

AN ABSTRACT OF THE THESIS OF

JOHN GAMBON for the MASTER OF SCIENCE

(name of student)

(degree)

in PSYCHOLOGY presented on April 19, 1991
(major) (date)

Title: A Validation Study of the Kansas Civil Service
Exam for Office Assistant, Level 2

Abstract approved: *James R. K...*

The use of testing in the selection of employees is a common practice. However, such testing must meet certain requirements set forth by the Uniform Guidelines for Employee Selection Procedures (1978). These guidelines require that all selection tests be periodically validated to ensure the test is a good predictor of how well someone will perform on the job.

The purpose of this study was to validate a state civil service exam. A concurrent validity study was conducted using the selection test (Civil service exam), and the level of education as predictors and employee's initial performance appraisal as the dependent variable.

Two regression analysis were conducted to determine: 1) the predictive ability of the Civil service exam on employee performance and 2) the predictive accuracy of the selection device and education on

employee's performance. Results of the analysis revealed that neither the civil service exam or the level of education served as accurate predictors of employee performance. This researcher concluded that the State of Kansas should examine its selection process in order to determine if the discrepancies were due to the selection test, the performance appraisal, or both.

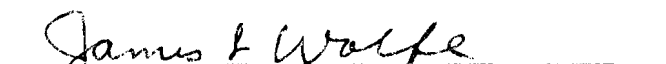
A VALIDATION STUDY OF
THE KANSAS CIVIL SERVICE EXAM
FOR OFFICE ASSISTANT, LEVEL 2

A Thesis Presented to
The Division of Psychology and Special Education
EMPORIA STATE UNIVERSITY

In Partial Fulfillment
of the Requirements for the Masters Degree
in Psychology

by
John P. Gambon
April 1991


Approved for the Major Division


Approved for the Graduate Council

ACKNOWLEDGEMENTS

I would like to extend my sincere appreciation to all those who helped make this thesis a reality. In particular, I appreciate the assistance of Dr. Tompkins in beginning me down this long road. I also appreciate the patience of Mr. Gary Shikels, Director of Personnel Services for the State of Kansas, who took time out of his schedule to periodically answer questions over the phone.

My sincerest appreciation also goes out to Nicole-Norian Baine and the personnel directors of the various Kansas universities who agreed to gather the data for me. Without their help, this thesis would never have evolved.

To my committee members, Dr. Dungan and Dr. Hawes. Their comments and insights were most beneficial, and I greatly respect their opinions.

To Dr. Jim Parks, I owe a special gratitude of thanks. Your assistance and perseverance cannot be measured. I appreciate everything you have done.

Finally, I would like to thank my family, who have been most supportive throughout my career. Thankyou...

TABLE OF CONTENTS

	Page
ABSTRACT.	i
ACKNOWLEDGEMENTS.	v
LIST OF TABLESviii
CHAPTER	
I. Introduction	1
Selected Characteristics	2
Basic Educational Requirement.	2
Employment Testing	4
Interviewing	6
Performance Evaluations.	9
Statement of the Problem	12
Hypothesis	14
Definition of Terms.	14
Statement of Significance.	17
Summary.	18
II. Method	19
Subjects	19
Design	20
Substantive Hypothesis	21
Steps and Procedures	21
Apparatus.	22
Statistical Design	25
Summary.	26
III. Results.	28
IV. Discussion	31

REFERENCES. 38

LIST OF TABLES

Table	Page
1. Means and Standard Deviations for the Selection Test Score, the Performance Score, and the Level of Education Score	30
2. Correlation Table for the Selection Test Score and Performance Score.	30
3. Correlation Table for the Selection Test Score, the Performance Score, and the Level of Education Score.	31
4. Multiple Regression Table for the Selection Test Score, and the Level of Education Score Dependent Variable, Performance Score.	31

Chapter I

Introduction

According to Anastasi (1988), methods of hiring can be traced all the way back to the Chinese Empire and their use of tests as selection devices. Today, there are still many methods of hiring in use. Some of the more common examples include the application blank and/or resume, one or more interviews, a fulfillment of basic educational or vocational requirements and possibly a series of tests. Different types of tests used in hiring include paper and pencil tests, stress tests, and problem solving tests.

Nicole Norian-Baine, Director of Personnel at Emporia State University stated (personal communication, October 1990) that three factors are used in the hiring process of Office Assistant, Level 2 (OA2). These factors include: (a) the validity and reliability of basic educational and/or vocational requirements, (b) content and criterion-related validity of the test being used, and (c) validity and reliability in the interviewing process. This researcher will analyze how these three factors are related to the subject's performance appraisal. Additionally, this researcher will examine how the validity of these procedures might generalize to similar hiring methods. Finally, the relevant court standards pertaining to such selection procedures will be discussed.

SELECTED CHARACTERISTICS

BASIC EDUCATIONAL REQUIREMENTS

For years, employers have been concerned about the basic educational and/or vocational requirements that are required for a job. Owens (1976) stated that the best predictor of what individuals will do in the future is what they have done in the past. Still today, society directly correlates a person's level of education with ability, status, and prestige. While a minimum level of education is required to function in society, Merritt-Haston and Wexley (1983) made it clear that educational achievement might merely be an indication of an individual's ability to read, write, and understand. Furthermore, possession of a degree may not guarantee that the individual has absorbed these basic skills. Educational requirements fall under the jurisdiction of Title VII of the 1964 Civil Rights Act (Twomey, 1990) and under the Uniform Guidelines on Employee Selection Procedures (1978).

To determine the legality of educational requirements, Merritt-Haston and Wexley (1983) reviewed a study of 83 Equal Employment Opportunity (EEO) cases and subsequent decisions to determine if education requirements were justified. One author examined the cases individually to determine what level of education was required for each particular EEO court hearing.

The categories were as follows: (a) Grade 10 education, (b) High school diploma, (c) apprenticeship (education provided by a union), (d) some college, (e) bachelor's degree, (f) master's degree, and (g) Ph.D. The results of the hearings were compared by determining the percentage of cases won by the defendant versus those won by the plaintiff in the respective categories.

These authors concluded from the results that there is a strong relationship between the level of education and the outcome of the case. The higher the level of education required, the greater the probability that the defendant would prevail in the case (i.e. the company or organization). Merritt-Haston and Wexley (1983) based these results on the idea that more complex jobs usually require a higher level of education and therefore, are less likely to be challenged and won by the plaintiff.

If a high correlation exists between job complexity and level of education, then all major companies should closely consider what amount of education is required for a successful employee. In addition to major companies, local, state, and national governments should also consider examining adequate levels of education for specific and sometimes complex jobs. With this in mind, the state of Kansas would be correct in requiring a high school diploma as a base level of education for the position of OA2.

In summary, basic educational and vocational requirements, while important, may be considered a questionable predictor of a person's true ability. The one exception to this is complex jobs requiring higher levels of education. The court system tends to agree with this point.

EMPLOYMENT TESTING

According to Ebel (1983), personnel testing is something fairly new for industry in the United States. Not until this century did employers begin to test applicants for ability and aptitude. Ghiselli, as cited in Landy (1989), stated personnel testing increased rapidly due to World War I, the Great Depression, and World War II. Furthermore, this flood of testing went unchecked until the early 1960's. During the 1960's, critics of testing began to voice their dissent claiming personnel tests were an invasion of privacy. They maintained that questions concerning a person's religion or nationality were not job relevant and should not be asked. Later, with the advent of Title VII of the Civil Rights Act of 1964, personnel testing began to make an astonishing comeback (Lee, 1988). This legislation created standards for which employers could build employment tests on. In essence, the federal government had set the rules for what personnel tests could and could not require.

Most of the larger corporations in the United States use some type of test when hiring. However, as with any test, the employer must provide a valid reason as to why a particular test is used.

More specifically, the government requires valid reasons for test use by a company. In 1978, the U.S. government created The Uniform Guidelines on Employee Selection Procedures. These guidelines require all companies to show both content validity in their selection tests and criterion-related validity when generalizing the results over to both the job and the evaluation of the employee in that job. The Uniform Guidelines state that a content validity study "...should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated" (p. 38296). Furthermore, a criterion-related validity study "...should consist of empirical data demonstrating that the selection procedure is predictive or significantly correlated with important elements of job performance" (p. 38298).

Another document used by both the U.S. government and private business is the Principles for the Validation and Use of Personnel Selection Procedures: Second Edition. This document was created by the Executive Committee of the Division of Industrial-Organization Psychology of the

American Psychological Association. The document gives both conventional and statistical recommendations for test design, implementation and validation.

Numerous studies have been conducted on personnel tests demonstrating their validity (Cope, 1982; Sackett, 1989; Safrit & Wood, 1981). Ghiselli, as cited in Reilly and Chao (1982), summarized the results of hundreds of criterion-related validation studies and found a mean validity coefficient of .45. Studies conducted by Schmidt and Hunter (1977) and Pearlman, Schmidt, and Hunter (1980) offer stronger support for test validity generalization. The results of these studies demonstrated, that even with the many restraints and requirements placed on corporations concerning validity in testing, employment tests are thriving. However, validation of testing procedures is not something that can be done once and left alone. According to Anastasi (1961), test validation is an ongoing process that only stops when the use of the test stops.

In summary, personnel testing is a widely used procedure for the process of hiring employees but it is not without regulations (i.e. Title VII of the 1964 Civil Rights Act, the Uniform Guidelines on Employee Selection Procedures and the Principle for the Validation and Use of Personnel Selection Procedures: Second Edition). The U.S. government requires all personnel tests to possess both content

validity and criterion-related validity. For employers, maintaining validity will always be an ongoing process.

INTERVIEWING

Industrial and Organizational (I/O) psychologists have been studying the use of the interview as a selection device for over 60 years. I/O psychologists have concluded that interview effectiveness has mixed results. The first comprehensive review of the research with interviews was published by Wagner in 1949. Wagner found that the earliest industrial research on the effectiveness of the interview began with Scott in 1915. Scott reported low reliability between the evaluations given to six personnel managers who had interviewed 36 sales applicants. Wagner noted that of the 106 articles he found dealing with the interview, only 25 contained empirically based findings. More specifically, these 25 articles assessed reliability by correlating evaluations of different interviewers who had assessed the same job applicants. Validity was assessed by correlating interview decisions with some criteria of the job. Wagner found that reliability correlations ranged from .23 to .97, with a median $r = .57$. Furthermore, Wagner found that validity coefficients ranged from .09 to .94 with a median $r = .27$. Neither median values for reliability or validity were considered particularly high.

Years later, Mayfield (1964) updated Wagner's findings. Mayfield noted that there had been little change in the reliability and validity coefficients for the employment interview. As a results of his follow up study, Mayfield offered the following observations: (a) Generally, unstructured interviews have low reliability while structured interviews results in normally higher inter-rater reliability; (b) Interview validity is usually low; and (c) Even when the interviewer has valid test information available to use along with the interview, the level of validity and prediction of the applicant's job success is usually no better and sometimes worse. Other research tends to substantiate both Mayfield's and Wagner's findings (Schmitt, 1976; Ulrich & Trumbo, 1965; Wright, 1969).

Recent research has had a more promising outlook toward the interview in terms of validity and reliability. One method of increasing reliability and validity has been to have a board or panel interview candidates. Landy (1976) reported favorable results when a board of interviewers selected police officers. The board interviewed a total of 399 applicants and hired 150. The factors that the board used in evaluating the applicants were called predictor variables. Furthermore, an analysis of supervisor ratings of performance identified four performance factors: (a) Professional maturity, (b) Technical competence,

(c) Demeanor, and (d) Communications. A validity analysis indicated that rated performance could be predicted from the predictor variables used by the board but the overall performance rating, from the board, could not. Due to restrictions of range the validity coefficients were small (.26, .29, .33 and .34), however, Landy still feels they are significant though not overwhelming.

Similar research on board interview methods has been done with testing and hiring procedures used by the Civil Service. Anstey (1977) did a 30 year follow-up on the British Civil Service Selection Board procedure. Using ranks from 30 years earlier as a criterion measure, Anstey analyzed a total of 301 employees and found a validity coefficient of .35. After correcting for restriction of range, the coefficient increased to $\underline{r} = .66$.

In another study, Reynolds (1979) investigated the inter-rater reliability of oral interviews used by the Louisiana State Civil Service Department. A three member panel was used to interview 67 job applicants. Reliability between the panel ranged from .78 to .85 with an overall composite of $\underline{r} = .80$.

In summary, certain methods of interviewing such as the structured interview are promising because of their higher levels of validity and reliability. Also the use of panel or board members can increase validity and reliability.

PERFORMANCE EVALUATIONS

Performance evaluations have been conducted as long as organized paying jobs have been in existence. Initially, performance evaluations were nothing more than an employee's supervisor offering them a pay increase. Such an increase was based on the supervisor's observations and feelings about the employee and their work. It was of little consequence if the supervisor was unable to observe the employee's work or simply did not like them. An employee's pay would very easily reflect these inadequacies. In more recent times, the U.S. government has stepped in and regulated the practice of evaluating employees. The initial concern of government was to control the discrimination against minorities. The Civil Rights Act of 1964 is a prime example of their effort. Minorities, as defined by The Civil Rights Act initially included a person's race, color, sex, national origin, or religion and later was amended to include age, handicap, and veteran status. In essence, those people not considered a protected minority were white males. Furthermore, Title VII of this act requires that those initial predictors of job ability (i.e., basic educational/vocational requirements, personnel tests and interviews) correlate with the evaluation criteria found on the performance appraisal.

There are many different methods and forms of evaluating employees and their performance. Some of the

more common methods include: Graphic Rating Scales, Forced Choice Method, Critical Incidents Method, Checklist and Weighted Checklist Method, Paired Comparison Method and Descriptive Essays. Also, there are different types of forms used for evaluation. They include: Behaviorally Anchored Rating Scales (BARS), Behavioral Observation Scales (BOS), and Management by Objective (MBO) (DuBrin, 1981).

Though some of these methods are more valid than others, the Federal Government only requires that the performance appraisal meet certain criteria. The Uniform Guidelines on Employee Selection Procedures states that all methods of evaluating personnel must contain some form of criterion-related validity. That is, the factors that the employee is evaluated on must be relevant to both the job and the selection procedure. Failure to periodically validate these procedures would put the company in violation of the Uniform Guidelines. Furthermore, companies may find that either their selection procedure or the evaluation method is no longer relevant to the job being done. For instance, if the position for secretary requires you to take a typing test, but typing is no longer a part of the job, then the typing test is an invalid predictor of performance. Furthermore, if employees are evaluated on their typing skills (when typing is no longer required), then the evaluation criteria is invalid.

When this happens, personnel offices are left with the task of either revising their procedures or creating new ones. Though creating a new method of selecting and evaluating employees may not be too difficult, defending it can be a different matter. Kleiman and Faley (1985) stated court judges are usually reluctant to accept new procedures and subsequent research findings if they are inconsistent with those in the Uniform Guidelines.

In summary, performance evaluations of some form have been used as long as regular paying jobs have been in existence. While initial evaluations were done through supervisor observations, the U.S. government now requires evaluations to reflect both the aspects of the job and the selection procedure. Therefore, selection procedures and performance evaluations must contain criterion-related validity as set forth by the Uniform Guidelines on Employee Selection Procedures. Any method or procedure not meeting the guideline's standards has usually been found unacceptable by the courts.

THE PROBLEM

The state of Kansas requires that all applicants applying for clerical positions complete a three-step process before being hired. Step one requires all employees to meet basic educational requirements (i.e., a high school diploma or its equivalent). Step two requires the applicant

take the Kansas State Civil Service Exam and pass with a score of 70 or better. Step three is a one-on-one interview between the employer and the applicant. Applicants must be able to successfully complete all three levels before being hired. An inability to meet the requirements of any of the three steps terminates the applicant's chance for job selection. For example, any applicant unable to fulfill the basic requirements of education will not be considered for the position of OA2. They will be unable to take the Kansas Civil Service Exam or go through the interviewing process.

There are numerous problems with the selection devices utilized by the State of Kansas. Specifically, Gary Shikels, Director of Test Administration and Validation for the state of Kansas reported that the Kansas Civil Service Exam for clerical workers has not been validated since 1985 (personal communication, November 1990). Cascio suggests a 'periodic audit' of selection procedures. It seems evident that a periodic audit has not been conducted and that a revalidation of the Civil Service Exam is in order.

Additionally, Nicole Norian-Baine, Director of Personnel at Emporia State University (Emporia, KS), has stated that no validation study has been done in the past five years on either the effectiveness of the interviewing process or the basic requirements needed for the position of OA2 (personal communication, October 1990). A validation study of both the interviewing process and the basic

education/vocational requirement is as equally essential as a validation of the civil service exam. This research, however, will focus on the exam and the strength of the relation it has with the Kansas performance appraisal used with OA2s.

An additional factor to be studied is the level of education completed by an applicant when they begin their work as an OA2. This study will also determine if there is a significant relationship between either an applicant's level of education and their selection test score or an applicant's level of education and the results of their initial performance appraisal.

STATEMENT OF HYPOTHESIS

The general education requirement and the administration and scoring of the Kansas civil service exam are both valid predictors of an individual's performance.

DEFINITION OF TERMS

Validity. The extent to which a measurement procedure actually measures what it is designed to measure (Cascio, 1987). For example, a math test given in an English class would not be a good device to measure a student's knowledge of English whereas the same math test might be an excellent measuring device in a math class.

have criterion-related validity because the criterion on the selection device would match the criterion on the performance appraisal. If the selection device only measured typing ability and the performance appraisal was only concerned with evaluating filing skills, the criterion would not match and either the selection test or the performance appraisal is invalid.

Title VII of 1964 Civil Rights Act. An employer cannot discriminate against an individual on the grounds of race, color, sex, national origin, or religion. Title VII has been amended to now include the following: age, handicap and veteran status.

Uniform Guidelines on Employee Selection Procedures. U.S. government guidelines designed in 1978 to provide a framework for the proper, nondiscriminatory use of tests and other selection procedures. Cascio asserts that the Supreme Court gives great deference to these guidelines.

Principles for the Validation and Use of Personnel Selection Procedures: Second Edition. A statement of principles adopted by the division of Industrial-Organizational Psychology of the American Psychological Association. Its purpose is to specify principles of acceptable practice in the choice, development, and evaluation of personnel selection procedures.

Basic Educational Requirements. Those requirements, either educational or vocational, that are used as a basis

for employment...(Merritt-Haston & Wexley, 1983). For example, the educational requirement of a university professor might be that he/she possesses a Ph.D or similar degree.

Employment Testing. Any device that identifies those individuals who will best fit into a particular job opening (Landy, 1989). For example, a taxicab driver must be able to effectively drive a car to do his/her job. Therefore, a driving test would be an employment selection device.

Performance Evaluation. Cascio (1987) defined performance evaluation (or appraisal) as "the systematic description of individual job-relevant strengths and weaknesses" (p. 73). A performance evaluation can be something as simple as telling a person how they did on a particular task. Usually, performance evaluations are done periodically to let an individual know how they are performing at their job.

Statement of Significance

The findings of this study will effect the state of Kansas in one of two ways. (a) The results will confirm the effectiveness of the Kansas Civil service exam and the performance appraisal used for selecting and evaluating OA2s. If so then this researcher would suggest that the state consider postponing the need for its own validation of both the civil service exam and the performance appraisal.

(b) The results will invalidate the effectiveness of the Kansas Civil service exam and the performance appraisal used by OA2s. If this occurs, this researcher would suggest that the Division of personnel services consider running a statewide job analysis on the position of OA2 to determine if the discrepancy lies in the civil service test or the performance appraisal. The state may also consider analyzing other positions that require the civil service exam as a factor in hiring applicants.

Summary

Clearly there is a need for an analysis of the civil service test scores and performance appraisals of OA2's. This includes the methods used to gather the data, statistically analyze it and present it in such a form that it will be both useful, understandable and replicable. The remainder of the thesis will specifically clarify the following: (a) the step-by-step method to be used to gather the data, or scores, including a summary on sampling procedures and levels of ecological validity; (b) the design used to categorize those scores; (c) the statistical method which will be used to analyze the data and the internal validity that this method possesses; and (d) an overall summary of the study.

Chapter II

METHODS

Subjects. The subjects for this study were 138 men and women who are or were employed on four university campuses in the state of Kansas. Subjects for this experiment met the following criteria:

- (a) All subjects should be previously or currently employed as Office Assistants, Level 2 (OA2s).
- (b) The subjects were all hired no earlier than 1985.
- (c) Subjects had taken the Kansas Civil Service Exam for OA2s.
- (d) Subjects had received at least one performance appraisal since hiring.

Selection of the subjects was left to the personnel department. Each department was given the above specifications in helping to select the sample. While a larger number of OA2s could be found in other state offices, the type of work required may vary. Therefore, a chance of a larger sample has been sacrificed at the expense of more control of those duties required of an OA2 in a university setting.

Another factor to consider was the restriction of range for both test scores and performance evaluations. Only scores and evaluations of employees who received the positions of OA2 were used. Those applicants who were .

rejected for the position could not be considered since they did not possess evaluation scores. Therefore, since only the applicants with the highest potential were hired, one would expect the performance evaluations to be higher and the overall correlation (r value) to be attenuated. (Cohen & Cohen, 1975)

Since the size of the sample was determined by each personnel department, the sample was not random. Therefore, the amount of external validity (the degree to which the results can be generalized to other groups of individuals) was limited to OA2s working in a university setting in the state of Kansas.

Design. The design in this study was ex-post facto. Kerlinger (1986) defines ex post facto research as:

...systematic inquiry in which the scientist does not have direct control of independent variables because their manifestations have already occurred or because they are inherently not manipulable. Inference about relations among variables are made, without direct intervention, from concomitant variation of independent and dependent variables. (p.348)

The independent variables in this study were the test scores received by the employees on the civil service exam and their education level. The dependent variable was the

scores each employee received on their performance evaluation.

Substantive Hypothesis.

1. There will be no correlation significantly different from zero between subject's civil service exam scores and their performance appraisal score.
2. There will be no correlation significantly different from zero between the education level of OA2s and their civil service exam scores.
3. There will be no correlation significantly different from zero between the education level of OA2s and their performance appraisal scores.
4. There will be no correlation significantly different from zero among any combination for the following variables: subject's civil service exam scores and the level of education of OA2s with the dependent variable performance appraisal scores.

Steps and Procedures. University personnel departments were contacted by phone. Those willing to cooperate provided the following data: (a) The last four digits of the subject's Social Security number (This was used to keep scores and performance appraisals with the appropriate individual), (b) The subject's test score on the Kansas Civil Service Exam for OA2s, (c) The subject's initial performance appraisal, and (d) The subject's level of education.

The level of education fell into one of six categories:

1. High School Diploma/GED
2. Vocational/Technical School
3. Some College/University
4. AA Degree
5. BA, BS or other Bachelor's Degree
6. MA, MS or other Master's Degree

The data was then entered and analyzed using the NCSS (Number Crunching Statistical System) program available on the Emporia State University campus (Hintz, 1988).

Apparatus. The test designed to measure an applicant's knowledge, skills, and abilities was the Kansas State Civil Service Exam for Office Assistants. It is a paper-and-pencil test made up of 90 multiple choice and True/False questions. The categories found on the test include the following: Math, filing, spelling, grammar and name & number matching. There is no ladder of difficulty to the test. That is, the last questions are no more difficult than the first. Each person taking the test has two hours to complete it. A percentage score of 70 or above is required to pass the test. The OA2 test was administered and supervised by G. Shikels, director of the State Division of Personnel Services office in Topeka Kansas.

Reliability of the test was determined using the split-half method. The split-half method breaks a test into 2

equal parts usually by separating the questions into one group with all the odd numbered questions and a second group with all the even numbered questions. These two smaller tests are administered and scored. The internal consistency of the Kansas Civil service exam, for OA2's was $r = .82$.

The performance appraisal used by OA2s is similar to what is known as a Management By Objectives (MBO). This method, according to Macdonald (1982), allows the supervisor and the employee to sit down and methodically work out a set of objectives that both the employee and the supervisor can agree to. This step alone may require several revisions. When both employee and supervisor agree on the objectives, the employee begins to implement the program. The employee receives periodic progress reports to determine if the objectives are being met. Upon completion of the specified period, the objectives are again reviewed to see if they were or were not fulfilled and why.

While the evaluation form used on the Kansas university campuses with OA2s is similar, the objectives are not necessarily agreed upon as they are in an MBO program. With this evaluation form, the supervisor writes down, in their own words, specifically what is required of the employee. Furthermore, the supervisor gives each required duty a "percentage" which indicates how important that requirement is. For example, a typist may have only two requirements on his/her evaluation: typing ability and filing. If the

majority of her job is typing, than the "percentage value" assigned to that task may be 70, 80, or 90%. The filing task may be seen as trivial and therefore assigned a low percentage value 10, 20, or 30%.

Studies on the MBO method of appraising employees has revealed serious flaws. According to Leonard (1986), nearly half of the firms that make up the Fortune 500 use the MBO program, however only 20-25% are considered successful. A study by Richards (1986), supported Leonard's findings. Richards stated that after 20 years of MBO programs, they are still not applied efficiently, effectively or easily as they could be.

Each OA2 is evaluated according to the objectives stated on their performance appraisal. New employees, on probationary status, are evaluated after three months. Other employees are evaluated every six months or annually depending on how long they've worked with the state. A cumulative score is totalled and applied to the following scale: 0-250 = Unsatisfactory; 251-450 = Satisfactory; 451-500 = Exceptional. This scale gives the employee their overall rating.

The rating method was designed by the Kansas Department of Personnel Services. The scale is weighted towards ratings of unsatisfactory and satisfactory. This is evident by the 250 point range for Unsatisfactory scores as opposed to 199 point range for Satisfactory scores and a 49 point

range for Exceptional scores. No reason was given by the Division for personnel services explaining why this method is used or how the 1, 2, 3 scale was developed.

Statistical Design. A regression analysis was used to study the data. The analysis consisted of the test scores of the OA2 employees, their level of education, and the overall rating received on their first performance appraisal. Since the sample size of the various universities ranged from 6 subjects to 50, subject scores from the various universities were analyzed cumulatively rather than separately.

Cascio (1987) lists 3 requirements that must be met to effectively conduct a criterion-related validation study using correlational analysis:

- (a) Criterion measures must be relevant and valid.
- (b) Criteria must be reliable.
- (c) Be cautious of possible criterion contamination.

Controlling for these three requirements was conducted as follows. Criterion measures can only be demonstrated as valid by an empirical analysis. This was accomplished in the content validity study conducted by Mr. Gary Shikels and the Division for personnel services. In a phone interview with G. Shikels (personal communication, November 28, 1990), he reported that the last validation and reliability study of the Kansas Civil Service exam was conducted in 1985. The

validation study was composed of a content validity analysis. Job incumbents and job superiors were asked to examine the test items to determine if the questions were representative of the knowledge, skills, and abilities requirements for the successful execution of the job. Both job incumbents and job supervisors were in agreement. The test questions represented the aspects of the OA2 position.

With regard to reliability of the criteria, this is typically done by giving supervisors training regarding employee evaluation. However, according to Gary Shikels, no periodic training has been administered to the various supervisors.

Criterion contamination can be controlled by preventing the evaluators knowledge of what the employees selection score was on their initial test. According to Mr. Shikels, this threat has been controlled for in that supervisors are not allowed to examine employee test scores thereby eliminating any chance of contamination.

Summary. Validation of different types of selection procedures is not only necessary but essential for companies for the selection of competent employees. While many methods of selection are in use, the State of Kansas relies on a three step procedure of: fulfilling basic educational requirements, passing a civil service test, and meeting the requirements of an interview to be considered for positions as office assistants. This research analyzed the

relationship between the Kansas civil service exam, the level of education of the applicants upon hiring, and the subsequent performance appraisal of applicants who were awarded the position of office assistant. Through regression analysis, this researcher was able to determine the predictive accuracy of employees scores on the civil service exam, level of education and their subsequent performance appraisals.

Chapter III

Results

A simple regression analysis was calculated between the selection test and the performance appraisal. A total of 138 scores met the requirement of both a test score and performance appraisal score. Means of 87.55, 2.18, and 5.08 were found for the selection test score, performance appraisal rating, and education level respectively (Table 1). A non-significant correlation of $r = .12$ was observed between the predictor (selection test score) and the dependent variable (employee's first performance appraisal) score (Table 2). A sample size consisting of 138 pairs of scores would require a correlation no less than $.26 (\alpha = .05)$.

A multiple regression was also conducted between the subject's selection test score, their performance appraisal, and their education level. A total of 80 scores met all 3 of these requirements. The analysis revealed that the three predictors did not reliably predict to employee performance. The specific correlations between level of education and either their score on the selection test or the performance appraisal can be found (Table 3). A correlation of $r = .07$ was found between the education level and test score. A negative correlation of $r = -.13$ was established between education level and the subject's performance appraisal.

Table 1

Means and Standard Deviations for the Selection Test Score,
the Performance Score, and the Level of Education Score

	Mean	Standard Deviation
Selection Test Score	87.55	4.97
Performance Score	2.18	0.47
Level of Educ. Score	5.08	1.57

Table 2

Correlation Table for the Selection Test Score and
Performance Score

	Selection Score	Performance Score
Selection Test Score	1.00	0.12
Performance Score		1.00

n = 138

Table 3

Correlation Table for the Selection Test Score, the
Performance Score, and the Level of Education Score

	Selection Score	Performance Score	Educ Level Score
Selection Test Score	1.00	0.01	0.07
Performance Score		1.00	-0.13
Level of Educ. Score			1.00
n = 80			

Table 4

Multiple Regression Table for the Selection Score, and the
Level of Education Score on the Dependent Variable,
Performance Score

Multiple R		.140
R Square		.020
Analysis of Variance		
	<u>df</u>	Mean Square
Regression	2	.173
Residual	77	.223
F = .77		

The original hypothesis stated that there would be no correlation significantly different from zero between the subject's test score, their performance appraisal score and their level of education rating. These results confirm the null hypothesis. No significant correlations were observed.

Furthermore, the original hypothesis stated that there would be no correlation significantly different from zero between any combination of predictor variables with the dependent variable, performance appraisal. These results also confirmed the null hypothesis. No significant correlations were observed.

Chapter IV

Discussion

The purpose of this study was to determine if the Kansas Civil Service Exam was a significant predictor of job performance. Additionally, the multiple predictors of selection test scores and level of education did not reliably predict to the OA2's initial performance appraisal.

The resulting correlations have shown that no significant correlations existed between any of these variables. This would lead one to believe that the civil service exam is a test with little value. However, Landy (1989) indicated that a test with no significant predictive ability is still better than nothing. For example, he cited a series of studies where the per employee cost of administering a selection was compared to the worth of the employee. Employee worth was determined by placing a dollar value on the employee's performance. This value was determined by Landy using expert judges. In nearly all of these studies, the value of the employee to the organization exceeded the cost of the selection test used to hire them. Even when the selection test was statistically shown to be a poor predictor of performance, the worth of the employee still far exceeded the cost of the selection test. In one particular study by Schmidt, Hunter and Gast-Rosenberg (1980), the supervisors of computer programmers working for

the federal government were asked to give a dollar value to the output of the typical computer programmer considering both quantity and quality of their work. In placing a dollar value on this output, supervisors were also asked to take into consideration the cost of hiring an outside firm to do the same work. The question used to obtain this information read as follows: "Based on my experience, I estimate the value to my agency of the average GS 9-11 programmer at _____ dollars per year." Schmidt et al. (1989) considered other factors such as: (a) the test used to hire these programmers (Programmer Aptitude Test) had an estimated validity of .76; (b) the testing cost was approximately \$10 per employee; (c) there were approximately 4,400 government programmers at the time; (d) new hires for this position averaged 618 per year and; (e) programmers remained with the government on average of 9.5 years. With these statistics, the researchers determined that if the government used a selection ratio of .05 (i.e., they hired only 1 applicant out of every 20), these programmers could still secure a cumulative value for their work of \$5.6 million over a 9 year period. Furthermore, if the government used a selection ratio of .80 (i.e., they hired 16 applicants out of every 20), the gain would reach approximately \$97.2 million over the same period as mentioned before.

Mr. Gary Shikels, Director of Personnel Services, was contacted to see if similar data, as stated in the above study, was available for OA2's (personal communication, April 16, 1991). Mr. Shikels stated that the Division of Personnel Services did not keep such information. Therefore, a study of the utility of the civil service exam is not possible at this time.

While the utility of the present civil service exam may enable the personnel manager to discriminate between good and bad applicants, the continued use of this test may result in an increased potential for discrimination against both higher quality applicants and minorities. Such discrimination against minorities could result in an "Adverse impact" lawsuit against the state. Cascio (1987) defines Adverse impact as "...selection in hiring, promotion, or other employment decisions that works to the disadvantage of members of a race, sex, or ethnic group" (p. 27).

One of the disturbing things about adverse impact is that few employers realize they have been discriminating against a protected minority until a lawsuit charging adverse impact has been brought against them. Employers may never realize that their selection test may be culturally biased. In other words, particular minorities may be unable to comprehend certain terms and examples contained in test item questions. A good example of this problem is Adrian

Dove's "Soul Folks Chitlin Test" (1968). The test was designed in Los Angeles in 1968 using the vocabulary of the black, inner-city subculture. While this test is not used for personnel selection, The Chitlin Test has become an excellent example of how cultural bias can invade our tests and tend to unfairly judge people as lacking in intelligence. One of Dove's questions ask:

1. A "hanky head" is a(n)
 - a. cool cat
 - b. Uncle Tom
 - c. porter
 - d. preacher

The correct answer is b. While this test may seem absurd to some, minority applicants may feel the same way concerning the selection tests they are required to take.

To correct this situation, organizations must adopt an Affirmative Action Program. Cascio (1987) defines this program as a set of measures written by the company designed to self-evaluate the company and eliminate employment discrimination. Affirmative action programs may include any one or more of the following:

1. Recruitment of applicants from groups that are underrepresented in the employer's work force.
2. Changing management attitudes: trying to eliminate conscious and unconscious prejudices within the organization.
3. Removing discriminatory obstacles: identifying employment practices that work to the disadvantage

of underrepresented groups and replacing them with acceptable nondiscriminatory practices.

4. Preferential treatment: hiring and staffing preferentially for groups that are underrepresented in the employer's work force.

This program, created and recommended by the Equal Employment Opportunity Commission, is in use in businesses throughout the country.

The non-significant correlations between the level of education and either the selection test or performance appraisal tend to show that an OA2's education has little bearing on their initial selection or their first appraisal. Still, the state of Kansas lists the following educational and vocational skills as necessary for the job:

Six months of experience in clerical work. Training in typing, general office practices, office administration, mathematics, English grammar or business at the high school level or at an accredited post high school academic/vocational institution may be substituted for the required experience at the rate of one half high school unit, 90 clock hours or two semester hours for one month of experience. Some positions in this class may require a valid driver's license or proficiency in the operation of specific types of office equipment, special knowledge and/or special abilities.

Other specific requirements which are helpful but not required include skills in dictation, bookkeeping, and library science.

The results of this research caused this researcher to question these requirements. It may be in the state's best interest to assess and possibly revise the requirements to better fit the position.

One question that arises from this research is: If the selection test and the performance appraisal show a non-significant statistical relationship, how will the state defend its selection process for OA2s? Additionally, how are other state positions requiring a selection test doing? Could other positions be having the same difficulty in hiring the best applicants for the job? Only similar analysis on other jobs can answer these questions.

Still another explanation for these findings may be found in the performance appraisal. The appraisal is based on criteria which is set up by each supervisor. There is no set rule for what a supervisor considers important to the job. Therefore, it is theoretically possible that the selection test accurately predicts quality performance but the supervisor is in disagreement with the selection test. A method for controlling errors in supervisor evaluations is to annually train these supervisors in proper evaluation techniques. Periodic training can reduce the amount of subjective and biased evaluation by re-instructing the

supervisor as to what standards their employees must strive to reach.

This researcher suggests the state replicate this study using a larger sample. If the results are in conflict, the state should consider contracting an outside consultant to conduct still another validation study. However, if the state's findings are similar to the findings of this research, the state should consider a full-scale job analysis on the position of OA2. A job analysis is "the process of describing and recording aspects of jobs. Typically described and recorded are the purposes of a job, its major duties or activities, and the conditions under which the job is performed" (Schuler, 1987). The job analysis will define exactly what an OA2 does. The job analysis will also accurately define what specific criteria are needed to effectively perform the job. From this information, the selection test can be revised to better cover the requirements of the position. Furthermore, the state can revise the performance appraisal so that it may more accurately measure those skills that are in use on the job.

References

- American Psychological Association, Division of Industrial and Organizational Psychology. (1987). Principles for the validation and use of personnel selection procedures. (3rd ed).
- Anastasi, A. (1961). Psychological testing. (2nd ed.). New York: The MacMillan Company.
- Anastasi, A. (1988). Psychological testing. (6th ed.). New York: The MacMillan Company.
- Anstey, E. (1977). A 30 year follow-up of the CSSB procedure, with lessons for the future. Journal of Occupational Psychology, 50, 149-159.
- Arvey, R.D., & Campion, J.E. (1982). The employment interview: A summary and review of recent research. Personnel Psychology, 35, 281-322.
- Baker, H.G., & Spier, M.S. (1990). The employment interview: Guaranteed improvement in reliability. Public Personnel Management, 19, 85-90.
- Borg, W.R. and Gall, M.D. (1983). Educational research: an introduction. New York: Longman.
- Cascio, W.F. (1987). Applied psychology in personnel management. Englewood Cliffs, NJ: Prentice Hall.
- Cohen, J. and Cohen P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates.

- Cope, R.T. (1982). Concurrent validity of a written test for entry level social workers. Educational and Psychological Measurement, 42, 1181-1188.
- Distefano, M.K. Jr., Pryer, M.W., & Erffmeyer, R.C. (1983). Application of content validity methods to the development of a job-related performance rating criterion. Personnel Psychology, 36, 621-631.
- Dove, A. (1968). Taking the Chitlin Test. Newsweek, July 15, 1968, 51-52.
- DuBrin, A.J. (1981). Personnel and human resource management (rev. ed.). New York: D. Van Nostrand Company.
- Ebel, R.L. (1983). The practical validation of tests of ability. Educational Measurement: Issues and Practice, 2, 7-10.
- Hintz, J.L. (1988). Number crunching statistical system. Kaysville, Utah: NCSS
- Kerlinger, F.N. (1986). Foundations of behavioral research. New York: CBS College Publishing.
- Kleiman, L.S., & Faley, R.H. (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: a review and analysis. Personnel Psychology, 38, 803-833.
- Landy, F.J. (1989). Psychology of work behavior. Belmont, California: Brooks/Cole Publishing.

- Landy, F.J. (1976). The validity of the interview in police officer selection. Journal of Applied Psychology, 61, 193-198.
- Landy, F.J., & Bates, F. (1973). Another look at contrast effects in the employment interview. Journal of Applied Psychology, 58, 141-144.
- Lawshe, C.H. (1983). A simplified approach to the evaluation of fairness in employee selection procedures. Personnel Psychology, 36, 601-608.
- Lee, C. (1988). Testing makes a comeback. Training, 25, 49-52, 56-59.
- Leonard, J.W. (1986). Why MBO fails so often. Training and Development Journal, 40, 38-39.
- Maier, M.H. (1988). On the need for quality control in validation research. Personnel Psychology, 41, 497-502.
- Macdonald, C.R. (1982). MBO can work. McGraw-Hill: New York.
- Mayfield, E.C. (1964). The selection interview: a re-evaluation of published research. Personnel Psychology, 17, 239-260.
- McDaniel, M.A., Schmidt, F.L., & Hunter, J.E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. Personnel Psychology, 41, 283-309.

- Merritt-Haston, R., & Wexley, K.N. (1983). Educational requirements: Legality and validity. Personnel Psychology, 36, 743-753.
- Owens, W. (1976). Handbook of industrial and organizational psychology (rev. ed.). Chicago: Rand McNally
- Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalizations results for tests used to predict training success and job proficiency in clerical occupations. Journal of Applied Psychology, 65, 373-406.
- Reilly, R.R., & Chao, G.T. (1982). Validity and fairness of some alternative employee selection procedures. Personnel Psychology, 35, 1-62.
- Reynolds, A.H. (1979). The reliability of a scored oral interview for police officers. Public Personnel Management, 8, 324-328.
- Richards, B. (1986). Three classes of objectives and plans make MBO more effective. Personnel Journal, 65, 28-30.
- Riggio, R.E. (1990). Introduction to industrial/organizational psychology. Glenview IL: Scott, Foresman, and Company.
- Sackett, P.R. (1989). Integrity testing for personnel selection: An update. Personnel Psychology, 42, 491-529.
- Safrit, M.J., & Wood, T.M. (1981). The use of motor ability tests to achieve sex-fair ability groupings. (Paper presented at the Physical Education Ability Grouping and

Performance Evaluation Symposium; Ottawa, IL, August 13-14, 1981).

- Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.
- Schmidt, F.L., Gast-Rosenberg, I., & Hunter, J.E. (1980). Validity generalization results for computer programmers. Journal of applied psychology, 65, 643-661.
- Schmitt, N. (1976). Social and situational determinants of interview decisions: implications for the employment interview. Personnel Psychology, 29, 79-101.
- Schuler, R.S. (1987). Personnel and human resource management (3rd ed.). St. Paul: West Publishing Company
- Thompson, D.E., & Thompson, T.A. (1982). Court standards for job analysis in test validation. Personnel Psychology, 35, 865-873.
- Twomey, D.P. (1990). Equal employment opportunity law. (2nd. ed.). Cincinnati: South-Western.
- Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. Psychological Bulletin, 63, 100-116.
- Uniform guidelines on employee selection procedures. (1978). Federal Register, 43, 38290-38315.
- Wagner, R. (1949). The employment interview: A critical review. Personnel Psychology, 2, 17-46.

Wright, C.R. (1969). Summary of research on the selection interview since 1964. Personnel Psychology, 22, 391-413.

TO: All Graduate Students Who Submit a Thesis or
Research Problem/Project as Partial Fulfillment of
the Requirements for an Advanced Degree

FROM: Emporia State University Graduate School

I, JOHN P. GAMRON, hereby submit this
thesis/report to Emporia State University as partial
fulfillment of the requirements for an advanced degree. I
agree that the Library of the University may make it available
for use in accordance with its regulations governing materials
of this type. I further agree that quoting, photocopying, or
other reproduction of this document is allowed for private
study, scholarship (including teaching) and research purposes
off a nonprofit nature. No copying which involves potential
financial gain will be allowed without written permission of
the author.

John P. Gamron
Signature of Author

4/25/91
Date
A Validation Study of the Kansas Civil Service
Exam for Office Assistant, Level 2
Title of Thesis/Research Project

Jackie Sollert
Signature of Graduate Office Staff Member

April 25, 1991
Date Received

Distribution: Director, William Allen White Library
Graduate School Office
Author