PERSONNEL PSYCHOLOGY
1988, 41

A META-ANALYSIS OF THE VALIDITY OF METHODS FOR RATING TRAINING AND EXPERIENCE IN PERSONNEL SELECTION

MICHAEL A. McDANIEL
Defense Personnel Security Research and Education Center

FRANK L. SCHMIDT
University of Iowa

JOHN E. HUNTER
Michigan State University

This paper reviews the validity evidence for systematic methods used to evaluate training and experience (T&E) ratings in personnel selection. Meta-analytic summaries of the data indicate that validity varies with the type of T&E evaluation procedure used. The Illinois job element and behavioral consistency methods each demonstrated useful levels of validity (.20 and .45, respectively) with small corrected standard deviations, thus supporting validity generalization. Both the point and task methods yielded low mean validities (.11 and .15, respectively) with larger variability. The authors hypothesized that both the point and task methods were affected by a job experience moderator. Partial support for this hypothesis was found. Moderator analyses suggested that the point method was most valid when the applicant pool had low mean levels of job experience and was least valid with an experienced applicant pool. Additional research is desirable on all T&E methods to decrease the potential impact of second-order sampling error in the meta-analytic results. Further research is also needed to explicate the constructs measured by T&E evaluations.

This paper summarizes and evaluates the validity evidence for methods of evaluating the training and experience (T&E) of applicants. T&E methods attempt to predict future job performance through systematic, judgment-based evaluations of information provided by applicants on resumés, applications, or other documents. This paper discusses the rationale for using T&E methods in applicant appraisal, reviews specific approaches to T&E evaluation, and presents a meta-analysis of T&E criterion-related validity coefficients. Meta-analytic results are presented for T&E methods as a whole and separately for specific T&E approaches.

The opinions expressed in this article are those of the authors and do not necessarily reflect the opinions and policies of the organizations with which they are affiliated.

Correspondence and requests for reprints should be addressed to Michael A. McDaniel, Defense Personnel Security Research and Education Center, 99 Pacific Street, Building 455 Suite E, Monterey, CA 93940.

Despite the fact that formal T&E evaluations are frequently used personnel selection methods, particularly in the public sector, personnel psychologists have conducted relatively little research on the validity of T&E evaluations, and most of this research has not been published. The present study cumulates and summarizes existing validity data on these methods.

The T&E evaluation methods considered in this study differ from empirically keyed biographical inventories (Owens, 1976) in that information is weighted on the basis of judgment rather than on empirically estimated validities. The judgmental weighting is based on implicit or explicit hypotheses about the relation between personal characteristics or experiences and later job performance. For example, a T&E evaluation method might determine an applicant's score by giving the applicant three points for each year of related job experience plus two points for each year of pertinent college education. This scoring scheme is typical of the point method approach to T&E evaluation. T&E methods are primarily used to rank order applicants following elimination of those who do not meet minimum required qualifications (e.g., a college degree, training in chemistry) and are used for selection into all types of occupations.

Several authors have examined the rationale for using systematic, formal evaluations of past experience and training as an applicant assessment method (Arvey, McGowen, & Horgan, 1981; Beardsley, 1976; Johnson, Guffey, & Perry, 1980; Mosier, 1946; Myers & Fine, 1981). Mosel (1952) argued that past experience and training provide evidence that job-related knowledges, skills, and abilities (KSAs) have been acquired or already are possessed. Schmidt, Caplan, et al. (1979) examined traditional T&E rating methods, which attempt to measure KSAs by crediting amount of education and experience. They reasoned that T&E rating methods are based on two assumptions. First, that the amount and "quality" of education and experience are indirect measures of KSAs, and second, that KSAs are correlated with job performance. While plausible, these assumptions lead one to predict relatively low validities. This framework assumes that the validity of the T&E rating is the product of the correlation between the rating and KSAs, on one hand, and the correlation between KSAs and job performance, on the other. Schmidt, Caplan, et al. (1979) stated that T&E ratings are unlikely to correlate more than about .40 with KSAs and KSAs are unlikely to correlate more than about .50 with job performance. Thus, the final validity coefficient of a traditional T&E rating is estimated to be, at best, about .40 × .50, or .20.

Training and Experience Rating Methods

Although T&E evaluation methods vary widely, they may be grouped into several categories. R. Ash (1981) summarized five approaches to T&E

evaluations: point, task, behavioral consistency, grouping, and the knowledge, skill, ability (KSA) methods. The point method of T&E evaluation uses a formula that determines an applicant's score by assigning points for years of specified training or experience. Different types of training or experience are assigned point values depending upon their judged worth. This rating method is essentially credentialistic. It assesses the amount of education and experience rather than focusing specifically on past achievements and accomplishments during the course of that experience and education. The point approach is the T&E method most frequently used in government at all levels (R. Ash, 1981).

The grouping method of T&E evaluation is a variant of the point approach. The grouping approach classifies applicants into several qualification categories, such as well qualified, qualified, or not qualified, based on consideration of their training and experience. Applicants within each group are assumed to be equally suited for employment. The authors located no validity coefficients for the grouping method that met our decision rules (described below).

The task method of T&E assessment evaluates applicants on the basis of their experience with job-specific tasks. Typically, applicants rate their experience or skill at each task. Different applications of the task method vary in their scoring procedures. While the typical use of the task method presents the applicant with a list of task statements and requests self-report data on each task, the type of self-report data varies. Lyons (1984) asked applicants to assess the relative amount of time spent performing each task. Ocasio (1983) requested applicants to assess their performance level on each task using a scale ranging from "unacceptable" to "outstanding." Anderson, Warner, and Spencer (1984), R. Ash (1981), and Farrell (1979) used scales measuring the amount of supervisory assistance or additional training one needs to perform the task. For example, R. Ash (1981) used a rating scale ranging from "have not performed the task" to "delegated task to subordinates and coordinated/reviewed/directed their performance." On the basis of job analysis information, some tasks may be weighted more than others, or all tasks may be given equal weight. This is the only T&E method in R. Ash's (1981) typology that is based entirely on self-ratings. Some authors have advocated the adjustment of self-ratings based on lie or overestimation scales (Anderson et al., 1984; Farrell, 1979).

The behavioral consistency approach to T&E evaluation was developed at the U.S. Office of Personnel Management (Schmidt, Caplan, et al., 1979). It requires applicants to describe their major achievements in several job-related areas. These job-related areas are those behavioral dimensions rated by experienced supervisors as showing maximal differences between superior and minimally acceptable performers. The applicant's achievement statements are evaluated using anchored rating scales for which the

anchors are achievement descriptors whose values along a behavioral dimension have been reliably determined by several subject matter experts. Hough, Dunnette, and Keyes (1983) and Hough (1984) have suggested that achievement scores should be adjusted by partialling out years of experience in order to enhance validity.

The most common KSA approach to T&E evaluation is an application of the job element method. The job element method (P. Ash, Taylor, & Hoel, 1973; Primoff, 1975) is a job analysis method used to develop personnel selection instruments and is therefore more than a T&E method, although T&E evaluations may be used to assess applicant standing on job elements (KSAs). Originally developed at the U.S. Office of Personnel Management, the method uses brain-storming job analysis meetings of experienced supervisors to generate statements of relevant KSAs. These subject matter experts rate the suggested KSAs for criticality to job performance. In a job element selection procedure, a variety of measurement procedures can be used to assess an applicant's job-related KSAs. When a job element measurement instrument involves an evaluation of education and experience, the instrument may be classified as a T&E procedure. The Primoff studies we located used T&E scoring methods approximating, in some cases, the behavioral consistency method and, in others, the point method.

P. Ash et al. (1973) proposed an approach to job element examination that is used at the University Civil Service System of Illinois. A number of validity studies have been conducted on examinations developed with this method (Benz, 1974a, 1974b, 1974c; London, 1975, 1976a, 1976b; Marusarz, 1974; Taylor & Zrout, 1974; Zrout, 1973). As in the Primoff (1975) approach, the Ash approach allows a variety of measurement methods to assess an applicant's job-related KSAs. The Illinois civil service examinations included T&E evaluations, written tests, interviews, and performance tests.

Although the point method is sometimes used in the Illinois examinations, the primary T&E review is a scored "biographic element." We will call this T&E method the Illinois job element approach. In this approach, applicants describe their work experience and then provide a self-assessment on a KSA dimension. A personnel staff member reviews the job experience description to determine if it supports the self-rating. If so, the self-rating is the applicant's score on the biographic element; if not, the applicant is given no credit.

The Validity of T&E Ratings Determined by Traditional Methods

The use of T&E ratings has traditionally been justified through content validity arguments (Beardsley, 1976; Cobb, Spool, & Pollock, 1974; Levine

& Flory, 1975; MacLane, 1982; Maslow, 1968; Porter, Levine, & Flory, 1976; Primoff, 1975; Sage, Cole, & Johnson, undated; Schmidt, Caplan, et al., 1979; State of Connecticut, 1978). There are only a few literature reviews of the criterion validity of T&E examinations. Schmidt, Caplan, et al. (1979) summarized much of the available literature and found the average uncorrected validity coefficient to be approximately .10. Hunter and Hunter (1984) reviewed the studies located by Schmidt, Caplan, et al. (1979) that involved traditional T&E methods, and they offered .13 as the average validity coefficient after correction for unreliability in supervisors' ratings. Most of the studies summarized by Schmidt, Caplan, et al. (1979) employed the point method.

MacLane (1982) has reviewed the criterion validity of selection procedures developed using the Primoff job element procedure. Two of the studies investigated multiple selection instruments including both T&E evaluations and written tests (Acuff, 1965; Ebright, 1959). In one, the job elements were measured solely by the T&E evaluations (Primoff, 1958).

Meta-Analysis as a Method of Determining Validity

The narrative literature review method used by Schmidt, Caplan, et al. (1979) and MacLane (1982) can be informative, but it is not the optimal method of integrating research findings. In recent years, a set of methods has evolved that allows for quantitative cumulation of results across studies. These methods, collectively called "meta-analysis," facilitate the development of accurate conclusions about validity based on a body of past studies (Hunter, Schmidt, & Jackson, 1982).

The Schmidt-Hunter meta-analytic method is based on the hypothesis that much of the variation in results across studies may be due to statistical and methodological artifacts rather than to substantive differences in underlying population relationships. Some of these artifacts also reduce the correlations (or effect sizes) below their true (e.g., population) values. The method determines the variance attributable to sampling error and differences between studies in reliability and range restriction and subtracts that amount from the total amount of variation, yielding estimates of the true variation across studies.

In the context of personnel selection, meta-analysis is used to evaluate the situational specificity of a selection procedure. If one assumes that the validity of a test or test type is dependent upon the situation in which it is used, one would expect that the observed variation in validities cannot be accounted for by variance attributable to statistical artifacts. Meta-analytic results typically address this situational specificity hypothesis by reporting the percentage of observed variance in the validity distribution that is attributable to statistical artifacts. Another way of expressing the

variation in the true validities is to report the validity value at or above which 90% of all estimates of true validity lie (the 90% credibility value).

This meta-analysis of formal T&E evaluation method validities is different from meta-analyses of ability constructs (e.g., verbal ability) in four major ways, and these differences will be shown to be important for the interpretation of meta-analytic results. Each T&E evaluation method is a measurement method—as is a paper-and-pencil test. When one meta-analytically summarizes the validity of paper-and-pencil tests, the analyses are conducted separately for the different constructs measured by the tests. Separate meta-analyses are performed because the construct distinctions are psychologically meaningful and because different constructs may have different correlations with performance. Like paper-and-pencil tests, T&E evaluations may measure different constructs (e.g., cognitive ability, interpersonal skills, manifest motivation). Unfortunately, there is no research on the constructs measured by T&E evaluations. While the present research partitions T&E evaluations by method, the data only permit a gross content division, and it is recognized that within each method category, heterogeneous groups of constructs are measured. Thus, the present research does not provide detailed information on the validity of specific constructs measured by T&Es; rather, it provides useful information on the validity of T&E evaluation as a (multi-construct) measurement method. That is, it is the validity of the methods, rather than the constructs or construct measures, that is evaluated. Furthermore, while information on constructs would be useful for advancing our knowledge about the prediction of human performance and in the design and refinement of new T&E methods, the present research has an immediate practical benefit for those who must evaluate formal T&E methods as selection tools. This is important, given the widespread use of T&E methods.

The present research also differs from most previous validity generalization work on ability measures because there is more variability in how T&E data are collected than there is in how ability data are collected. While different paper-and-pencil measures of a given ability may vary slightly in the measurement process (e.g., they may use different item types), the measurement process across written tests for a given ability is very similar. In contrast, T&E evaluations vary widely in data collection processes. Some T&E evaluation data are obtained from resumés. Other data are gathered from traditional job application forms. Still other T&E data are gathered with very structured supplemental application forms. While the T&E evaluation method categories used in this analysis are meaningful ones, we recognize that the categories are not perfectly homogeneous. For example, among task method evaluations, some scoring schemes focus on time spent performing the task, while others focus on self-assessments of skill with the task.

A third distinction between the present meta-analysis and meta-analyses of test validity coefficients is the heterogeneity of the occupations in the analysis. Most validity generalization studies have been conducted on data drawn from a specific job classification. These classifications have been based on either job content (e.g., secretaries, police officers, petroleum workers) or job attributes such as the level of cognitive demands placed on an employee (Gutenberg, Arvey, Osburn, & Jeanneret, 1983; Hunter, 1980). Classification schemes based on job content may typically provide more control over sources of validity variance caused by job attributes because many job attributes have little or no variance within a job family. This control is gained at the sacrifice of detailed information about the attributes of jobs that may moderate validity. In contrast, job attribute classifications provide a better opportunity to discover why the validity of a given measure varies across jobs. However, the success of a job attribute analysis is dependent upon reliable measurement of the attribute and explicit hypotheses regarding the attribute. Furthermore, the effect of any given job attribute on validity may be obscured by the effects of other unmeasured job attributes. Too few validity studies have been conducted on T&E measures to permit separate meta-analyses by job attributes or job content category.

The first three of the four major differences between the present research and most past validity generalization studies suggest that the present research has less control over three sources of variance: heterogeneity of the (1) constructs measured, (2) measurement process, and (3) occupational categories. These three sources of uncontrolled variance make conclusions regarding the evaluation of the situational specificity hypothesis and the extent of validity generalization more conservative than in past validity generalization studies. In past evaluations of the situational specificity hypothesis, the construct being measured, the measurement process, and occupational category were held constant. Thus, any variance remaining after correcting for statistical artifacts could be attributed to situational or other moderators and uncorrected-for artifacts. In the present research, the variance remaining after correcting for statistical artifacts could be due to these sources or to differences among studies in (1) the constructs measured by the T&E evaluation, (2) the T&E evaluation measurement process, and (3) uncontrolled job attribute and job content differences. Thus, relative to past meta-analytic summaries of validity data, the interpretation of the situational specificity results in the present research will be ambiguous if the validity variance remaining after correction for artifacts is substantial. That remaining variance may be due to a situational moderator, but other causes of this remaining variance cannot be rejected.

The first three of the four major differences between the present research and past validity generalization studies will also affect how one can interpret

the validity generalization results. Validity generalization can be defined in many ways (Pearlman, 1982). The present research concludes that a measure shows generalization of validity when the lower 90% confidence value of the true operational validity is above zero (Callender & Osburn, 1981). This lower bound is dependent on the amount of variance remaining after correction for statistical artifacts.

The final difference between the present research and past validity generalization studies is the number of validity coefficients available for analysis. While we have assembled a substantial number of coefficients (132), the number is small relative to some past validity generalization research. Furthermore, when we divide the coefficients into formal T&E method categories, we are left with fewer coefficients to analyze. This leaves our results open to distortion caused by second-order sampling error (Schmidt, Hunter, Pearlman, & Hirsh, 1985, Q&A No. 25). The most commonly discussed form of sampling error is first-order sampling error. This is the random sampling error that affects individual correlation coefficients. It is a function of sample size and the size of the population correlation. The greater the sample size and the greater the absolute value of the population correlation, the smaller the expected sampling error. Second-order sampling error is conceptually similar to, yet distinct from, first-order sampling error. As noted by Schmidt et al. (1985, Q&A No. 25), the outcome of any meta-analysis depends to some extent on which studies happen to be randomly available. This is true even if the studies analyzed are all that exist at that point in time. This phenomenon, called "second-order sampling error," is a function of the number of studies analyzed. Thus, particularly with meta-analyses based on a small number of studies, the estimates of the population mean and variance may be higher or lower than the actual population values due to random error. As with first-order sampling error, second-order sampling error has a greater distorting effect on the variance estimates than on the mean estimates. Thus, in the present study, some of the estimates of population means and standard deviations may differ from actual values because of the small number of coefficients on which those analyses were based.

General Design of this Study

In the present study, meta-analytic techniques were applied to all known validity coefficients of training and experience evaluations. The situational specificity hypothesis was examined for these coefficients, and the generalizability of the validities estimated. In addition to examining the set of studies as a whole, four of the T&E methods described by R. Ash (1981) were examined separately to determine if the means and standard deviations of the distribution of validity coefficients varied as a function of the

type of T&E method (no validity coefficients were found for the grouping method).

Method

Validity coefficients for T&E ratings were obtained by an extensive literature search, beginning with a review of references cited in the Schmidt, Caplan, et al. (1979) study. The earliest published validity study of T&E evaluations that was located was Mosel (1952). The *Social Science Citation Index* was then used to locate articles that cited the Mosel article. No articles containing validity coefficients for T&E measures were found. Unpublished literature was found by reviewing the *International Personnel Management Association Assessment Council (IPMAAC) Proceedings*, and by searching the IPMAAC research library. Additional reports of criterion validity studies were located by contacting public-sector testing consortia, individuals cited in reference sections of reports, authors of T&E manuals, and authors referenced by other contacts or reports. Primoff's job element studies were located in files at the U.S. Office of Personnel Management.

A total of 132 validity coefficients, based on 12,048 observations, were located. Appendix A displays information about these studies. The observation count of 12,048 is based on the adjusted sample size resulting from the decision rules used to establish the data base for this analysis. The actual sample size was slightly smaller (as explained below).

Determination of Coefficients to be Included

The criterion of interest was a measure of overall job performance, rather than any specific aspect of performance. Most of the criteria were performance ratings (see Appendix A), typically task- and duty-based multiple scale instruments completed by one supervisor. The correlation with the sum of the task or duty ratings was used if presented or computable (see below). If neither of these was available, the correlation with a single-scale rating of overall job performance was used. If the criteria were multiple rating scales measuring different aspects of job performance, the correlation between the sum of the ratings and the T&E evaluation (Ghiselli, Campbell, & Zedeck, 1981) was computed and used when the correlations among the criteria were known. This was the case for 20 coefficients. Such a composite correlation was preferred over a mean coefficient because its sampling error was known (See Hunter et al., 1982, Chapter 5.). The two coefficients from the Acuff (1965) study were each based on the correlation between a predictor composite (i.e., the sum of T&E evaluations over several dimensions) and a criterion composite.

Thirty-one validity studies were obtained for examinations developed with the Illinois job element approach. As noted earlier, some of these elements were measured with T&E reviews. Other elements were measured with other assessment methods, including aptitude and performance tests. These were not used. Each of the 31 validity studies used criterion ratings based on the same performance evaluation instrument. In eight of the studies, the intercorrelation matrix of the criterion scales was provided. The sample-size-weighted means of the elements of these eight intercorrelation matrices were used to derive an average intercorrelation matrix. This matrix was then used in computing composite correlations for the Illinois studies that did not report a criterion intercorrelation matrix. The Illinois examination studies were reviewed to identify validity coefficients for T&E reviews; usable data were obtained from 19 of the 31 studies. Three studies reported point method validities. Sixteen of the 31 studies reported validity data for the Illinois job element T&E approach. For eight of the studies, the reported correlation was a composite between a predictor score and a composite criterion.

The remaining eight coefficients from the Illinois T&E distribution were obtained from examinations where some elements were measured by a T&E review and others were measured by another instrument (e.g., aptitude tests). For these studies, the number of elements measured by T&E predictors ranged from two to five. A composite correlation could not be computed because the intercorrelations among the T&E predictors were not known. For each T&E predictor in a study, a composite correlation was computed between the single predictor and a composite criterion. The mean of the composite validity coefficients in each study was reported as the study's validity coefficient. The exact sampling error of such a mean correlation is not known (Hunter et al., 1982). The sampling error of these coefficients was estimated conservatively by using as the sample size the original sample size times the number of coefficients that were averaged to compute the reported coefficient.

When multiple criteria were based on separate job performance measures, the correlations were reported separately. For these coefficients, sample size was reported once for each coefficient. Four jobs in the study by McKinney and McCormick (1976) had two criterion measures. When a phi or point-biserial correlation was reported, the coefficient was corrected to the value expected had the sample sizes been equal in the two groups (Hunter et al., 1982). This adjustment was made for six coefficients (Haynes, undated; McKinney & McCormick, 1976). This adjustment increases the sampling error in the correlation. To estimate the sampling error in the analysis properly, the sampling error variance of these coefficients was multiplied by the squared ratio of the corrected to the uncorrected coefficient (see Hunter et al., 1982, chap. 3). Appendix A identifies the

coefficients that are composites, are adjusted, or are based on a second criterion. The type of criterion for each coefficient is also listed in Appendix A.

Assignment of Coefficients to Distributions

Correlations included in the analysis were grouped into the categories of point, task, behavioral consistency, and Illinois job element on the basis of type of T&E method employed, and each category was analyzed separately. The type of T&E method used in a study was usually apparent. Since the vast majority of T&E procedures are point method systems, a study was assigned to the point method category when no method was specified. Two coefficients in the task category distribution should be noted. For the task method coefficient reported by Farrell (1979), it is unclear whether the coefficient was derived from raw T&E scores or from scores that were adjusted for overestimation. Farrell proposed both rationally constructed and empirically derived methods to correct the scores of applicants who overestimated their ability to perform job tasks. Anderson et al. (1984) reported coefficients for both raw and adjusted data; the coefficient reported for the raw data was used in this analysis. One correlation (for an attorney position) in the behavioral consistency category (Hough, 1984; Hough et al., 1980; Hough et al., 1983) was based on a score that was adjusted for years of experience. This correlation was used because the coefficient derived from the unadjusted data was not reported. The remaining coefficients from the Hough et al. (1980) study were zero-order validities.

The six validity coefficients for T&E examinations based on Primoff's job element procedures were assigned either to the behavioral consistency distribution or to the point distribution, depending upon the manner in which applicant information was rated. The studies assigned to the behavioral consistency method were Primoff (1958), Haynes (undated), and Acuff (1965). In this analysis we have used a validity coefficient for the Primoff et al. (1958) study that is higher than that reported by Schmidt, Caplan, et al. (1979) (.59 vs. .42) because it is a composite based on two performance appraisal ratings; Schmidt, Caplan, et al. (1979) reported the validity coefficient for the first performance rating only. Schmidt, Caplan, et al. (1979) reported only one estimated validity coefficient for the Haynes study (undated), but the present analysis included two coefficients from this study (one for each of the two jobs). A job element study (Ocasio, 1983) was assigned to the point distribution on the basis of a description of the

scoring process (B. P. Ocasio, personal communication, January 20, 1984). The study by Groll (1975) was also assigned to the point category.[1]

R. Ash (1981) reported validity coefficients from four T&E methods for each of three jobs. The study used a simulated criterion: predicted performance as a supervisor as determined from peer nomination rankings. Since it was not a measure of actual job performance, the 12 validity coefficients were not included in the analysis.

Analysis of the Distributions

In addition to the distribution of the total population of coefficients and the four distributions grouped by T&E type, two additional distributions of coefficients were analyzed. The purpose of these additional analyses was to examine the effects of outlier data errors on meta-analytic results. Tukey (1960) has argued that all real data distributions contain erroneous data, which when located at the tails of the distribution, severely distort the variance estimate and, to a lesser degree, the mean estimate. Tukey has suggested that the top and bottom 5% of any distribution be truncated to minimize the disproportionate influence of outliers. Following Tukey's suggestion, a truncated distribution of the total population of coefficients and a truncated version of the point distribution were prepared. Truncated versions of the remaining distributions were not feasible because of the small number of studies in the distributions.

Given that most of the cumulated studies lacked criterion reliability and range restriction data, it was not possible to correct each coefficient individually for the effects of these factors. Instead, meta-analytic procedures based on assumed distributions of artifact values were used. Two artifact distributions constructed by Schmidt, Hunter, Pearlman, and Shane (1979) were utilized in the present research. The distribution of predictor reliabilities was that used by Schmidt, Hunter, et al. (1979) for tests of cognitive ability. The review by Schmidt, Caplan, et al. (1979) found interrater reliabilities in the .80s to be typical of point system T&E methods. These reliabilities were the correlations between evaluators scoring the same applicant responses (i.e., conspect reliability, see Cattell, 1971). A more appropriate (and probably lower) estimate would have been obtained if applicants had retaken the T&E instrument. However, such reliability estimates were not available. Interrater reliabilities of other T&E methods

[1] Researchers closely associated with the Primoff job element research program did not agree with our assignment of these studies. They thought that all studies using the Primoff job element method of job analysis should be placed in a separate category (job element category), regardless of the resulting T&E instrument. We judge this to be inappropriate, since it would lead to a very heterogeneous category of T&E methods that would cut across the T&E categories.

TABLE 1

Assumed Distribution of Range Restriction Across Studies

Prior selection ratio	*SD* of test	Relative frequency
1.00	1.000	5
.92	.854	11
.82	.770	16
.72	.708	18
.62	.655	18
.52	.607	16
.42	.563	11
.32	.519	5

Note: Expected value (*SD*) = .70.

may be lower. For the reasons given in Pearlman, Schmidt, and Hunter (1980) the validity distributions were corrected for validity variance due to study differences in predictor reliability, but they were not corrected for mean predictor unreliability.

The distribution of criterion reliabilities is also from Schmidt, Hunter, et al. (1979). The mean reliability of .60 is based on King, Hunter, and Schmidt (1980), which showed that the mean interrater reliability of performance ratings was .60 if the individual supervisor's judgment was measured with perfect reliability. Since performance rating instruments do not have perfect intrarater reliability, this reliability distribution is probably an overestimate of the true mean reliability in these studies. Thus, criterion reliability corrections based on this distribution are likely to underestimate the true mean validity.

The data were analyzed twice, once with range restriction corrections and once without. In developing the range restriction artifact distribution, the authors reviewed Hunter (1980). Hunter examined empirical data on range restriction in over 400 U.S. Employment Service validity studies from the private sector and found that the average range restriction value (ratio of restricted to unrestricted test standard deviation) was .67 for measures of cognitive ability. On the basis of this finding, an assumed distribution of range restriction values was constructed. Table 1 shows this range restriction distribution, which has an average standard deviation of .70. The figures in the range restriction distribution correspond to a mean selection ratio of 70% (Schmidt, Hunter, & Urry, 1976). Ordinarily, selection ratios are far lower in government hiring, averaging approximately 15% or less (Schmidt, Hunter, Outerbridge, & Trattner, 1986). Thus the distribution may underestimate the actual level of range restriction. If so, the result would be an underestimation of true validities.

It may be argued however, that it is inappropriate to correct for range restriction when no empirical data on range restriction is available specifically for T&E measures. To address this potential criticism, we repeated all analyses with no range restriction corrections. The results from these

analyses are very conservative (i.e., underestimate true validity), since there is almost certainly a substantial amount of range restriction in these data.

This study employed the interactive validity generalization equations described in detail in Schmidt, Gast-Rosenberg, and Hunter (1980). The computer program used incorporated the structural equations given in the appendix of that article. The result was a computation algorithm that differed in trivial ways from the computational method used by Schmidt et al. (1980). These differences are described in Appendix B.

Results and Discussion

The results of the primary analyses to assess situational specificity and validity generalization for each T&E distribution are presented in Tables 2 and 3.

Situational Specificity

The analyses addressing situational specificity are presented in Table 2. These analyses address the question of whether artifacts explain all of the variance in the observed validity coefficients. While a variety of research methods have addressed (and disconfirmed) this hypothesis for employment tests (Schmidt, Hunter, et al., 1985), a common evaluation method involves the application of the 75% rule. Schmidt and Hunter (Hunter et al., 1982) argued that if 75% or more of the observed variance in a distribution of validities can be explained by typically corrected-for statistical artifacts (i.e., sampling error and differences among studies in measurement error and range restriction), one should conclude that a situational moderator is not present and that the unaccounted-for variance is likely to be due to uncorrected-for artifacts. Recent research, however, has noted problems in the interpretation of the 75% rule (McDaniel & Hirsh, 1986; McDaniel, Hirsh, Schmidt, Raju, & Hunter, 1986). The percentage of observed variance accounted for by artifactual variance is a function of sample size. As the mean sample size in the meta-analysis increases, the percentage of observed variance accounted for by sampling error variance decreases. Thus, with a moderator of a constant magnitude, a meta-analysis with a small average sample size could find a large percentage of variance due to sampling error, while a meta-analysis with a large average sample size, but with the same amount of residual variance, would find a smaller percentage of variance due to sampling error.

A preferred method of determining the presence of a moderator is a direct examination of the residual variance (or residual standard deviation). If a moderator is affecting the correlation, it will cause variance in the distribution. A useful yardstick for the size of the residual standard deviation

TABLE 2

Situational Specificity Hypothesis Results for T&E Methods Using Two Sets of Artifact Distributions

					Includes range restriction corrections			No range restriction corrections		
T&E method distribution	Total *N*	No. *r*'s	Mean *r*	Obs. *SD*	Pred. *SD*	% var. acc. for	Res. *SD*	Pred. *SD*	% var. acc. for	Res. *SD*
All cases	12,048	132	.090	.157	.102	42	.119	.101	41	.120
All cases (truncated)	11,607	119	.092	.128	.101	62	.078	.100	61	.080
Point	6,741	91	.058	.171	.113	44	.129	.113	43	.129
Point (truncated)	6,362	81	.060	.137	.111	66	.080	.111	66	.080
Illinois job element	3,168	16	.106	.067	.073	119	.000	.071	113	.000
Task	991	10	.082	.175	.097	31	.145	.096	30	.146
Behavioral consistency	1,148	15	.248	.124	.112	82	.052	.106	73	.064

in validity generalization studies can be found in McDaniel et al. (1986). Depending on the category of predictor examined, the residual standard deviation across several meta-analyses ranged from .053 to .108. Thus, we can compare these values with those obtained in the present research to make judgments about the relative magnitude of any potential situational moderator. As stated earlier, the present research has less control over three sources of validity variance (i.e., heterogeneity of constructs measured, measurement process, and occupational categories). Thus, even in the absence of a situational moderator one would expect these results to show larger residual standard deviations. The first column in Table 2 identifies the T&E distribution analyzed. The next five columns of data present the total sample size and number of validity coefficients on which each distribution was based, the uncorrected mean and observed standard deviation of each distribution, and the standard deviation predicted on the basis of the four artifacts for which corrections were made [sampling error and differences among studies in (1) reliability of the predictor, (2) reliability of the criterion, and (3) range restriction], the percentage of observed variance in the distribution accounted for by these artifacts, and the residual standard deviation. The data in the next three columns show the comparable statistics when no range restriction corrections were made.

As expected from previous research (McDaniel et al., 1986; McDaniel & Hirsh, 1986), the standard deviation predicted on the basis of sampling error is a strong monotonic function of the average sample size. While the percentage of variance accounted for by sampling error varies widely across analyses, the residual standard deviation shows less variability across analyses. Since the residual standard deviation reflects the amount of variance attributable to moderators and artifacts not corrected for and is not affected by the average sample size in the analysis, one can compare the residual standard deviations obtained in this study with those of past meta-analytic studies to determine the relative magnitude of unexplained variance. A comparison of the range reported in McDaniel et al. (1986) with the values obtained in this study indicate that, after corrections, some distributions are more variable and some less variable than those in past validity generalization research. In particular, the residual standard deviations are quite small for the behavioral consistency (.052) and the Illinois job element (.000) methods, indicating that there may be no nonartifactual validity variance (i.e., situational specificity) for these two methods. On the other hand, the value for the task method (.145) is considerably larger. The results of these analyses should be interpreted with caution due to the small number of coefficients in each distribution.

Validity Generalization

Validity generalizability may be defined in several ways (Pearlman, 1982). Using the most conservative definition, validity generalization may be applied to a measure when the true variance of the distribution of its validities equals zero. In that case there is no room for variables to moderate the relationship. By a second definition, validity may be generalized when most of the true validities in the distribution are greater than a minimum useful validity. The variance remaining in the distribution may result from moderators or from uncorrected artifacts, but this remaining variance is sufficiently small to allow the measure to be valid for the vast majority of its applications. For the purposes of this study, predictors will be said to have generalizable validity if the value at the lower 10th percentile of the distribution of estimated true validities is greater than zero (Callender & Osburn, 1981). Thus, our definition of validity generalizability is directly analogous to significance testing (in significance testing, a correlation is statistically significant when the lower bound of its confidence interval does not include zero).

While a validity coefficient that is only slightly greater than zero is meager relative to that found for ability tests or other assessment measures, it can have more practical utility than random selection. This fact would be especially relevant where selection methods with higher validity and utility (e.g., ability tests) could not be used for some reason (e.g., a court decision). Also, a measure with a low validity may sometimes provide a useful increment to the validity of a selection battery.

Table 3 presents the validity generalization results. The first column of the table identifies the T&E distribution analyzed. The next two columns of data show the sample size and the number of validity coefficients in each distribution. Columns 4, 5, and 6 present the estimated mean $(\bar{\hat{\rho}})$, standard deviation $(SD_{\hat{\rho}})$ and 90% credibility value for the distribution of true validities as computed using the set of artifact distributions that include range restriction corrections. Columns 7, 8, and 9 present the comparable data for the set of artifact distributions that do not include range restriction corrections.

For the first set of analyses (i.e., those reported in columns 4, 5, and 6), including those involving truncated distributions, the mean true validity is corrected for range restriction and unreliability in the criterion. As well, for the first set of analyses, the variance of the true validity distribution is corrected for sampling error and for differences among studies in predictor reliability, criterion reliability, and range restriction. For the second set of analyses (i.e., those reported in columns 7, 8, and 9), including those involving truncated distributions, the mean true validity is corrected for criterion reliability only. The variance of the true validity distribution for

TABLE 3

Validity Generalization Results for T&E Methods Using Two Artifact Distributions

T&E method distribution	Total N	No. r's	Includes range restriction corrections $\bar{\hat{\rho}}$	$SD_{\hat{\rho}}$	90% C.V.	No range restriction corrections $\bar{\hat{\rho}}$	$SD_{\hat{\rho}}$	90% C.V.
All cases	12,048	132	.17	.22	−.12	.12	.16	−.08
All cases (truncated)	11,607	119	.17	.15	−.01	.12	.10	−.01
Point	6,741	91	.11	.24	−.20	.07	.17	−.14
Point (truncated)	6,362	81	.11	.15	−.08	.08	.11	−.06
Illinois job element	3,168	16	.20	.00	.20	.14	.00	.14
Task	991	10	.15	.27	−.19	.11	.19	−.14
Behavioral consistency	1,148	15	.45	.10	.33	.32	.08	.22

Note: $\bar{\hat{\rho}}$ = estimated mean population correlation coefficient; $SD_{\hat{\rho}}$ = estimated population standard deviation; 90% C.V. = 90% credibility value.

the second set of analyses is corrected for sampling error and for differences among studies in predictor reliability and criterion reliability.

The results with and without range restriction corrections are very similar, although the first set of analyses yielded somewhat higher true validity estimates than the second set. Because the same conclusions would be drawn from either set of analyses, only the first set will be discussed.

The meta-analysis of the distribution of all validity coefficients yielded a mean true validity of .17, with a standard deviation of .22. These results indicate that T&E ratings as a whole have only moderate mean validity that cannot be generalized across situations since the value at the 10th percentile of the distribution was negative (−.12). When the truncated distribution of all cases was analyzed, the situation was somewhat better, with a mean true validity of .17 and a standard deviation of .15. The 90% credibility value was not above zero (−.01). An inspection of the results for specific T&E methods reveals that some of the variation in the distribution of all coefficients apparently resulted from the type of T&E method employed.

T&E method, however, is not a moderator. A moderator variable is one whose value is correlated with the relationship between two other variables. T&E method is not a third variable; it is a variant of the predictor. To further illustrate this distinction, consider the findings of Pearlman (1979) in the prediction of clerical performance. Pearlman found the validity of verbal ability to be .40, while the validity of perceptual speed is .48. One would conclude that one predictor is better than another; one would not conclude that test type is a moderator.

The point method studies, which constitute 69% of all known validity coefficients for T&E ratings, show a mean true validity of .11 and a

relatively large standard deviation of .24. The mean true validity of the truncated point distributions is .11 and the standard deviation is .15. The 90% credibility values for the point and truncated point distributions are −.20 and −.08, respectively. Thus, the point method has a low mean validity and lacks generalizability. The mean validity of the truncated distribution (.11) is lower than the figure of .20 advanced by Schmidt, Caplan, et al. (1979) as an *upper limit* on the validity of traditional (i.e., point method) T&E evaluations. Thus our estimated true validity of .12 is not inconsistent with the Schmidt et al. prediction.

The Illinois job element and behavioral consistency methods showed larger validities. The corrected standard deviation and 90th credibility values of the validities suggest that the validities are generalizable; however, these findings are based on distributions of relatively small numbers of studies. As discussed earlier, distributions with few coefficients allow more scope for the operation of second-order sampling error, which can distort true validity variance estimates and, to a lesser extent, true mean validity estimates. As more studies become available in the future, the analyses should be rerun to permit firmer conclusions about the mean level of validity and the extent of validity generalization.

The Illinois job element method showed substantially better results than the point method. A mean true validity of .20 with a standard deviation of zero indicated a useful degree of validity and suggested that the method shows validity generalization. The task method distribution had a mean true validity of .15, with a standard deviation of .27. The task method validity distribution did not meet the 90% credibility value criterion for validity generalization. The behavioral consistency method yielded a mean true validity of .45 with a standard deviation of .10. From initial results, it appears that both the Illinois job element and the behavioral consistency methods are far superior in predictive validity to the traditionally used point method. While the mean validity of the task method is superior to the point method, no support for the generalization of task method validities was found.

Five of 15 validities in the behavioral consistency distribution are based on the Primoff job element method, which predates the behavioral consistency method by many years (Primoff, 1958). Although the job element method differs in a number of respects from the behavioral consistency method (E. Primoff, personal communication, October 29, 1986; Schmidt, Caplan, et al., 1979), the behavioral consistency method incorporated many of the concepts underlying the job element method. The behavioral consistency method is very similar to some applications of the job element method. The critical characteristic shared by the 15 coefficients in the behavioral consistency distribution is the evaluation of individuals on the basis of past accomplishments and achievements rather than on credentials.

The task method validity can be compared to the results of a meta-analysis of self-evaluation of ability (Mabe & West, 1982) because, as noted earlier, task method T&E scores are based on self-ratings. In the Mabe and West study, ability and performance measures were used as criteria in evaluating self-ratings of ability (written ability tests, scores on typing tests, etc.).

Mabe and West (1982) report an estimated true correlation coefficient of .36, corrected for criterion reliability but not for restriction in range. Mabe and West judged their range restriction data to be too scanty to hypothesize an accurate range restriction distribution. The corresponding statistic from this study's task distribution is .11. If the range restriction effects in the two distributions are similar, then the true correlation between task ratings and job performance is smaller than the true correlation between self-ratings of ability and other measures of that same ability. Such an outcome appears reasonable. However, it is also likely that range restriction is greater in the present data and that this fact accounts for some of the difference.

This difference in the magnitude of relationships may be due in part to the lack of favorable measurement conditions that characterize self-ratings for personnel selection. Mabe and West (1982) found four conditions that enhance the validity of self-assessments. These are (1) expectation of self-evaluation verification, (2) self-evaluation instructions using social comparison terminology, (3) self-evaluation experience, and (4) instructions as to anonymity. Task methods vary in the extent to which the applicants expect self-evaluation verification. Some developers of task method measures request information that would permit verification of self-evaluations (Anderson et al., 1984; Farrell, 1979), while others do not. Typically, the last three conditions identified by Mabe and West (1982) are absent when self-ratings of task performance are collected for applicant screening. Although social comparison instructions (e.g., "How do you perform compared with others?") could be easily incorporated into task inventories, none of the analyzed task studies appeared to use them. Also, self-assessment experience of applicants is rarely under the control of the employer using the T&E method. Finally, since task ratings are used for hiring purposes, the applicant's self-ratings are not anonymous.

This study's task method results may also be compared with a review of self-ratings for personnel selection conducted by Reilly and Chao (1982). They report a sample-size-weighted mean uncorrected validity of .15 for a set of three self-assessment studies. Although this value is higher than the observed (uncorrected) validity of .08 for the task distribution in the present study, both correlations are low.

Both the point and task distributions showed substantial unexplained variance. McDaniel and Schmidt (1985) report a more detailed analysis of

TABLE 4

Analyses of Grade Level Moderator for the Point Method

T&E method distribution	Total N	No. r's	Mean r	$\bar{\rho}$	$SD_{\hat{\rho}}$	90% C.V.
All Molyneaux studies	3,744	51	.06	.11	.29	−.26
Grades 3–6	1,203	19	.16	.29	.11	.14
Grades 7–8	1,100	14	.05	.10	.32	−.31
Grades 9–12	1,441	18	−.02	−.03	.28	−.39

Note: Population estimates are corrected for criterion unreliability and range restriction; $\bar{\rho}$ = estimated mean population correlation coefficient; $SD_{\hat{\rho}}$ = estimated population standard deviation; 90% C.V. = 90% credibility value.

the point method. Using the same point method coefficients as in this study, plus six coefficients from the R. Ash (1981) study, the authors attempted to assess three potential moderators as possible sources of variance in the point method distribution. Neither study source, variation in scoring procedures, nor job attributes (e.g., job complexity) appeared to moderate point method validities.

Recent theoretical and causal modeling work by Schmidt, Hunter and Outerbridge (1986) provides a clue to the source of this variance. Schmidt, Hunter, and Outerbridge (1986) argue that it is *relative* individual differences in job experience that cause individual differences in job performance. They further argue that these relative individual differences in job experience decrease as the mean level of job experience in a sample increases. In brief, this theory predicts that, given constant variance of absolute job experience levels, the validity of job experience is highest in applicant pools where the mean level of job experience is low. As the sample's mean level of job experience increases, the validity of job experience is expected to decay. Strong empirical support for this decay in the validity of job experience validities is reported by McDaniel (1986) and McDaniel, Schmidt, and Hunter (in press). Since point method scoring strategies give substantial weight to the amount of job experience, one would expect the validity of point method evaluations to vary with the mean level of job experience in the study. Likewise in the case of the task method, we suspect that scores on this T&E method are at least moderately correlated with the length of job experience. The longer one works in an occupation, the more opportunity one will have to perform and gain skill at various tasks.

The 51 point method validity coefficients from Molyneaux (1953) permit a partial test of the job experience moderator hypothesis. Although mean length of job experience was not provided for any of his samples, Molyneaux did report the grade level for each of his samples. To the extent that grade level is positively correlated with job experience, the moderating effect of job experience should be reflected in varying levels of mean validity across grade level. Table 4 shows the results of meta-analyses to

test this hypothesis. The mean validity was substantially higher for the samples from the lower grade levels.[2]

If the validity of point and tasks methods decays with increasing levels of job experience, our earlier conclusions about the lack of validity generalization for these two methods could be too broad. The point method may be generalizable for lower-level jobs. Future research should examine whether point and task methods of training and experience evaluations show higher validities for applicant pools with low mean levels of job experience.

In addition to a potential job experience moderator of point and task method validities, content differences within each method are very likely to have contributed variance to the distributions. Some point methods emphasize experience; others emphasize educational credentials. To the extent that different content (construct) domains have different validities, the validity of a given point or task measure will vary with the emphasis placed on different content areas.

Some may contend that the results for the Illinois job element method and the point method may misrepresent the true validities for these methods because most of the results included in these analyses were reported by a few researchers (i.e., London, Molyneaux, and Mosel). For example, one may argue that studies done by one researcher may share systematic contaminants or follow a particular variant of methodology (use of certain rating dimensions, certain criteria, etc.) that may influence the results in some manner. This hypothesis is blunted by an inspection of the primary studies. Neither London nor Molyneaux conducted the studies that they reported. Specifically, they relied on archival data and did not develop either the T&E predictors or the job performance criteria for the studies that they described. Their involvement in the results they reported was limited to the calculation of the validity coefficients. Mosel relied on archival sources for the predictor data; it is not clear whether Mosel developed the criteria or relied on archival criteria. In brief, there is little support for researcher-specific contamination.

Summary and Conclusions

This report has analyzed the validity of T&E methods for personnel selection. Meta-analytic summaries revealed that validity varies with the type of T&E method used. Both the point method and the task method yielded low validity (.11 and .15) with substantial variance. Recent empirical research and theorizing on the effect of job experience on performance (McDaniel, 1986; McDaniel, Schmidt, & Hunter, in press; Schmidt, Hunter,

[2] We thank an anonymous reviewer for suggesting this analysis.

& Outerbridge, 1986) and moderator analyses presented in this study suggest that the point and task methods may show useful levels of validity for samples with low mean experience levels. The Illinois job element method yielded low-moderate validity (.20). The behavioral consistency method yielded the highest validity (.45). Based on existing data, the Illinois job element and behavioral consistency methods appear to display validity generalizability. Additional validity studies on all T&E methods are needed so that analyses may be rerun with larger distributions, and firmer conclusions can be drawn regarding mean validity and generalizability.

More research is needed to determine the constructs measured by T&E instruments. As Hunter and Hunter (1984) noted, the evaluation of alternative predictors often fails to distinguish two issues: specification of what is to be measured (content) and the method used for measurement. While T&E methods may be in part measures of abilities, they also may measure noncognitive traits that correlate with job performance. An understanding of the constructs measured by T&E evaluations and their relationships with cognitive and other constructs may suggest better measurement methods for T&E evaluations. It will also provide the information needed to assemble selection systems composed of T&E evaluations combined with other selection methods to maximize the prediction of job performance.

REFERENCES

*Acuff R. (1965). *A validity study of an aircraft electrician and electrical worker job element examination.* Pasadena, CA: U.S. Naval Laboratories in California, Test Development and Personnel Research Station.

*Anderson CD, Warner JL, Spencer CC. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology, 69,* 574–580.

Arvey RD, McGowen S, Horgan D. (1981). *The use of experience requirements in selecting employees.* Presented at the annual conference of the International Personnel Management Association Assessment Council, Denver, CO.

Ash P, Taylor N, Hoel L. (1973). The University Civil Service System of Illinois: Updating a merit system. *Public Personnel Management, 2,* 456–461.

Ash RA. (1981). Comparison of four approaches to the evaluation of job applicant training and work experience. *Dissertation Abstracts International, 42,* 4606B. (University Microfilms No. DE082-07909).

*Bean KL. (1958). When should an unassembled examination be used? *Public Personnel Review, 9,* 92.

Beardsley VA. (1976). *A study of the rating of education and experience as an examination method in the Pennsylvania State Civil Service Commission.* Commonwealth of Pennsylvania State Civil Service Commission, Bureau of Examinations, Research and Special Projects Division.

*Studies that contributed data to the meta-analysis are marked with an asterisk.

*Benz MP. (1974a). *Validation of the job element examination for admissions and records officer I.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

*Benz MP. (1974b). *Validation of the job element examination for kitchen laborer.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

*Benz MP. (1974c). *Validation of the examination for health service nurse.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

Callender JC, Osburn HG. (1981). Testing the constancy of validity with computer-generated sampling distributions of the multiplicative model variance estimate: Results for petroleum industry validation research. *Journal of Applied Psychology, 66,* 274–281.

Cattell RB. (1971). *Abilities: Their structure, growth , and action.* Boston: Houghton Mifflin.

Cobb CL, Spool MD, Pollock AL. (1974). *A flexible and comprehensive system for rating training and experience.* Presented at the meeting of the Southeastern Psychological Association, Atlanta, GA.

*Delaney EC. (1954). *Teacher selection and evaluation with special attention to the validity of the personnel interview and the national teacher examinations as used in one selected community (Elizabeth, New Jersey).* Doctoral dissertation, Columbia University, New York, NY.

*Ebright E. (1959). *A follow-up validity of an engineering draftsman examination developed by the J-element technique.* Pasadena, CA: U.S. Naval Laboratories in California, Personnel Research Section.

*Farrell BM. (1979). *Task performance self-evaluation: An alternative selection procedure to traditional experience and training ratings.* Minneapolis, MN: Selection Research Unit, Minnesota Department of Personnel.

Ghiselli EE, Campbell JP, Zedeck S. (1981). *Measurement theory for the behavioral sciences.* San Francisco, CA: W. H. Freeman Co.

*Groll MF. (1975). *Low skill manual labor classes validity study.* Boise, ID: Idaho Personnel Commission.

Gutenberg RL, Arvey RD, Osburn HG, Jeanneret PR. (1983). Moderating effects of decision-making/ information-processing job dimensions on test validities, *Journal of Applied Psychology, 68,* 602–608.

*Halinski L. (1974). *Validation of the job element examination for nursing assistant.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

*Haynes E. (Undated). *Tryout of job element exa ination for shipfitter.* Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center.

Hough LM. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69,* 135–146.

*Hough LM, Dunnette MD, Bartlett CJ. Goldstein IL, Keyes MA, King KW, Weiss SJ, Levine GS, Buxton VM. (1980). *Validation of selection and promotion procedures for FTC professional employees* (Institute Report No. 50). Personnel Decisions Research Institute.

Hough LM, Dunnette MD, Keyes MA. (1983). An evaluation of three "alternative" selection procedures. PERSONNEL PSYCHOLOGY, *36,* 261–276.

*Hough LM, Keyes MA, Dunnette MD. (1984). *Development and validation of personnel selection systems for eight Library of Congress professional series* (Institute Report No. 91). Personnel Decisions Research Institute and Personnel Decisions, Inc.

Hunter JE. (1980). *Test Validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery (GATB).*

Washington, DC: Division of Counseling and Test Development, Employment and Training Administration, U.S. Department of Labor.

Hunter JE, Hunter RF. (1984). The validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.

Hunter JE, Schmidt FL, Jackson GB. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage Publications.

*Johnson JC, Guffey WL, Perry RA. (1980). *When is a TE rating valid?* Presented at the annual conference of the International Personnel Management Association Assessment Council, Boston.

King LM, Hunter JE, Schmidt FL. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology, 65*, 507–516.

Levine EL, Flory A III (1975). Evaluation of job applications: A conceptual framework. *Public Personnel Management, 4*, 378–385.

*London M. (1975). *Validation of selection procedures developed by the University Civil Service System of Illinois: A summary of studies conducted during spring 1975.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

*London M. (1976a). *Summary of validity study findings for clerk II, typing clerk II, typing clerk III, clerk-stenographer II/ secretary II (stenographic), library clerk II.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

*London M. (1976b). *Summary of validity study findings for secretary (stenographic)/ secretary IV (stenographic), secretary (transcribing)/ secretary IV (transcribing), cashier II.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

*Lyons TJ. (1984). *Development and validation of the unassembled clerical exams for clerk typist (GS-4) and clerk stenographer (GS-5): A status report.* Washington, DC: U.S. Office of Personnel Management, Office of Staffing Policy.

*Lyons TJ. (1985). *Accountant/ auditor examination validity study: An interim report* (Draft). Washington, DC: U.S. Office of Personnel Management, Office of Staffing Policy. (Data updated via personnel communication to senior author, June 1986.)

Mabe PA, West S. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280–296.

MacLane CN. (1982). *Evidence relating to the utility of Primoff's job element method* (OPRD-82-2). Washington, DC: U.S. Office of Personnel Management, Office of Personnel Research and Development. (NTIS No. PB 82 233256).

*MacLane CN. (1986). Personal communication to senior author.

*Marusarz T. (1974). *Validation of the job element examination for janitor.* Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

Maslow AP. (1968). The unassembled examination. In Donovan JJ (Ed.), *Recruitment and selection in the public service* (pp. 255–275). Chicago: Public Personnel Association.

McDaniel MA. (1986). The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance. Doctoral Dissertation, The George Washington University. *Dissertation Abstracts International, 47*, AAD86–08356.

McDaniel MA, Hirsh HR. (1986). Methods of moderator detection in meta-analysis. In McDaniel MA (Chair), *An overview and new directions in the Hunter, Schmidt, Jackson meta-analysis technique.* Symposium presented at the first annual conference of the Society for Industrial/ Organizational Psychology, Chicago, IL.

McDaniel MA, Hirsh HR, Schmidt FL, Raju NS, Hunter JE. (1986). Interpreting the results of meta-analytic research: A comment on Schmitt, Gooding, Noe, and Kirsh (1984). PERSONNEL PSYCHOLOGY, *39*, 141–148.

McDaniel MA, Schmidt FL. (1985). A meta-analysis of the validity of training and experience ratings in personnel selection (OSP 85-1). Washington, DC: U.S. Office of Personnel Management, Office of Staffing Policy. (NTIS No. PB 86 109 618/ AS).

McDaniel MA, Schmidt FL, Hunter JE. (in press). Job experience correlates of job performance. *Journal of Applied Psychology*.

*McKinney TS, McCormick JM. (1976). *Experience/ interest measure as a determinant of job success* (TR-4-76). Employee Selection Research, Personnel Department, City of Phoenix, AZ.

*Molyneaux JW. (1953). *An evaluation of unassembled examinations*. Unpublished master's thesis, The George Washington University, Washington, DC.

*Mosel JN. (1952). The validity of rational ratings of experience and training. PERSONNEL PSYCHOLOGY, *5*, 1–10.

Mosier CI. (1946). Rating experience and training in public personnel selection. *Educational and Psychological Measurement*, *6*, 313–329.

Myers DC, Fine SA. (1981). *Development of a methodology to obtain and assess applicant experience for employment*. Presented at the annual conference of the International Personnel Management Association Assessment Council, Denver, CO.

*Ocasio BP. (1983). *Validation of the custodial unassembled examination procedure*. Washington, DC: U.S. Postal Service, Office of Test Development and Validation.

*Ocasio BP, Hillery JM. (1983). *An experimental analysis of self-assessment for employee selection*. Washington, DC: U.S. Postal Service, Office of Test Development and Validation.

Owens WA. (1976). Background data. In Dunnette MD (Ed.), *Handbook of industrial and organizational psychology* (pp. 609–644). Chicago: Rand McNally.

Pearlman K. (1979). *The validity of tests used to select clerical personnel: A comprehensive summary and evaluation* (TS-79-1). U.S. Office of Personnel Management, Personnel Research and Development Center, (NTIS No. PB 80 102650).

Pearlman K. (1982). The Bayesian approach to validity generalization: A systematic examination of the robustness of procedures and conclusions. *Dissertation Abstracts International*, *42*, 49609A.

Pearlman K, Schmidt FL, Hunter JE. (1980). Validity generalization results for tests used to predict training success and job proficiency in clerical occupations. *Journal of Applied Psychology*, *65*, 373–406.

Porter WR, Levine EL, Flory A III. (1976). *Training and experience evaluation*. Tempe, AZ: Personnel Services Organization.

*Primoff ES. (1958). *Report on validation of an examination for electrical repairer, McClellan Field California*. Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center.

Primoff ES. (1975). *How to prepare and conduct job element examinations* (TS-75-1). Washington, DC: U.S. Government Printing Office.

Reilly RR, Chao GT. (1982). Validity and fairness of some alternative employee selection procedures. PERSONNEL PSYCHOLOGY, *35*, 1–62.

Sage DR, Cole J, Johnson B. (undated) *Manual for the development and administration of education and experience rating guides*. Examination Development and Analysis Section, Division of Personnel, Office of Administration, State of Missouri.

Schmidt FL, Caplan JR, Bemis SE, Decuir R, Dunn L, Antone L. (1979). *The behavioral consistency method of unassembled examining* (TM-79-21). Washington, DC: U.S. Office of Personnel Management, Personnel Research and Development Center, (NTIS No. PB 268785/AS)

Schmidt FL, Gast-Rosenberg I, Hunter JE. (1980). Validity generalization: Results for computer programmers. *Journal of Applied Psychology, 65*, 643–661.

Schmidt FL, Hunter JE, Outerbridge AN. (1986). The impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432–439.

Schmidt FL, Hunter JE, Outerbridge A, Trattner MH. (1986). The economic impact of job selection method on size, productivity, and payroll costs of the federal work force: An empirical demonstration. PERSONNEL PSYCHOLOGY, *39*, 1–30.

Schmidt FL, Hunter JE, Pearlman K, Hirsh HR. (1985). Questions and answers about validity generalization and meta-analysis. PERSONNEL PSYCHOLOGY, *38*, 697–798.

Schmidt FL, Hunter JE, Pearlman K, Shane GS. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. PERSONNEL PSYCHOLOGY, *32*, 257–281.

Schmidt FL, Hunter JE, Urry VW. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61*, 476–485.

State of Connecticut, Department of Administrative Service, Personnel Division, Psychometric Unit. (1978). *Development and administration of experience/ training evaluations (factor method)*. Hartford, CT: Author.

*Taylor N, Zrout T. (1974). *Validation of job element examination for building service worker*. Urbana, IL: Testing Research Program, University Civil Service System of Illinois, University of Illinois.

Tukey JW. (1960). A survey of sampling from contaminated distributions. In Olkin I, Ghurye SG, Hoeffding W, Madow WD, Mann HB (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

*U.S. Civil Service Commission. (1958). *Historical development of a method of quality rating of training and experience*. Washington, DC: Bureau of Programs and Standards, Standards Division, Test Development Occupational Research Section.

*Zrout T. (1973). *Validation of the job element examinations for hospital service worker and building service worker*. Urbana, IL: Testing Research Program. University Civil Service System of Illinois, University of Illinois.

Appendix A

Listing of Information on Validity Studies Included in the Analyses

Author	Job title	DOT code	*n*	*r*	Criterion[a]	Adjustments[b]
	. . . Behavioral consistency . . .					
Acuff, R., 1965	Air craft electrical worker	825.381-010	22	.783	SR	CPC
Acuff, R., 1965	Air craft electrician	825.381-010	24	.137	SR	CPC
Haynes, E., undated	Shipfitter	806.381-046	21	.661	SR	U
Haynes, E., undated	Shipfitter helper	806.687-050	10	.616	SR	U
Hough, L. et al., 1980	Attorney	110.107-010	220	.22	SR	N
Hough, L. et al., 1980	Economist	050.067-010	23	.20	SR	N
Hough, L. et al., 1980	Consumer protection specialist	096.121-014	32	.41	SR	N
Hough, L. et al., 1980	Accountant	160.167-010	10	.13	SR	N
Hough, L. et al., 1980	Manager	186.167-026	20	-.03	SR	N
Hough, L. et al., 1984	Librarian	100.127-014	298	.22	SR	N
Hough, L. et al., 1984	Supervisory librarian	100.117-010	77	.26	SR	N
Hough, L. et al., 1984	Other professional	not available	165	.19	SR	N
Johnson, J.C. et al., 1980	Sr. eligibility counselor	195.267-010	104	.260	SR	N
MacLane, C.N. , 1986	Claims representative	169.277-014	101	.21	SR	N
Primoff, E.S., 1958	Electrical repairer	825.281-010	22	.590	SR	C
	. . . Illinois job element . . .					
Benz, M.P., 1974a	Admissions & records officer I	205.367-010	53	.304	SR	C
Benz, M.P., 1974b	Kitchen laborer	318.687-010	175	.098	SR	C
Halinski, L., 1974	Nursing assistant	355.674-014	72	.306	SR	C
London, M., 1975	Animal caretaker	410.674-014	49	.003	SR	CG
London, M., 1975	Building service worker	382.664-010	278	.181	SR	CG
London, M., 1976a	Clerk II	219.362-010	112	.143	SR	CGM
London, M., 1976a	Clerk-stenographer II	202.362-101	266	.115	SR	CGM
London, M., 1976a	Library clerk II	249.367-046	308	.063	SR	CGM
London, M., 1976a	Typing clerk II	219.362-010	182	.075	SR	CGM
London, M., 1976a	Typing clerk III	219.362-010	249	.134	SR	CGM
London, M., 1976b	Cashier II	211.462-010	405	.133	SR	CGM
London, M., 1976b	Secretary (stenographic)	202.362-014	309	.092	SR	CGM
London, M., 1976b	Secretary (transcribing)	201.362-030	243	.141	SR	CGM
Marusarz, T., 1974	Janitor	382.664-010	372	-.014	SR	C
Taylor, N. et al., 1974	Building service worker	382.664-010	111	.069	SR	C
Zrout, T., 1973	Hospital service worker	355.674-010	24	.090	SR	C
	. . . Point . . .					
Bean, K.L., 1958	Graduate nurse	075.364-010	46	-.370	SR	N
Benz, M.P., 1974c	Health service nurse	075.124-014	59	.077	SR	C
Delaney, E.C., 1954	School teacher	092.227-014	93	.158	SR	N
Ebright, E., 1959	Engineering draftsman	005.281-010	18	.137	SR	N
Farrell, B.M., 1979	Accounting officer senior	160.167-010	60	-.010	SR	N
Groll, M.F., 1975	Custodian worker	382.664-010	19	.47	SR	N
Groll, M.F., 1975	Custodian worker	382.664-010	17	.10	SR	N
Groll, M.F., 1975	Food service worker	313.361-014	12	.34	SR	N
Johnson, J.C. et al., 1980	Sr. eligibility counselor	195.267-010	104	-.010	SR	N
London, M., 1975	Library technical assistant I	100.367-010	190	.040	SR	CG
London, M., 1975	Medical technologist	078.361-014	120	-.032	SR	CG

Appendix A (continued)

Author	Job title	DOT code	*n*	*r*	Criterion[a]	Adjustments[b]
Lyons, T.J., 1984[c]	Clerk-typist, stenographer	201.362-010	179	.020	SR	N
Lyons, T.J., 1985[d]	Accountant	160.167-010	34	-.061	SR	N
Lyons, T.J., 1985	Accountant	160.167-010	73	-.106	SR	N
Lyons, T.J., 1985	Auditor	160.162-014	212	-.053	SR	N
Lyons, T.J., 1985	Auditor	160.162-014	31	-.086	SR	N
McKinney, T.S. et al., 1976	Engineering technician I	638.261-010	48	.032	PR	U
McKinney, T.S. et al., 1976	Engineering technician I	638.261-010	18	.010	SR	N,2
McKinney, T.S. et al., 1976	Groundskeeper	406.687-010	12	-.261	SR	N
McKinney, T.S. et al., 1976	Groundskeeper	406.687-010	27	-.126	PR	U,2
McKinney, T.S. et al., 1976	Semi-skilled worker	955.687-010	13	.465	SR	N
McKinney,T.S. et al., 1976	Semi-skilled worker	955.687-010	41	.472	PR	U,2
McKinney, T.S. et al., 1976	Street maintenance worker I	955.687-018	33	-.021	SR	N
McKinney, T.S. et al., 1976	Street maintenance worker I	955.687-018	48	.111	PR	U,2
Molyneaux, J.W., 1953	Adjucator GS-ll	119.167-010	122	.000	SR	N
Molyneaux, J.W., 1953	Adjucator GS-12	119.167-010	58	.350	SR	N
Molyneaux, J.W., 1953	Adjucator GS-9	119.167-010	256	-.150	SR	N
Molyneaux, J.W., 1953	Athletic director GS-7	153.224-010	28	.200	SR	N
Molyneaux, J.W., 1953	Athletic technician GS-5	153.224-010	24	.090	SR	N
Molyneaux, J.W., 1953	Athletic technician GS-6	153.224-010	39	.020	SR	N
Molyneaux, J.W., 1953	Correctional therapist GS-5	045.107-010	94	.110	SR	N
Molyneaux, J.W., 1953	Correctional therapist GS-7	045.107-010	34	.180	SR	N
Molyneaux, J.W., 1953	Correctional therapist GS-9	045.107-010	30	.300	SR	N
Molyneaux, J.W., 1953	Dental technician GS-6	712.381-018	51	.450	SR	N
Molyneaux, J.W., 1953	Dental technician GS-7	712.381-018	22	.390	SR	N
Molyneaux, J.W., 1953	Dietician GS-5	077.127-014	96	.280	SR	N
Molyneaux, J.W., 1953	Dietician GS-7	077.127-014	83	.140	SR	N
Molyneaux, J.W., 1953	Dietician GS-9	077.127-014	20	-.130	SR	N
Molyneaux, J.W., 1953	Loan guarantee off. GS-11	186.267-018	49	.100	SR	N
Molyneaux, J.W., 1953	Loan guarantee off. GS-12	186.267-018	26	.380	SR	N
Molyneaux, J.W., 1953	Loan guarantee off. GS-7	186.267-018	64	.090	SR	N
Molyneaux, J.W., 1953	Loan guarantee off. GS-9	186.267-018	50	.230	SR	N
Molyneaux, J.W., 1953	Manual arts therapist GS-6	076.124-010	36	-.030	SR	N
Molyneaux, J.W., 1953	Manual arts therapist GS-8	076.124-010	38	-.550	SR	N
Molyneaux, J.W., 1953	Medical technician GS-3	079.367-018	27	.340	SR	N
Molyneaux, J.W., 1953	Medical technician GS-4	079.367-018	48	.310	SR	N
Molyneaux, J.W., 1953	Medical technician GS-5	079.367-018	145	.130	SR	N
Molyneaux, J.W., 1953	Medical technician GS-6	079.367-018	136	.070	SR	N
Molyneaux, J.W., 1953	Medical technician GS-7	079.367-018	24	.420	SR	N

Appendix A (continued)

Author	Job title	DOT code	*n*	*r*	Criterion[a]	Adjustments[b]
Molyneaux, J.W., 1953	Occ. therapist aide GS-3	355.377-010	20	.000	SR	N
Molyneaux, J.W., 1953	Occupational therapist GS-5	076.121-010	98	.250	SR	N
Molyneaux, J.W., 1953	Occupational therapist GS-7	076.121-010	41	-.150	SR	N
Molyneaux, J.W., 1953	Pharmacist GS-5	074.161-010	36	.360	SR	N
Molyneaux, J.W., 1953	Pharmacist GS-7	074.161-010	63	.480	SR	N
Molyneaux, J.W., 1953	Pharmacist GS-9	074.161-010	26	.290	SR	N
Molyneaux, J.W., 1953	Physical therapist GS-5	076.121-014	83	.120	SR	N
Molyneaux, J.W., 1953	Physical therapist GS-7	076.121-014	64	.200	SR	N
Molyneaux, J.W., 1953	Physical therapist GS-9	076.121-014	20	.230	SR	N
Molyneaux, J.W., 1953	Recreational director GS-6	195.227-014	41	-.200	SR	N
Molyneaux, J.W., 1953	Recreational director GS-7	195.227-014	37	-.390	SR	N
Molyneaux, J.W., 1953	Recreational tech. GS-5	341.367-010	24	.270	SR	N
Molyneaux, J.W., 1953	Recreational tech. GS-6	341.367-010	73	.250	SR	N
Molyneaux, J.W., 1953	Registration Officer GS-10	205.367-042	59	-.450	SR	N
Molyneaux, J.W., 1953	Registration Officer GS-11	205.367-042	20	-.210	SR	N
Molyneaux, J.W., 1953	Registration Officer GS-7	205.367-042	284	.030	SR	N
Molyneaux, J.W., 1953	Registration Officer GS-9	205.367-042	434	-.060	SR	N
Molyneaux, J.W., 1953	Social Worker GS-11	195.107-034	24	.120	SR	N
Molyneaux, J.W., 1953	Social Worker GS-7	195.107-034	227	-.020	SR	N
Molyneaux, J.W., 1953	Social Worker GS-9A	195.107-034	74	.090	SR	N
Molyneaux, J.W., 1953	Social Worker GS-9B	195.107-034	31	.410	SR	N
Molyneaux, J.W., 1953	Social Worker GS-9C	195.107-034	28	.030	SR	N
Molyneaux, J.W., 1953	Training Officer GS-7	166.227-010	41	.110	SR	N
Molyneaux, J.W., 1953	Training Officer GS-9	166.227-010	114	-.160	SR	N
Molyneaux, J.W., 1953	X-Ray Technician GS-5	199.361-010	99	.150	SR	N
Molyneaux, J.W., 1953	X-Ray Technician GS-6	199.361-010	33	-.150	SR	N
Mosel, J.M., 1952	Auto mechanic	620.261-010	99	.270	SR	N
Mosel, J.M., 1952	Aviation metalsmith	806.682-010	98	.120	SR	N
Mosel, J.M., 1952	Carpenter	860.664-010	51	.190	SR	N
Mosel, J.M., 1952	Equipment repairman	806.684-118	40	.160	SR	N
Mosel, J.M., 1952	Hand compositor	973.381-010	116	.160	SR	N
Mosel, J.M., 1952	Highlift fork operator	921.683-050	116	-.040	SR	N
Mosel, J.M., 1952	Machine operator	619.360-018	108	.140	SR	N
Mosel, J.M., 1952	Machinist (I)	600.280-022	76	.000	SR	N
Mosel, J.M., 1952	Machinist (II)	619.360-018	100	.040	SR	N
Mosel, J.M., 1952	Ordnanceman	632.261-010	100	-.090	SR	N
Mosel, J.M., 1952	Ordnancemen-torpedo	632.261-018	125	.090	SR	N
Mosel, J.M., 1952	Painter	840.381-010	75	.260	SR	N
Mosel, J.M., 1952	Radio mechanic	833.261-018	116	-.170	SR	N
Ocasio, B.P., 1983	Custodial worker	382.664-010	67	.240	SR	N
U.S. Civil Service, 1958	Auto mechanic	620.261-010	88	.250	SR	N
U.S. Civil Service, 1958	Metalsmith	619.360-014	115	.120	SR	N

Appendix A (continued)

Author	Job title	DOT code	*n*	*r*	Criterion[a]	Adjustments[b]
	. . . *Task* . . .					
Anderson, C.D. et al., 1984	Clerical jobs	203.362-010	66	.270	TY	N
Farrell, B.M., 1979	Accounting officer intermediate	160.167-010	50	.640	SR	N
Farrell, B.M., 1979	Accounting officer senior	160.167-010	60	.340	SR	N
Johnson, J.C. et al., 1980	Sr. eligibility counselor	195.267-010	104	.100	SR	N
Lyons, T.J., 1984[c]	Clerk-typist, stenographer	201.362-010	179	.029	SR	N
Lyons, T.J., 1985[d]	Accountant	160.167-010	34	-.012	SR	N
Lyons, T.J., 1985	Accountant	160.167-010	73	-.013	SR	N
Lyons, T.J., 1985	Auditor	160.162-014	212	-.100	SR	N
Lyons, T.J., 1985	Auditor	160.162-014	31	-.056	SR	N
Ocasio, B.P. et al., 1980	Maintenance clerk	239.367-014	182	.090	SR	N

[a]Criterion codes: SR = Supervisory Rating; PR = Promotion; TY = Typing (Job Sample Test)

[b]Adjustment codes: C = Composite—one predictor with composite criteria; CG = Composite correlation coefficient based on mean criterion data (see text page 292 for explanation); CGM = Mean of two or more composite correlations based on mean criterion data (see text page 292 for explanation); N = No adjustments; U = Adjustment for unequal sample sizes in dichotomous variable; 2 = 2nd coefficient for same job; CPC = Correlation of composite predictor with composite criterion.

[c]Two coefficients were reported for this study. The self-report measure contained multiple item types grouped into subsections. On the basis of content, we classified one subsection as a point method measure and another as a task method measure. Data from the other subsections were not used because the item types did not fit the T&E categories covered in this research. The validity of the total self-report measure, including all subsections, was .21.

[d]For each sample, two coefficients were reported. This self-report measure contained multiple item types grouped into subsections. On the basis of content, we classified four subsections as point method measures and one section as a task method measure. The task method coefficient was assigned to the task category. Data from other subsections were not used because the item types did not fit the T&E categories covered in this research.

Appendix B

Differences Between Computer Programs

The analyses presented in this paper were performed using a computer program that operationalized the structural equations for the interactive meta-analytic procedures (Schmidt, Gast-Rosenberg et al., 1980). This program uses a slightly different computational method from that used in previous research by Schmidt, Hunter, and associates. The original program uses the statistic U to record range restriction information. U was defined as the ratio of restricted to unrestricted standard deviations. The quantity $\bar{r}$ is the mean observed coefficient. The present program uses c, which is a function of U and the size of the true r. It is defined by Hunter et al. (1982, p. 86) as

$$c = \sqrt{U^2 + (1 - U^2)\bar{r}^2}.$$

This computational difference results in small differences in the true mean correlation and true standard deviation estimates. Six analyses were run using both programs. The results for the two programs are identical for all practical purposes. The average true mean difference was .00016, and the true standard deviation difference was .00226.