. Unpublished doctoral dissertation, Georgia State University.
- MacKinnon, D. W. (1975, May). An overview of assessment centers (CCL. Tech. Rep. 1). Berkeley: University of California, Center for Creative Leadership.
- Martin, W. J., & Willower, D. J. (1981). The managerial behavior of high school principals. Educational Administration Quarterly, 17(1), 69-90.
- McIntyre, K. E. (1966). Selection of educational administrators. A University Council for Educational Administration (UCEA) Position Paper, Columbus, OH.
- Melton, J., & Miller, R. (1987, February). Assessment project makes major milestone. The School Administrator. Arlington, VA: American Association of School Administrators, 20.

- Merino, A. (1973). The development of an assessment center for the selection of school administrators. Unpublished doctoral dissertation, University of Massachusetts.
- Micek, T. J. (1983). Teacher perceptions of performance of the NASSP assessment center behavior dimensions as it relates to the conceptual complexity of principals. Unpublished doctoral dissertation, University of Nebraska.
- Mitchel, J. (1975). Assessment center validity: A longitudinal study. Journal of Applied Psychology, 60(5), 573-579.
- Mo. Rev. Stat. #168.405 (1985).
- Nathan, B., & Alexander, R. (1988). A comparison of criteria for test validation: A meta-analytic investigation. Personnel Psychology, 41, 517-535.
- National Association of Secondary School Principals. (1988). NASSP Assessment Handbook. Reston, VA.
- National Commission on Excellence in Education. (1983). A nation at risk: The imperative for education reform. Washington, DC: U.S. Government Printing Office.
- Neuman, G., Edwards, J., & Raju, N. (1989). Organizational development interventions: A meta-analysis of their effects on satisfaction and other attitudes. Personnel Psychology, 42(3), 461-491.
- Ogawa, R. T., & Oxaal, I. (1986, April). A reexamination of NASSP assessment center ratings: research notes. Paper presented to the Annual Meeting of the American Educational Research Association, San Francisco.
- Pearlman, K., Schmidt, F., & Hunter, J. (1980). Validity generalization results for tests used to produce job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.
- Peters, S., & Bagenstos, N. (1988, March). State-mandated principal evaluation: A report on current practice. (Report No. EA 292 889). New Orleans, LA: American Educational Research Association, 12.

- Pitner, N. J., & Russell, J. S. (1985-86). Structured observation of school administration work activities: Methodological limitations and recommendations for research, part I. Educational Research Quarterly, 10(2), 13-24.
- Pokorny, N. G. (1985). Techniques used in the selection of public school principals in the United States. Unpublished doctoral dissertation, University of Colorado.
- Reighard, C. H. (1977). A procedural model for selecting elementary school principals utilizing the assessment center technique. Unpublished doctoral dissertation, Miami University.
- Rolland, L. M. Y. (1990). A study of the Minnesota Principal Assessment Center and its impact on career opportunities for its participants. Unpublished doctoral dissertation, The University of Minnesota.
- Rosenthal, R. (1987). Judgment studies: Design, analysis, and meta-analysis. New York: Cambridge University Press.
- Rosenthal, R. (1989). Meta-analytic procedures for social research. Applied Social Research Method Series. Newbury Park, CA: Sage Publications.
- Rosenthal, R., & Rosnow, R. (1975). Primer of methods for the behavioral sciences. New York: John Wiley & Sons.
- S.C. Code Ann., ch. 24, 59-24-10 (1984).
- Schmitt, N., & Cohen, S. A. (1990). Criterion-related validity of the NASSP assessment center. East Lansing, MI: Michigan State University, Department of Psychology.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.
- Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP assessment center. East Lansing, MI: Michigan State University, Department of Psychology.

- Schmitt, N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the NASSP assessment center. Reston, VA: NASSP Assessment Center.
- Scholl, S. S. (1980). Criteria for selection of Ohio middle school principals with particular attention to the leadership role in curriculum development. Unpublished doctoral dissertation, Miami University.
- Schrup, M. G. (1989). Role of reflection on administrative job performance following National Association of Secondary School Principals' assessor training and assessment center service. Unpublished doctoral dissertation, Indiana University.
- Sergiovanni, T. J. (1987). The principalship: A reflective practice perspective. Boston: Allyn & Bacon.
- Snedecor, G. W., & Cochran, W. G. (1980). Statistical methods. Ames, IA: Iowa State University Press.
- Southern Regional Education Board. (1986). Effective school principals. A report to the Southern Regional Education Board by Its Commission for Educational Quality. Atlanta, GA: Author.
- Storlien, H. T. (1983). Recruitment and selection of secondary principals in Alberta: An analysis of procedures, criteria, and identified problems. Unpublished doctoral dissertation, University of Montana.
- Streitman, W. H. (1975). The use of simulation techniques to identify potentially effective educational administrators. Unpublished doctoral dissertation, Georgia State University.
- Thomas, T. (1983). The testing of teachers' legal parameters and procedures, school law-update 1982, 189-202.
- Thornton, G., & Byham, W. (1982). Assessment centers and managerial performance. New York: Academic Press.
- Thornton, G. C., III, & Byham, W. C. (1982). Assessment centers and managerial performance. New York: Academic Press.
- Trevetto, S. K. (1990). An analysis of school system and participant use of assessment center for professional growth and development. Unpublished doctoral dissertation, University of Maryland.

- Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of New Jersey. Trenton: New Jersey School Boards Association.
- Van Newkirk, C. J. (1984). A study of the assessment center as a process for identifying prospective school principals and assistant principals. Unpublished doctoral dissertation, College of William and Mary.
- Voit, L. B. (1989). Perceptive and evaluation of university principal preparation programs by Michigan Public School Principals. Unpublished doctoral dissertation, Western Michigan University.
- Walden, D. L. (1985). The effects of the NASSP Assessment Center process on self-improvement of assessees. Unpublished doctoral dissertation, University of Colorado.
- Wendel, F. C., Gappa, L. R., & Yusten, C. R. (1988). Use of the assessment center method in administrator preparation programs. Tempe: Arizona State University, The University Council for Educational Administration.
- Wendel, F., & Sybouts, W. (1988). Assessment center methods in educational administration: Past, present, and future. UCEA Monograph Series. Tempe, AZ: University Council for Educational Administration. (Pub. No. 85287-2211)
- Witters-Churchill, L. J. (1988). University preparation of the school administrator: Evaluations by Texas principals. Unpublished doctoral dissertation, Texas A & M University.
- Wolf, F. M. (1987). Meta-analysis, quantitative methods for research synthesis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-059. Beverly Hills: Sage Publication.
- Zubay, A. H. (1976). The use of the assessment center to identify potentially effective educational administrators. Unpublished doctoral dissertation, Georgia State University.

Vita

Linda Dykes Pantili
406 East Drive
Princeton, WV 24740
304-425-0749

Education:

- Ph.D. 1992 Doctor of Philosophy in Educational Administration,
Virginia Polytechnic Institute and State University,
Blacksburg, Virginia
- 1992 Certificate of Advanced Studies, Virginia Tech,
Blacksburg, Virginia
- 1992 Certified as National 4MAT Trainer and Implementation
Leader, Excel, Inc., Barrington, Illinois
- 1988 Licensed Professional Counselor, West Virginia
- 1983 National Certified Counselor
- 1974 Endorsement in Administration
- M.A. 1972 Marshall University, Huntington, West Virginia Guidance
and Counseling
- B.S. 1968 Concord College, Athens, West Virginia
Teacher Education
Comprehensive Social Studies
Elementary Certification

Professional Experience:

- 1992-Present Vice-President, Associated Development Services,
Athens, Ohio, 45701

| | |
|------------------|--|
| 1990-1992 | Assistant Editor <u>Journal of Educational Finance</u> Virginia Tech Blacksburg, Virginia |
| 1979-1990 | Counselor, Oakvale High School Oakvale, WV 24739 Grades 7-12 |
| 1968-1979 | Counselor/Teacher McDowell County, WV |

Recognition:

Recipient of 1991-92 Educational Administration Alumni Scholarship

**National Secondary School Counselor of the Year, 1987-88
American School Counselor Association**

**Anne H. Maynard Award, 1986-87
West Virginia School Counselor Association**

U. S. Air Force Dedication and Devotion Award, 1987

Nominated as Teacher of the Year, 1985

**Outstanding Service in Guidance and Counseling in WV, 1983
West Virginia Personnel and Guidance Association**

Educational Activities:

**January, 1992: NASSP Assessor, Southwest Regional Assessment
Center, Blacksburg, Virginia**

**February, 1992: Assessor Trainer, NASSP Assessment Center, Ft.
Worth, Texas**

**February, 1992: Representative, Virginia Tech, NASSP National
Conference, San Francisco, California**

- September, 1992: Presenter, NASSP Assessment Center Directors' Meeting, Williamsburg, Va.
- November, 1992: Participant, Ohio School Boards Association Conference, Columbus, Ohio.
- March, 1991: Representative, Virginia Tech: Norham Centre for Leadership Studies, University of Oxford Department of Educational Studies, Oxford, England
- April, 1991: Presenter, American Educational Research Association, Chicago, Illinois
- May, 1991: NASSP Assessor, West Virginia Department of Education, Bridgeport, West Virginia
- May, 1991: Coaching and Mentoring Training, NASSP, Virginia Beach, Virginia
- May, 1991: NASSP Assessment Center Directors Training, Louisville, Kentucky
- June, 1991: NASSP Assessor, South Carolina Department of Education, Columbia, South Carolina
- July, 1991: Coordinator of NASSP Coaching and Mentoring Training for Virginia Tech, Blacksburg, Virginia
- September, 1991: Proof reader, Dr. David Alexander, Public School Law, Virginia Tech, Blacksburg, Virginia.
- November, 1991: Project Evaluator, Drug Free Schools Program, McDowell County, West Virginia
- December, 1991: Educational Administration Strategic Planning Department Meeting, Virginia Tech, Virginia Beach, Virginia
- October, 1990: National Association of Secondary School Principals Assessment Center; Assessor Training
- December, 1990: Educational Administration Strategic Planning Department Meeting, Virginia Tech, Staunton, Virginia

Publications:

Pantili, Linda, Williams, John and Fortune, Jim. (May, 1991).
Principal Assessment: Effective or Not? A Meta-Analytic Model,
ERIC Document No. EA023 031.

Williams, John, Fortune, Jim, and Pantili, Linda. (April, 1992).
Operationalization of Leadership by Black and White Leaders:
Implications for Validity. Presented to the American
Management Association, Las Vegas.

Pantili, Linda, Williams, John and Fortune, Jim. (July, 1992). A
Meta-Analytic Model Principal Assessment, Journal of School
Leadership, 2(3), 256-279.

Professional Memberships:

American School Counselor Association
American Association for Counseling and Development
Association for Supervision and Curriculum Development
West Virginia School Counselor Association
West Virginia Association for Counseling and Development



Linda D. Pantili