Validation and Validity Generalization of Placement Exercises Used in the Selection of Clerical Employees

Keith Edward Mitchell

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Industrial and Organizational Psychology Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/1401

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
VALIDATION AND VALIDITY GENERALIZATION
OF PLACEMENT EXERCISES USED IN THE
SELECTION OF CLERICAL EMPLOYEES

by

Keith Edward Mitchell

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Arts
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
August 1985

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A follow-up study was conducted to determine the predictive validity of a group of job-related Placement Exercises used in the selection of clerical employees for a northeastern utility company. The present study was also to determine if the obtained validities could be generalized to similar organizational settings.

The subjects included in this study were 98 clerical employees working in the job classifications of Junior Clerk, Clerk, Intermediate Clerk, and Secretary-Stenographer. The obtained results supported the hypothesis that a positive and significant relationship existed between performance on the Placement Exercises and supervisory ratings of employee job performance. After applying the results of this study to the Bayesian validity generalization procedures provided by Pearlman, Schmidt, and Hunter (1980), it was concluded that the validity of the placement exercises could be generalized to similar organizational settings.
ACKNOWLEDGEMENTS

In completing this thesis, I must first thank the Eternal Father whose spirit has guided me, and whose tender mercies removed all obstacles. I must also extend my gratitude to Drs. Asher, Nangle and Brethower for their advice, patience, encouragement, and insight. From these gentlemen I have learned that a man can accomplish nothing alone.

Keith Edward Mitchell
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Mitchell, Keith Edward

VALIDATION AND VALIDITY GENERALIZATION OF PLACEMENT EXERCISES USED IN THE SELECTION OF CLERICAL EMPLOYEES

Western Michigan University M.A. 1985

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................. ii

LIST OF TABLES ...................................................... iv

INTRODUCTION .......................................................... 1
  Problem ............................................................. 1
  Literature Review .................................................. 4

METHODOLOGY .......................................................... 13
  Research Site and Subjects ....................................... 13
  Predictor Measures ................................................ 14
  Criterion Measures ................................................ 15
  Statistical Analysis .............................................. 17

RESULTS ................................................................. 19

DISCUSSION ............................................................ 24

APPENDICES ............................................................ 27
  A. Non-Exempt Salaried Employee Appraisal ...................... 28
  B. Validity Generalization Results for Proficiency Criterion Distribution ................................................. 30

BIBLIOGRAPHY .......................................................... 32
LIST OF TABLES

1. Comparison of Predictive and Concurrent Validation Results .................................. 19
2. Placement Exercise Intercorrelations ................................................................. 20
3. Intercorrelations Between Job Performance Factors ........................................... 21
4. Predictive Validity Coefficients ........................................................................... 22
5. Prior Distribution Mean Validities and Tailored Mean Validities ....................... 23
6. Validity Generalization Results for Proficiency Criterion Distributions ............... 30
INTRODUCTION

Problem

The objective of a personnel selection program is to predict, in advance of hiring, which job applicants will become successful and productive employees and which, if hired, would be inadequate or poor performing employees. To assist in the identification and hiring of potentially successful employees, many organizations have utilized a variety of instruments (objective and subjective) as part of their personnel selection program.

Selection instruments used in evaluating information from reference checks, academic achievement, previous work experience, interviews, and interest inventories have shown limited success in the prediction of job performance. On the other hand, the research on biographical data and peer ratings has reported more positive results. However, the use and validation of objective selection instruments have been much more prolific and promising (Dunnette, 1972; Reilly and Chao, 1982; Hunter and Hunter, 1984).

Brogden (1949) has noted that where objective selection instruments are used, their use could have a significant impact upon an organization. Hiring decisions made on the basis of a selection instrument that is inadequate or invalid could hurt an organization through the hiring of unqualified applicants who later become poor performers. Conversely, the failure to hire applicants possessing the skills and abilities needed within the organization is equally costly.
Several researchers have quantified the cost of selection procedures in dollar terms, through utility analysis, and have concluded that good selection procedures are critical to organizational success (Brogden, 1949; Cascio and Silbey, 1979; Cronbach and Gleser, 1965; Hunter and Hunter, 1984).

The possible impact that selection instruments could have, as illustrated above, serves to emphasize the necessity for organizations to insure that the selection instruments they employ are validated. In addition to the practical objectives an organization may attempt to meet through its selection procedures, government regulations also require that selection instruments be validated.

The organization in which the present research was conducted has utilized objective selection instruments in the hiring of its clerical employees for over ten years. These instruments were developed to measure the job related traits that have been identified as being essential for acceptable performance in the organization's clerical job classifications. The initial validation of the selection instruments was conducted under a concurrent validation design and the instruments were reported to be valid.

The objective of the present research is to determine the predictive validity of these instruments, and to provide empirical data to help broaden our understanding of the relationship between the instruments and measures of on-the-job performance. It is expected that the predictive validity of the instruments will indicate a significant and positive relationship with the supervisory rating criteria.
There has been much discussion in the research literature on whether the predictive design is the only appropriate validation method, or whether the concurrent validation design provides an acceptable approximation of predictive validity. The present study will review this issue and compare the obtained results with the results reported in the initial concurrent validation study. It should be noted that though the present study is predictive in nature, it is not exactly parallel to the methodologically pure predictive validation design.

The operating environment of an organization typically does not lend itself to the methodologically pure validation study. Factors such as the availability of an adequate sample size, limited availability of accurate and reliable criteria measures, and the availability of only present employees can place limitations on validation studies. The possible impact of these factors upon validation research will be examined within the context of the statistical power issue.

The results of the present study will also be evaluated in light of the recent findings of Pearlman, Schmidt, and Hunter (1980). In their study on validity generalization of employment tests used in clerical occupations, they provide prior distributions of validity coefficients for testing the generalizability of individual validation results. The results of the current study will be applied to their Bayesian prior distributions to estimate the validity of the selection instruments in similar job-test situations. The present study will

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
attempt to test the research hypothesis that the relationship between the selection instruments and supervisory ratings of employee performance is both positive and significant.

Literature Review

Few personnel psychologists would disagree with the necessity to validate selection instruments used in making hiring decisions. However, the research literature indicates that there has been much debate concerning the selection of an appropriate design strategy for obtaining validation information.

Researchers have differentiated between the design strategies of predictive and concurrent validation as early as Tiffin (1946). The issues surrounding the appropriate use of the two designs concern the accuracy of the validity estimates and the quality of the information obtained through the respective designs.

Discussion on the two designs has centered on whether concurrent validity is equivalent to, or is an inadequate substitute for predictive validity. Guion (1965) has taken the position that the concurrent validation design is a clear violation of scientific principles. In support of this position, Guion and Cranny (1982) argue that the population of interest in employee selection is a population of potential, not actual employees. Therefore, the sample population of present employees utilized in a concurrent design appears to be an inappropriate population.
Guion and Cranny (1982) have also argued that the validation designs which permit correction for range restriction will provide the most accurate estimates of true validity. Conversely, the authors point out that the nature of the data available in a concurrent validity study makes correction for range restriction virtually impossible. Several other researchers have supported Guion's contention and have concluded that the predictive validation methodology is superior to that of concurrent validation (Anastasi, 1976; Cascio, 1978; Dunnette, 1966).

While acknowledging that the predictive and concurrent designs are not equivalent, Barrett, Phillips, and Alexander (1981) have argued that the conceptual and empirical differences between the designs have been exaggerated. They contend that the weaknesses historically associated with the concurrent design (the missing persons problem, restriction of range, motivational and demographic differences, and confounding by job experience) affect the predictive design as well. These researchers also postulate that the effects of the above factors are minimal and act equally in the two designs or are counter-balanced with other factors across the two designs.

Barrett's et al. (1981) position appears to be supported by the work of Pearlman, Schmidt, and Hunter (1980). In their analysis of predictive and concurrent validity coefficients included in their Bayesian prior distributions, the researchers concluded that there were no appreciable differences in observed validity coefficients for the two designs. They further concluded that it was appropriate to
include both predictive and concurrent validities in their study. Jensen (1980) has reached similar conclusions in his analysis of a large number of predictive and concurrent validity studies.

The proponents on both sides of this design issue have recognized, as does the present research, that the practical considerations and limitations existing within an organizational setting will dictate the methodological strategy that can be employed in obtaining validity information. As will be seen, these factors could also have a significant impact upon the statistical power that can exist within a criterion-related validation study.

Researchers have only recently come to recognize the importance of having an adequate level of statistical power in conducting criterion-related validation research. Prior to the work of Cohen (1970), and Schmidt, Hunter, and Urry (1976) it seems little concern had been shown for the statistical power necessary for testing research hypotheses. Trattner and O'Leary (1980) have reached similar conclusions in their research.

Schmidt et al. (1976) have postulated that the lack of concern for statistical power observed in past research studies stems from many researchers' willingness to believe in what they call "the erroneous law of small numbers". They argue that in accepting this law, researchers are assuming that small samples (30 to 50 subjects) are as representative of population parameters as large random samples. The results of an earlier study indicated that psychol-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ogists, in general, overestimated the power of statistical tests in small samples (Tversky and Kahneman, 1971).

Prior reviews of validity studies have found the median sample sizes to fall in the range from 40 to 68 subjects (Jones, 1950; Guion, 1965; Lent, Aurbach, and Levin, 1971). Schmidt and Hunter (1980) report that when sample sizes are limited to 30-50 subjects, the level of statistical power available in such studies will detect a truly significant validity only 25 to 50 percent of the time. These findings have led the researchers to conclude that criterion-related validity studies are technically feasible much less frequently than is commonly assumed.

In addition to the low statistical power available in the typical validation study (due to small sample size), several researchers have argued that the impact of range restriction, criterion reliability, and predictor reliability also contribute to the underestimation of true validity coefficients (Schmidt et al., 1976; Lee, Miller and Graham, 1982). It is believed that the impact of these factors is to significantly increase the sample size needed to detect statistically significant validity coefficients, when they exist.

Along with increasing sample sizes for greater statistical power, recent studies have suggested that validity coefficients should be corrected for range restriction and criterion unreliability in obtaining accurate estimates of true predictor-criterion validities (Schmidt et al., 1976; Pearlman et al., 1980; Lee et al., 1982). These findings are in contrast to the Standards for Educational and
Psychological Tests (American Psychological Association, 1974) which has taken the position that both corrections should not be made. However, Lee et al. (1982) have concluded that range restriction and criterion unreliability are simultaneously present in most validity studies, and that coefficients corrected for attenuation and restriction in range are better estimates of true validities.

Consistent with the above findings, Schmidt and Hunter (1977) report that when the effects of statistical artifacts such as range restriction and criterion unreliability are taken into consideration, validity coefficients become much more stable across studies. They believe that correcting the observed coefficients for these artifacts results in a more accurate estimate of the true validity and makes validity generalization to other settings more feasible.

Researchers have traditionally believed that test validation results were situation specific and could not be generalized to other settings, even when test types and organizational settings were similar. Ghiselli (1966) noted that there is a great deal of variability in the magnitude of obtained validity coefficients from one study to another. This variability has led him and other researchers to conclude that it is not possible to generalize test validity results from one setting to another, and that empirical validation is necessary in each organizational setting that a test is to be used. (Albright, Glennon, and Smith, 1963; Guion, 1965).

In recent years, however, several researchers have presented
empirical evidence challenging this long standing belief. Schmidt and Hunter (1977) have concluded that the results of their research demonstrate that a substantial amount of the variability observed between empirical validation studies can be explained when the effects of statistical artifacts are considered. They have identified seven artifacts that they believe contribute to this variability; they are: criterion unreliability, predictor unreliability, range restriction, sampling error due to small sample size, computational and typographical errors, criterion contamination and deficiency, and slight differences in factor structures between different tests thought to measure similar constructs.

In a series of studies Schmidt and his colleagues have demonstrated that, through the Bayesian approach to statistical probability, validity coefficients could be corrected for four of the above sources of variance resulting in validity coefficients which were quite stable across similar job-test studies (Schmidt et al., 1976; Schmidt and Hunter, 1977, 1978; Pearlman et al., 1980; Schmidt, Gast-Rosenberg, and Hunter, 1980; Schmidt, Hunter, and Pearlman, 1981; Schmidt, Hunter, Pearlman, and Shane, 1979). The researchers have conceptualized their validity generalization approach as the accumulation of data from prior empirical validation studies in such a way as to eliminate the effects of artifacts produced by inevitable imperfections in the individual studies.

It is believed that the superiority of the Bayesian approach in estimating true test validity lies in its use of both sample-derived
information and relevant information from prior studies on the same test or test type on similar jobs (Schmidt and Hunter, 1977). In this approach, after correcting for error variance due to the statistical artifacts of criterion reliability, predictor reliability, range restriction, and sampling error, the mean of the resulting posterior distribution of validity coefficients is then considered to be the best estimate of true test validity across all similar job-test situations. The Bayesian approach does not correct for the remaining artifacts due to the fact that they would be difficult if not impossible to estimate. By not correcting for these artifacts the researchers contend that the procedure insures conservative estimates (Pearlman, Schmidt, and Hunter, 1980).

Of particular relevance to the present study are the results reported by Pearlman et al. (1980) in their research on the Bayesian validity generalization procedure for clerical occupations. Their study represents the first large-scale test of this validity generalization model.

Pearlman et al. (1980) collected data from both published and unpublished validity studies on five clerical job families, and their study included 10 test types as predictor variables. Their data included 2,786 prior validity coefficients on overall job proficiency or performance criteria. Based on the data collected on each job-test combination, prior distributions were constructed along with assumed distributions of criterion reliability, test reliability, and range restriction. The researchers found it necessary to rely on assumed
distributions of these artifacts because they found the majority of the validity studies included in their prior distributions of validity coefficients did not contain the information necessary to determine the actual values for these artifacts (Pearlman et al., 1980; Jones, 1950).

As noted earlier, the prior distributions were corrected for the assumed mean values of the above artifacts and for variance due to sampling error. The mean value of the corrected Bayesian prior distribution of validity coefficients was then considered the best estimate of true validity, and used in generalizing to other settings. The present study utilized the proficiency criterion prior distributions provided by Pearlman et al. (1980) to evaluate the generalizability of obtained results from the present study to similar organizational settings.

Pearlman's et al. (1980) report that the results of their study supported their conclusion that most of the variability in empirical validity coefficients is the result of statistical artifacts. They also concluded that in most cases validity generalization to similar clerical jobs or settings was possible.

Using similar approaches other researchers have obtained comparable results, thus lending methodological support to the Bayesian validity generalization model (Glass, 1976; Callender, Osburn, Greener, and Ashworth, 1982; Roju and Burke, 1983). Promising results using the Bayesian model have also been reported outside of the area of employee selection (Terborg, Lee, Smith, Davis, and Turbin, 1982).
The objective of an employment test or any instrument used in the selection process is to predict future job performance. Of equal importance in this prediction equation is the choice of appropriate criteria by which job performance can be accurately evaluated.

The test validation literature has identified many types of objective and judgmental measures that have been used as criteria in evaluating job performance. Data obtained from production records, absenteeism rates, turnover rates, supervisory and peer ratings, and other sources have been used as criteria in studying the relationship between predictors and on-the-job performance (Smith, 1976). However, it appears that a majority of the test validation studies have been conducted utilizing supervisory ratings as the criteria in evaluating employee job performance (Landy and Farr, 1980). This is consistent with the use of supervisory ratings by a vast majority of business organizations in evaluating employee job performance.

The organization in which the present study was conducted has utilized supervisory ratings to evaluate its employees for many years. The present study will examine the relationship between the selection instruments currently in use by the organization and supervisory ratings of employee job performance.
METHODOLOGY

Research Site and Subjects

The organization in which this study was conducted is a north­eastern utility company. The subjects were selected from the two office locations having the largest number of clerical employees on their payroll. In that there was a limited number of subjects in the available employee population it was not possible to select the subjects by random sampling.

The subjects included in the present study were 98 clerical employees who were performing a wide variety of clerical and non-physical assignments within the organization. At the time this study was conducted the subjects were working in the job classifications of Junior Clerk, Clerk, Intermediate Clerk, and Secretary-Stenographer. This study included all clerical employees in the two locations for which placement exercise scores were available. However, to obtain reliable data on the performance rating criteria, subjects employed by the organization for less than one year were excluded from the study. It should also be noted that none of the subjects included in the present study were found to be participants in the initial concurrent validation study.
Predictor Measures

The predictor measures this study focused upon were a group of four paper and pencil instruments, constructed to measure the job-related traits associated with successful performance in the organization's clerical job classifications. These instruments have been designated as Placement Exercises, and have been used as an integral part of the organization's selection procedures since 1974.

The placement exercises are identified as PMC-3, PD-7, NC-5, and WE-9/10, and were developed based on the job analysis procedure called Threshold Traits Analysis (Lopez, Kesselman, and Lopez, 1981). The exercises are said to contain samples of the work performed in the organization's clerical classifications, and were constructed to measure the job-related traits of Perception, Concentration, Memory and Comprehension (PMC-3); Planning and Decision Making (PD-7); Numerical Computation (NC-5); and Written Expression (WE-9/10). Each exercise contains thirty items.

Prior to the addition of the placement exercises to the organization's selection program, concurrent validation procedures were carried out. The results of this initial validation study (completed in 1973) indicated that the exercises (PMC-3, NC-5, PD-7 and WE-9/10) were valid and possessed acceptable levels of reliability. Though included in the initial validation study, and found to be valid, placement exercise PD-7 was the most recent instrument to be added to the organization's selection program.

The placement exercises were administered to each clerical
employee prior to being hired. In the organization's selection procedures, performance on the placement exercises, employment interviews, past work experience, and other relevant information are considered in arriving at the hiring decision. The placement exercises primarily serve as instruments for screening out those job applicants who are believed to be the least qualified.

The specific scores an individual receives on the placement exercises do not determine whether the individual is hired. However, the use of the exercises as screening instruments does, to some degree, influence the overall hiring decision. Taking this into consideration, the researcher recognizes that the present study is not exactly parallel to the methodologically pure predictive validation design. However, as previously emphasized, the practical considerations and limitations existing within most organizational settings typically do not permit a methodologically pure design strategy. The placement exercise scores on each subject were gathered from personnel files maintained on each of the organization's employees.

**Criterion Measures**

As part of a new salary administration program instituted in 1977, the organization developed a new appraisal form (PSC-82) for evaluating the job performance of its clerical employees. Each clerical employee is annually evaluated on their job performance by their immediate supervisor and the appraisal form is included in the
The PSC-82 contains a five-degree graphic scale, by which an employee's performance is evaluated on nine job performance factors and an overall or summary rating. In the present study the performance factors Work Output-Quality, Work Output-Quantity, Job Knowledge, Job Skill, and Overall Rating were used as criterion measures (See Appendix A).

Utilizing the graphic scale each employee's supervisor annually rates the employee's performance on the nine factors as either unsatisfactory, below average, satisfactory, good, or excellent. These ratings have been scored from 1 to 5 respectively for each of the performance factors. The performance factors were found to have mean ratings from 3.673 to 3.898, and were found to be skewed from .073 to -.453. Four of the five performance factors were negatively skewed.

In the present organization newly appointed supervisors receive individualized training on evaluating employee performance, and in completing the employee appraisal form (PSC-82). The researcher used the most recent employee appraisals available at the time criterion data were collected.

In the present study sufficient data is not available to determine the reliability of the ratings. However, based on the data reported by Pearlman (1979), and Pearlman et al. (1980), we can make certain assumptions concerning the reliability of the ratings. The above researchers were able to construct assumed distributions of criterion reliabilities, test reliabilities, and range restriction effects across validation studies. The distributions were based on
the best estimates of such effects available from the research literature and test manuals. They also point out that such data is generally not provided in the majority of research studies. The assumed distribution of proficiency criterion reliabilities refers to criterion reliabilities in the applicant population, that is, reliabilities corrected for restriction in range (Schmidt, Hunter, and Urry, 1976). The expected mean value for proficiency criterion reliability is .60.

Statistical Analysis

Pearson product-moment correlation coefficients were calculated, under the assumption of linearity, in determining the relationship between placement exercise scores and the five performance rating factors. Means, standard deviations, and variances for the predictor variables were also calculated using the SPSS program (Klecka, Nie, and Hull, 1975). Split-Half Reliability coefficients (odd-even) were calculated for each placement exercise, and the Spearman-Brown correction formula was applied to the obtained coefficients.

The Bayesian validity generalization procedure was applied to the obtained results to determine the generalizability of the placement exercises to similar organizational settings. Applying these procedures to individual validation results requires utilizing the data on prior empirical validation studies accumulated by Pearlman et al. (1980). These researchers have provided distributions of mean validity coefficients for each test type/job category used in their
study on clerical occupations (see Appendix B).

The prior distributions provided by these researchers have been corrected for criterion unreliability and range restriction, but not for test unreliability. Pearlman et al. (1980) point out that the means of the prior distributions of validity coefficients were computed assuming that the specific tests to which the distributions will be applied have a reliability of .80. This value (.80) represents the assumed average test reliability across studies.

In applying the Bayesian procedure to the results obtained in the present study, the means of the prior distributions were first corrected for assumed average test reliability (divide the mean validity coefficients by the square root of .80). The fully corrected means were then tailored to the placement exercises by attenuating the corrected prior distribution means by the square root of the placement exercise reliabilities (Pearlman et al., 1980; Schmidt and Hunter, 1977). It is believed that this procedure produces more accurate estimates of true test validities in similar job-test settings.
RESULTS

A comparison of placement exercise means, standard deviations, reliabilities, and validity coefficients for the present predictive validation study and the initial concurrent validation study are presented in Table 1. Predictive and concurrent validity coefficients ($r_{xy}$) for the criterion of job performance rating (overall) and scores on placement exercises PMC-3 and NC-5 are all significant at the .01 level (one-tailed test). The Predictive and concurrent validity coefficients for WE-9/10 are significant at the .05 level (one-tailed test), and the predictive and concurrent validity coefficients for PD-7 are significant at the .01 and .05 levels, respectively (one-tailed test).

Table 1  
Comparison of Predictive and Concurrent Validation Results

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Study</th>
<th>Number of Subjects</th>
<th>Mean</th>
<th>S.D</th>
<th>$r_{XX}$</th>
<th>$r_{xy}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMC-3</td>
<td>Predictive</td>
<td>98</td>
<td>27.3</td>
<td>3.33</td>
<td>.92</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>Concurrent</td>
<td>119</td>
<td>26.6</td>
<td>4.07</td>
<td>.88</td>
<td>.24</td>
</tr>
<tr>
<td>NC-5</td>
<td>Predictive</td>
<td>98</td>
<td>22.7</td>
<td>5.47</td>
<td>.90</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td>Concurrent</td>
<td>119</td>
<td>22.9</td>
<td>5.74</td>
<td>.88</td>
<td>.23</td>
</tr>
<tr>
<td>WE-9/10</td>
<td>Predictive</td>
<td>98</td>
<td>22.5</td>
<td>3.94</td>
<td>.77</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>Concurrent</td>
<td>93</td>
<td>21.6</td>
<td>4.22</td>
<td>.76</td>
<td>.22</td>
</tr>
<tr>
<td>PD-7</td>
<td>Predictive</td>
<td>79</td>
<td>24.3</td>
<td>3.20</td>
<td>.73</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>Concurrent</td>
<td>61</td>
<td>22.7</td>
<td>4.86</td>
<td>.88</td>
<td>.25</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The intercorrelation matrix for the four placement exercises is presented in Table 2. The intercorrelation coefficient between placement exercises PMC-3 and PD-7 is significant at the .05 level (one-tailed test). All other placement exercise intercorrelations are significant at the .01 level (one-tailed test).

Table 2
Placement Exercise Intercorrelations

<table>
<thead>
<tr>
<th></th>
<th>PMC-3</th>
<th>NC-5</th>
<th>WE-9/10</th>
<th>PD-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMC-3</td>
<td>-</td>
<td>.42</td>
<td>.27</td>
<td>.19</td>
</tr>
<tr>
<td>NC-5</td>
<td>-</td>
<td>-</td>
<td>.41</td>
<td>.42</td>
</tr>
<tr>
<td>WE-9/10</td>
<td></td>
<td>-</td>
<td></td>
<td>.52</td>
</tr>
<tr>
<td>PD-7</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

The intercorrelations between the job performance factors; Work Quality, Work Quantity, Job Knowledge, Job Skill and Overall Rating obtained from the appraisal form PSC-82 are presented in Table 3. All job performance factor intercorrelations are significant at the .01 level (one-tailed test).
The predictive validity coefficients presented in Table 4 indicate the correlation between the placement exercises and the job performance factors used as criteria in this study. The coefficients range from -.09 to .36 and as would be expected the higher correlations were obtained between the predictors and the overall job performance rating. For placement exercise PMC-3, NC-5, and WE-9/10 a correlation of approximately .17 is required for significance at the .05 level (one-tailed test). For placement exercise PD-7 a correlation coefficient approaching .19 would be required for significance at the .05 level (one-tailed test).
Table 4  
Predictive Validity Coefficients

<table>
<thead>
<tr>
<th>Criterion</th>
<th>PMC-3</th>
<th>NC-5</th>
<th>WE-9/10</th>
<th>PD-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Ql.</td>
<td>.14</td>
<td>.21</td>
<td>.15</td>
<td>.29</td>
</tr>
<tr>
<td>Work Qn.</td>
<td>.15</td>
<td>.15</td>
<td>.17</td>
<td>.28</td>
</tr>
<tr>
<td>Job Know.</td>
<td>.07</td>
<td>.11</td>
<td>-.09</td>
<td>.14</td>
</tr>
<tr>
<td>Job Sk.</td>
<td>.13</td>
<td>.10</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>Overall</td>
<td>.26</td>
<td>.33</td>
<td>.18</td>
<td>.36</td>
</tr>
</tbody>
</table>

To determine the extent to which use of the predictors can be generalized to similar job-test situations, data provided by Pearlman et al. (1980) is utilized (See Appendix B). The Bayesian prior distributions for job proficiency criteria is used in estimating the true validity of the predictors in new settings.

The data presented in Table 5 indicate the appropriate prior distribution mean validities for job proficiency (Pearlman et al., pp. 388–389, 1980) and the prior mean validity coefficients tailored to the reliability of the placement exercises. The prior distributions utilized were based on composites of job categories (A–E) representing a variety of clerical occupations. The tailored mean validity coefficients are considered to be the most accurate estimates of the validity of the predictors in similar settings. Table 5 also lists the placement exercises and the test type classifications.
(Pearlman et al., 1980) consistent with the traits measured by the placement exercises. It should be noted that the Planning and Decision Making traits as measured by placement exercise PD-7 have been included by Pearlman et al. (1980) as a subset of the Reasoning Ability test type classification.

Table 5
Prior Distribution Mean Validities
And Tailored Mean Validities

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Prior ( r_{xy}^{a} )</th>
<th>Tailored ( r_{xy}^{a} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory (PMC-3)</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>Quantitative Ability (NC-5)</td>
<td>0.50</td>
<td>0.53</td>
</tr>
<tr>
<td>Verbal Ability (WE-9/10)</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td>Reasoning Ability (PD-7)</td>
<td>0.44</td>
<td>0.42</td>
</tr>
</tbody>
</table>

\(^a\)Mean validity coefficients for proficiency criterion provided by Pearlman et al. (1980, pp. 338-389), for the composite of clerical job categories A through E (see appendix B).
DISCUSSION

The results of the present study appear to support the hypothesis that the relationship between the placement exercises and supervisory ratings of employee performance is a positive and significant one. The magnitude of the predictive validity coefficients ranged from .18 for the Written Expression exercise to .36 for the Planning and Decision Making exercise. Three of the predictive validity coefficients were significant at the .01 level and the fourth was significant at the .05 level (one-tailed test).

A comparison of the results from the present study with the results reported in the initial concurrent validation study indicate, for the most part, that the results are comparable (see Table 1). This appears to support Pearlman's et al. (1980) conclusion that there is no appreciable differences in observed validity coefficients for predictive and concurrent designs.

The intercorrelations obtained between the four placement exercises ranged from .19 to .52. If the placement exercises are in fact measures of job-related traits, a moderate degree of interrelationship between the measures would normally be expected. All intercorrelations were found to be significant at either the .05 level or the .01 level (one-tailed test).

In the present organization, newly appointed supervisors receive individualized training on evaluating employee performance, and in completing the employee appraisal form (PSC-82). Though the
performance factors are intended to measure different aspects of an employee's job performance, some degree of interrelationship would logically be expected.

The intercorrelations obtained between the job performance factors ranged from .55 to .75, and all intercorrelations were found to be significant at the .01 level (one-tailed test). With intercorrelations of this magnitude, consideration should be given to the possible impact of halo and other rating errors on the observed correlations. Such results should not be surprising however, in that, a review of the literature on this problem reveals that halo is probably the most pervasive error in performance appraisal, and one that is difficult if not impossible to control (Borman, 1975; Cascio, 1978; Smith, 1976; King, Hunter and Schmidt, 1980; Nathan and Lord, 1983).

In order to determine whether the validity of the placement exercises could be generalized to similar organizational settings, the results obtained in the present study were applied to the validity generalization procedures presented by Pearlman et al. (1980). Tailoring the appropriate mean prior distribution validity coefficients to the reliability of the placement exercises resulted in validity coefficients ranging from .39 to .53 in magnitude.

It is believed that the tailored coefficients obtained through this procedure represent the best estimate of the placement exercises validity in similar settings. The magnitude of the tailored coeffi-
cients appear to support the generalizability of the placement exercise validities.

Based on the results of the present study, the continued use of the placement exercises, as part of the selection procedures for clerical employees is recommended. It is also concluded that the use of the placement exercises in similar organizational settings would be appropriate and beneficial. However, it is recommended that the placement exercises be used in conjunction with other selection procedures (interviews, work history data, etc.) for optimum usefulness.
APPENDIX A

NON-EXEMPT SALARIED EMPLOYEE APPRAISAL

EMPLOYEE NAME ___________________________ DATE _____________

JOB CLASSIFICATION ________________________ SALARY GRADE ______

DEPARTMENT/SECTION/MINE ___________________ DATE OF EMPLOYMENT______

DIVISION/PLANT _____________________________ DATE ON THIS JOB __________

LOCATION _________________________________ DATE OF LAST APPRAISAL __________

PLEASE "X" APPROPRIATE BOXES BELOW:

<table>
<thead>
<tr>
<th>JOB PERFORMANCE FACTOR</th>
<th>UNSATISFACTORY</th>
<th>BELOW AVERAGE</th>
<th>SATISFACTORY</th>
<th>GOOD</th>
<th>EXCELLENT OR OUTSTANDING</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORK OUTPUT - QUALITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORK OUTPUT - QUANTITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOB KNOWLEDGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOB SKILLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INITIATIVE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COOPERATION &amp; ATTITUDE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEPENDABILITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTENDANCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PUNCTUALITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERALL APPRAISAL*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Or Summary Rating*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explain Overall Appraisal Rating - Also* If Unsatisfactory or Below Average performer (i.e., in overall appraisal or summary rating), what have you done or what do you plan to do to obtain improved performance or to correct the situation?

________________________________________________________________________

________________________________________________________________________

DEGREE OF SUPERVISION REQUIRED:
(1) Requires Frequent and/or Close Supervision. □ (3) Little or Infrequent supervision needed. □
(2) Needs Moderate or Average amount of □ (4) Needs only Minimum supervision. □ Supervision
Almost None required.

JOB TRAINING OR QUALIFICATION PROGRAMS:
Any completed in past 12 months? YES _________ NO _________

If "Yes." indicate type and content of training
________________________________________________________________________

________________________________________________________________________

(over)
Appendix A – continued

APPRAISAL INTERVIEW SUMMARY:

Date of Interview _______________________

Summary of Major Points Made By:

EMPLOYEE:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

SUPERVISOR:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

OTHER REMARKS:

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

Supervisor ______________________ Date ______________________________

Department Head and Other Reviewers ________________________________

(If needed, please use additional sheet for any extra comments, etc.)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### APPENDIX B

Validity Generalization Results for Proficiency Criterion Distributions (Pearlman et al., 1980, pp. 388-389)

<table>
<thead>
<tr>
<th>Test Type/Job Category</th>
<th>Total N</th>
<th>No. Total</th>
<th>Prior Distribution</th>
<th>90% c.v.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{p}$</td>
<td>$SD\hat{p}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{p}$</td>
<td>$SD\hat{p}$</td>
</tr>
<tr>
<td>General Mental Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4,487</td>
<td>76</td>
<td>.50</td>
<td>.24</td>
</tr>
<tr>
<td>B</td>
<td>4,432</td>
<td>47</td>
<td>.49</td>
<td>.24</td>
</tr>
<tr>
<td>E</td>
<td>718</td>
<td>10</td>
<td>.43</td>
<td>.00</td>
</tr>
<tr>
<td>A-E</td>
<td>10,564</td>
<td>144</td>
<td>.50</td>
<td>.22</td>
</tr>
<tr>
<td>A-H</td>
<td>17,539</td>
<td>194</td>
<td>.52</td>
<td>.24</td>
</tr>
<tr>
<td>Verbal Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>18,227</td>
<td>215</td>
<td>.39</td>
<td>.23</td>
</tr>
<tr>
<td>B</td>
<td>6,712</td>
<td>97</td>
<td>.41</td>
<td>.25</td>
</tr>
<tr>
<td>C</td>
<td>1,331</td>
<td>28</td>
<td>.37</td>
<td>.00</td>
</tr>
<tr>
<td>A-E</td>
<td>27,352</td>
<td>355</td>
<td>.40</td>
<td>.22</td>
</tr>
<tr>
<td>A-H</td>
<td>39,187</td>
<td>450</td>
<td>.39</td>
<td>.23</td>
</tr>
<tr>
<td>Quantitative Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>13,658</td>
<td>155</td>
<td>.49</td>
<td>.13</td>
</tr>
<tr>
<td>B</td>
<td>9,001</td>
<td>121</td>
<td>.52</td>
<td>.16</td>
</tr>
<tr>
<td>C</td>
<td>1,347</td>
<td>33</td>
<td>.60</td>
<td>.09</td>
</tr>
<tr>
<td>E</td>
<td>1,271</td>
<td>17</td>
<td>.45</td>
<td>.00</td>
</tr>
<tr>
<td>A-E</td>
<td>25,850</td>
<td>333</td>
<td>.50</td>
<td>.13</td>
</tr>
<tr>
<td>A-H</td>
<td>39,584</td>
<td>453</td>
<td>.47</td>
<td>.14</td>
</tr>
<tr>
<td>Reasoning Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3,069</td>
<td>36</td>
<td>.38</td>
<td>.00</td>
</tr>
<tr>
<td>B</td>
<td>1,240</td>
<td>29</td>
<td>.63</td>
<td>.12</td>
</tr>
<tr>
<td>C</td>
<td>739</td>
<td>10</td>
<td>.31</td>
<td>.18</td>
</tr>
<tr>
<td>A-E</td>
<td>5,377</td>
<td>80</td>
<td>.44</td>
<td>.14</td>
</tr>
<tr>
<td>A-H</td>
<td>11,586</td>
<td>116</td>
<td>.39</td>
<td>.15</td>
</tr>
<tr>
<td>Perceptual Speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>28,824</td>
<td>368</td>
<td>.45</td>
<td>.24</td>
</tr>
<tr>
<td>B</td>
<td>17,043</td>
<td>251</td>
<td>.50</td>
<td>.14</td>
</tr>
<tr>
<td>C</td>
<td>2,951</td>
<td>50</td>
<td>.45</td>
<td>.00</td>
</tr>
<tr>
<td>D</td>
<td>878</td>
<td>10</td>
<td>.40</td>
<td>.22</td>
</tr>
<tr>
<td>E</td>
<td>1,665</td>
<td>23</td>
<td>.39</td>
<td>.12</td>
</tr>
<tr>
<td>A-E</td>
<td>51,361</td>
<td>702</td>
<td>.47</td>
<td>.20</td>
</tr>
<tr>
<td>A-H</td>
<td>70,935</td>
<td>882</td>
<td>.47</td>
<td>.22</td>
</tr>
</tbody>
</table>
### Appendix B - continued

<table>
<thead>
<tr>
<th>Test Type/Job Category</th>
<th>Total N</th>
<th>No. rs</th>
<th>Prior Distribution</th>
<th>90% c.v.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>^p</td>
<td>SDp</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3,323</td>
<td>49</td>
<td>.38</td>
<td>.20</td>
</tr>
<tr>
<td>B</td>
<td>1,970</td>
<td>39</td>
<td>.42</td>
<td>.00</td>
</tr>
<tr>
<td>C</td>
<td>726</td>
<td>11</td>
<td>.44</td>
<td>.14</td>
</tr>
<tr>
<td>A-E</td>
<td>6,278</td>
<td>102</td>
<td>.39</td>
<td>.12</td>
</tr>
<tr>
<td>A-H</td>
<td>7,764</td>
<td>117</td>
<td>.38</td>
<td>.17</td>
</tr>
<tr>
<td><strong>Spatial/Mechanical Ability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4,247</td>
<td>38</td>
<td>.20</td>
<td>.12</td>
</tr>
<tr>
<td>B</td>
<td>3,782</td>
<td>47</td>
<td>.42</td>
<td>.19</td>
</tr>
<tr>
<td>C</td>
<td>537</td>
<td>12</td>
<td>.48</td>
<td>.06</td>
</tr>
<tr>
<td>A-E</td>
<td>9,240</td>
<td>107</td>
<td>.30</td>
<td>.20</td>
</tr>
<tr>
<td>A-H</td>
<td>9,306</td>
<td>108</td>
<td>.30</td>
<td>.19</td>
</tr>
<tr>
<td><strong>Motor Ability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>7,662</td>
<td>95</td>
<td>.29</td>
<td>.25</td>
</tr>
<tr>
<td>B</td>
<td>8,405</td>
<td>97</td>
<td>.30</td>
<td>.14</td>
</tr>
<tr>
<td>C</td>
<td>1,521</td>
<td>21</td>
<td>.27</td>
<td>.00</td>
</tr>
<tr>
<td>D</td>
<td>834</td>
<td>12</td>
<td>.15</td>
<td>.23</td>
</tr>
<tr>
<td>E</td>
<td>1,360</td>
<td>21</td>
<td>.26</td>
<td>.00</td>
</tr>
<tr>
<td>A-E</td>
<td>19,782</td>
<td>246</td>
<td>.29</td>
<td>.18</td>
</tr>
<tr>
<td>A-H</td>
<td>21,277</td>
<td>257</td>
<td>.30</td>
<td>.21</td>
</tr>
<tr>
<td><strong>Performance Tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4,416</td>
<td>55</td>
<td>.50</td>
<td>.43</td>
</tr>
<tr>
<td>A-E</td>
<td>5,316</td>
<td>61</td>
<td>.47</td>
<td>.42</td>
</tr>
<tr>
<td>A-H</td>
<td>6,265</td>
<td>67</td>
<td>.44</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Clerical Aptitude</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4,127</td>
<td>63</td>
<td>.50</td>
<td>.22</td>
</tr>
<tr>
<td>B</td>
<td>1,674</td>
<td>26</td>
<td>.53</td>
<td>.26</td>
</tr>
<tr>
<td>A-E</td>
<td>5,989</td>
<td>94</td>
<td>.51</td>
<td>.23</td>
</tr>
<tr>
<td>A-H</td>
<td>11,927</td>
<td>142</td>
<td>.48</td>
<td>.24</td>
</tr>
</tbody>
</table>

Note: c.v. = credibility value.

Job categories are: A = stenography, typing, filing, and related occupations (DOT Occupational Groups 201-209), B = computing and account-recording occupations (DOT Occupational Groups 210-219), C = production and stock clerks and related occupations (DOT Occupational Groups 221-229), D = information and message distribution occupations (DOT Occupational Groups 230-239), E = public contact and clerical service occupations (DOT Occupational Groups 240-248), F = miscellaneous clerical occupations (DOT Occupational Group 249), G = unspecified clerical occupations, and H = mixed samples. bFinger, hand, and arm dexterity tests and motor coordination tests.

Tests comprised of verbal, quantitative, and perceptual speed components.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.