

JOB APPLICANT TRAINING AND WORK EXPERIENCE EVALUATION IN PERSONNEL SELECTION

Ronald A. Ash, James C. Johnson, Edward L. Levine,
and Michael A. McDaniel

INTRODUCTION

Biographical data are pervasive in their effect upon human lives. Vagaries of birth in many societies may still dictate whether a person is designated an untouchable, a noble, or even a king. People select their mates on such grounds as race or religion, their friends on the basis of similarities on such factors as age, and their leaders on the basis, often, of their having attended the right schools, their holding memberships in the right clubs and perhaps even their very height.

Research in Personnel and Human Resources Management, Volume 7, pages 183-226.
Copyright © 1989 by JAI Press Inc.
All rights of reproduction in any form reserved.
ISBN: 1-55938-005-5

Tolstoy's famous novel *Anna Karenina* is one of many literary sources that depicts how the proper connections may make or break an individual's career. Though the characters are fictitious, it appears that the selection processes employed in Russian society at the time the novel unfolds were veridically depicted.

Another well known illustration of this type of effect is the continuing controversy over the selection of the "true" author of Shakespeare's plays and sonnets. Marchette Chute (1949), author of the well-known biography *Shakespeare of London*, described the controversy, when he wrote (p. 350, Dutton edition), "Many people felt almost relieved when Delia Bacon took the final logical step in 1857 and suggested that the plays of Shakespeare were written by someone else entirely. Delia chose Sir Francis Bacon for the honor. Later candidates to be suggested have been the Earl of Oxford, Sir Edward Dyer, the fifth Earl of Rutland, the sixth Earl of Derby, and even the Countess of Pembroke. . . ." Chute ascribed the controversy to the "respect for the literary value of noble birth. . . ."

Lest an assumption be made too quickly that there is an easy acceptance of the legitimacy and validity of such usage of biographical data, Chute illustrated the case for the doubters, noting (p. 360), "The respect for the literacy value of noble birth is . . . a little hard to explain logically, since the most learned of Elizabethan dramatists was a bricklayer, and the most poetic, next to Shakespeare, was the son of a cobbler."

This is the context that must be conjured with when the modern day use of biographical data is explored for making critical selection decisions. Biographical data used for such decision making nowadays may include, but is not limited to previous jobs held, past education and training, membership in professional, ethnic, religious, and political groups, demographic characteristics, such as height and weight, credit standing, marital status, and place of residence. For purposes of this paper, primary attention is focused upon work experience, work-related achievements, and educational history. The use of such information in personnel selection is discussed, including underlying theory, practices and procedures, and the results of evaluative research. The discussion is believed to have implications for personnel practitioners and researchers alike.

One perspective on the use of biographical data for selecting employees was provided by Levine and Flory (1975). They proffered a conceptual framework to categorize the types of biographical data that might be used, the manner of interpretation of the data, and the method of evaluating the data. Data may be: (1) job related (e.g., previous work experience) or not (e.g., marital status); (2) interpreted as evidence of straightforward personal characteristics (e.g., closeness of match between past jobs and the one to be filled) or as indicators of inferred traits like leadership ability; and (3) evaluated judgmentally or statistically.

Combining these three dichotomous variables in all possible ways yielded eight categories, which Levine and Flory (1975) viewed as a complete rendering of all approaches to biographical data usage in selection. The categories of interest in this paper, which capture the training and experience (T&E) evaluation approaches covered here include all four of those designated by high job relatedness. However, some methods rely on surface characteristics, others upon inferred traits. Most methods are judgmental in nature, but a few rely on statistical scoring methods.

Porter, Levine, and Flory (1976) have defined and described T&E evaluation processes on a number of facets. To summarize and update their description in accordance with Levine and Flory's (1975) framework, it should be noted first that T&E evaluations rely on written information provided by applicants on application blanks, questionnaires, or computer interpretable input modes. The types of *job-related* information employed include performance on tasks in previous jobs, brief narrative descriptions of past jobs, past accomplishments on job-related dimensions, such as leadership ability, summaries of educational history, and reported in-service training. Self-ratings of levels of knowledge, skills, and abilities (KSAs) that applicants possess, which are corroborated with previous, relevant work or educational experiences to pinpoint where these KSAs were acquired, might also be encountered. Thus, information may be evaluated on its face or used to infer the degree to which an applicant possesses important KSAs.

The information may be evaluated judgmentally by a rater who follows a scoring plan that is based on assumptions about how previous work experience and education are related to future job success. Or, the scoring of the information may be done by a fixed formula. An example of the latter may be seen in the following: An applicant is asked to review a list of tasks relevant to a job in question, and to check those which he or she has performed in the past. The number of checks is merely counted. More complicated versions of this approach are covered later in this paper.

T&E evaluation is an examination in the sense that, by means of the various rating or scoring approaches, applicants are formally assigned a rating or score that ranks them according to their prospects for success on the job. Score designations may be as crude as "qualified" versus "not qualified" or as refined as scores carried to hundredths of a point on a 0-100 scale.

As pointed out by Porter et al. (1976), the T&E is, legally speaking, a test and is subject to the *Uniform Guidelines on Employee Selection Procedures* (1978). This is the case whether the T&E is referred to as an E&E (education and experience evaluation), TRAEX (training and experience evaluation), or simply as a scored application blank.

There are other modern approaches to the use of biographical data that must be distinguished from T&E evaluation. One is the weighted application form or the closely related biographical information blank (Asher, 1972; Cascio,

1987; England, 1971). Both of these devices are questionnaires that contain job-related and non-job-related items. By *empirically* keying individual item responses to criteria of job success or longevity, scoring keys are established. These keys are then applied to the forms completed by applicants.

Two primary distinctions between the weighted biodata approach and T&E evaluation are: first, that weighted biodata often include non-job-related items, and second, that validity studies are done to link item responses and criteria empirically prior to establishing scoring procedures. To illustrate the former point, some of the most valid items in biographical information blanks are ones that bear no obvious relationship to the job. Levine and Flory (1975) cited two examples of valid items from a study completed by Scollay (1956): "If subject's father is dead, subject's age at the time," and "Mother's occupation." Asher (1972) cited one item that was reported to be almost as good a predictor of success in flight training as the entire Air Force Test Battery: "Did you ever build a model airplane that flew?" Such questions as these are *not* employed in T&E evaluation.

As to the second point, T&E evaluations rely on rationally-based rather than empirically-based linkages between item responses and prospects of future job success. In the example of the task checklist cited above, the assumption behind the scoring scheme is that the more job-relevant the tasks a worker performed in the past, the better qualified he or she is for the job in question. Of course, the empirical underpinnings of weighted biodata virtually guarantee a highly valid selection process, whereas validity studies on the T&E evaluation approach are relatively few in number and generally suggest that T&E scores are only moderately valid at best (Ash & Levine, 1981; Asher, 1972; Hunter & Hunter, 1984). A good deal more will be reported on this and related topics in later sections of this paper.

Another approach to be contrasted with formal T&E evaluation is the cursory review of the resume or application form where, unlike T&E evaluation, no formal scoring scheme is employed and no rating or score is formally assigned. An illustration of this approach is scanning a resume to determine whether or not an applicant might be a good bet to invite for an interview. This cursory approach is discussed later under the subheading *holistic judgment*.

Given its roots in the public sector and its reliance on formal scores, T&E evaluation is predominantly associated with and widely used in governmental agencies, particularly where a formal merit system is established. Saso and Tanis (1974) reported survey results showing that T&E's were widely used in the public sector. Nor is there any evidence to dispute their findings in terms of more recent patterns of usage (cf. Ash & Levine, 1981). This is why T&E evaluation is an important area of study despite the apparently limited validity of many approaches to such evaluation.

The appeal of T&E evaluation in the public sector rests primarily on grounds other than validity. Porter et al. (1976) suggested that T&E evaluations yield scores needed for public merit system selection with less developmental cost and effort than is the case with other methods, such as written tests. A long tradition of use also accounts for the popularity of T&E evaluation in the public sector. T&E evaluation is sometimes preferable to written tests because it does not require that candidates appear at a certain time and place to be examined or tested. Hence, T&E evaluation processes are sometimes referred to as unassembled examinations.

In terms of the role of T&E evaluation in public sector selection, T&E evaluations may be employed as the sole rating or ranking device. Alternatively, T&E evaluations may be employed as a preliminary screening device to reduce the applicant pool prior to the administration of a more expensive, more time consuming procedure, such as an assessment center, an interview, or a written test. On occasion, the T&E score may be incorporated into a composite that includes scores on other devices.

Because of its relative efficiency and its widespread use, T&E evaluation demands research attention. Research could unearth the more valid methods; ones that might find use even in the private sector, and discern the proper role of T&E evaluation in the selection process. Moreover, such research might overcome the problem that T&E evaluation, despite its widespread use, is all but ignored in otherwise excellent treatments of personnel selection (Cascio, 1987; Dreher & Sackett, 1983; Schneider & Schmitt, 1986).

This paper has several purposes, then. T&E evaluation has thus far been described, defined, and differentiated from the empirically weighted biographical data approach to personnel selection. Next, theories and assumptions underlying the scoring of job applicant training and work history for use in personnel selection are discussed. This is followed by a discussion of the use of training and experience as minimum qualifications in employment screening. Then come descriptions of various T&E evaluation methods used for the purpose of ranking job applicants. Subsequently, research on the reliability and validity of various T&E evaluation methods is summarized. Finally, recommendations for practice and future research on T&E evaluation are offered.

THEORIES AND ASSUMPTIONS UNDERLYING T&E EVALUATION

One major assumption underlying T&E evaluation is that the written exchange of biographical information, where the source is the applicant, is likely to be more accurate than is similar information provided by other vehicles of information exchange. Studies have suggested that the biodata supplied on

the application blank is more accurate than that supplied in interviews (Cascio, 1987; Levine & Flory, 1975). Moreover, the applicant obviously has an easier and more direct access to such information as compared to a background investigator.

Another major assumption underlying T&E evaluation is the necessity of job analysis. Since the hallmark of T&E evaluation is matching backgrounds of applicants to job requirements, job analysis must play an indispensable role. When raters have little job knowledge to inform their evaluations, the likely outcome is unreliable ratings or reliance on irrelevant information (Dipboye, Fromkin, & Wiback, 1975; Langdale & Weitz, 1973; Wiener & Schneiderman, 1974). Levine (1983) provided a method of job analysis called B-JAM (Brief Job Analysis Method), which includes a segment devoted to gathering information that can be used to devise a rating plan specifically for T&E evaluation. However, any of the well known job analysis approaches reviewed by Levine can provide a useful foundation for T&E evaluation.

A third important assumption of current T&E evaluation methods is that standardization be incorporated to the fullest extent possible (Levine & Flory, 1975). The standardization issue is attacked from three perspectives: (1) standardization of the forms and questionnaires completed by applicants, (2) standardization of the rating forms, (3) standardization of rating processes through reliance on training programs for raters and/or documentation of rating/scoring procedures in protocols which serve as job aids.

Arvey, McGowen, and Horgan (1981), Ash and Levine (1981), and Porter et al. (1976) all have offered overlapping discussions about the assumptions underlying a belief in the presumptive predictive power of T&E evaluation. These assumptions suggest that T&E measures may work because they measure the degree to which a complex of job-relevant abilities, skills, and motivational patterns are possessed by applicants. The notion that past behavior is a powerful predictor of future behavior reflects this thinking. If an applicant has performed tasks relevant to a job in the past, and further performed them well, he or she should be able and willing to do so again. There are many studies in various behavioral domains supporting this notion. For example, Hunter and Hunter (1984) found past work experience to be predictive, though modestly so, of future job performance.

Wernimont and Campbell (1968), among others, have argued that the predictive power of past performance may be greatest when previous accomplishments match job demands for accomplishment in all their complexity. Therefore, directly relevant job experience that matches closely the work required in the job for which screening is being done should be credited more highly than less closely matched experience.

The motivational side of the picture has to do with the issue of commitment to an occupation. As experience and education directed toward a particular

occupation grow, commitment to remain in the occupation, and to seek success within it should likewise grow.

A more tenuous assumption is that T&E evaluations provide measures of specific knowledge, skills, and abilities (KSAs) needed for job performance. This is most explicit, of course, in T&E evaluation methods that ask for information directly related to KSAs. The reason that this assumption is more tenuous than the others is the obvious lack of precision in measuring KSAs, which the reliance on global reports of previous work experience and accomplishments provides. T&E evaluation is not nearly as well suited to measuring cognitive abilities, for example, as are written tests. In addition, the manner in which KSAs combine to produce skilled performance at various points in someone's career is not easily specified (Adams, 1987).

There are other, more specific assumptions about T&E evaluation that are applicable to particular T&E evaluation methods. Some are merely different iterations of the items already discussed. These assumptions include the following:

1. Training and experience that shows a pattern of progressive increase in complexity and responsibility makes an applicant better qualified than one who does not exhibit such a pattern. This might be looked at in the framework of learning, where it is not clear that someone who worked as a supervisor in an occupation for a time and then became a worker is less qualified than a candidate who did the reverse. In his review of motor skills learning, Adams (1987) considered an analogous notion—adaptive training where difficulty of tasks to be performed by the learner are adapted to each learner's level of mastery. Unfortunately, the approach did not produce better trained people.
2. Training and experience obtained more recently is likely to make an applicant better qualified than one whose experience is stale. For those occupations that have undergone massive changes, it is difficult to argue with this assumption. For example, a personnel testing technician with experience in the post EEO Act era is likely to be more qualified to do testing than is an applicant whose testing experience was obtained prior to 1965. An issue that must be considered in this context is how long and difficult a retraining period would need to be to overcome the projected disparity in qualifications.
3. More complex or responsible training and experience is more valuable than less complex or responsible training and experience. This assumption seems supportable on the basis of skill acquisition (Adams, 1987), so long as the complexity matches that required by the job being filled.
4. Greater length of training and experience is more valuable than lesser length of training and experience. On rational grounds, this assumption

is defensible only to the extent that greater length continues to provide opportunities for increased levels of competence in a given occupation. There will often be a point at which new learning reaches its limits beyond which additional training and experience is not instrumental to successful performance. This latter notion undergirds those T&E scoring systems that credit, for example, experience of up to three or five years, but do not additionally credit experience beyond this level.

5. Length and complexity of experience or education beyond a certain optimum amount should result in less qualified applicants. Such an assumption underlies the thinking that a veteran performer is likely to be overqualified for an entry-level position, and so should receive a lower score on a T&E evaluation than an applicant with less experience. There is some anecdotal evidence bearing on this point, both positive and negative, but little in the way of strong research. People have heard of the accomplished worker who quits an entry-level job due to boredom, but they have also heard of the accomplished veteran who, desirous of a less stressful job, performs admirably in a non-supervisory position. In an era of concern about age discrimination, this assumption is very tenuous, both because of lack of research and legal vulnerability.

What would a complete theory of T&E evaluation consist of? It would contain a set of carefully researched assumptions, linkages between the various T&E evaluation procedures and the constructs they measure, and linkages between these constructs and job success in a variety of occupations and settings. At present, despite the fairly lengthy history of the use of biographical data in selection, the field is not yet close to such a theory. There is important work underway in the area of biodata which may ultimately yield a theory applicable to T&E evaluation and selection (cf. Owens & Schoenfeldt, 1979), but such a theory does not appear to be on the immediate horizon.

TRAINING AND EXPERIENCE AS THE MINIMUM QUALIFICATIONS FOR EMPLOYMENT

The Use of Minimum Qualifications in the Personnel Selection Process

A *minimum* amount or kind of prior training, education, or experience may well be the most common employment requirement in the United States. Newspaper employment ads typically include such statements by employers as "College degree a must!" and "Five years experience in restaurant management preferred." Certainly the belief that education is strongly associated with future employment prospects seems almost universal in most segments of our society, a reason why many parents seek "the best" schools

for their children and insist on a college education for them. But there are competing views. The assumption that education is associated with job performance has often been challenged as unfairly discriminatory, whether intentional or not, serving to screen out of the employment process minority group members and others who have tended in the past to end their educations before obtaining high school diplomas or entering and completing college degree programs.

Organizations differ widely in their use of education and experience requirements. Some private and most public employers follow formal, systematic procedures to recruit and ultimately select employees for vacant positions, while others follow procedures which are far less formal, consistent, and systematic.

In practice, use of a minimum qualification (MQ) is part of an employment *system* composed of interacting components. *Recruitment* is the initial step through which prospective candidates are induced to enter the employment process. Through newspaper advertisements and notices in other publications, posted announcements, recruiters and external employment agencies, or other means, a pool of candidates is sought from which one or more persons can be selected for employment. Salary and other benefits are often highlighted to maximize interest among potential candidates, while “qualifications” are expressed to limit the pool to those most likely to possess the attributes sought by the organization. The “qualifications” sought by employers are typically expressed in one of three ways: (1) as a minimum amount and kind of education and experience; (2) as a preferred amount and kind of education and experience; or (3) as a statement of the job competencies the organization seeks, regardless of how these might have been acquired. As discussed in greater detail shortly, these alternative means of asking prospective candidates to “pre-select” themselves may have paradoxical effects on the efficiency and effectiveness of the employment process.

Preliminary screening procedures also vary. If a minimum qualification of education and experience is used, candidates who do not meet the requirement are eliminated from further consideration—a *screen out* process. Another practice is the review of application materials for the purpose of identifying a few candidates to be further assessed—a *screen in* process.

Ranking candidates to reach a final hiring decision is usually a more extensive process, particularly in public organizations. Tests, assessment centers, performance tests, employment interviews, and other methods may be used in varying combinations to assess candidates more thoroughly before reaching final hiring decisions.

Use of MQ requirements as a preliminary screen—in effect, a pass-fail test which is the first hurdle in a multiple-hurdles selection strategy—may heavily impact the utility of the rank-ordering process. Though obvious, this fact is often unrecognized by many employers, with the result that personnel selection

problems are unexpectedly created, or are incorrectly attributed to other components of the employment process.

The nature and stringency of education and experience requirements affect both the number of candidates likely to be recruited, and the number available for more comprehensive assessment. Employers adopting minimal preliminary requirements (or none at all) rely on subsequent phases of the assessment process to identify candidates who will be employed; adverse impact and the need for validity thus rest primarily on those methods of assessment, and a relatively large number of candidates must be assessed. With stringent education and experience requirements, however, fewer candidates need to be assessed in the later phases of the selection process, but the requirement is more likely to adversely impact the employment prospects of members of legally protected groups. The importance of insuring the validity of the education and experience requirement is thus greater. Stringent requirements are also potentially more costly, not only because of the possibility of unfair discrimination, but because of increased need for recruitment to insure adequate numbers of candidates, and because higher salaries are often required to attract competent candidates who possess the credentials sought.

In practice, it appears that most employers determine education and experience requirements on the basis of a rational judgment of job requirements, comparison with requirements for other jobs in the organization and the labor market generally, and compensation policy. Education and experience have "market value" and consequently, persons with more education and experience seek and expect greater compensation than persons with less "value." These linkages between education and experience requirements for employment, compensation policy, and conditions in the labor market affecting recruitment impact both the determination and "validation" of MQs as employment requirements in complex ways (Johnson & Waldron, in preparation). For example, stringent credential requirements associated with relatively low pay may result paradoxically in employment of persons with *lower* ability than would be the case if the credential requirements were *less* stringent. Given stringent credential requirements and *low* pay, the employer may attract applicants primarily from the lower portion of the ability distribution among those who meet the stringent requirements. With *less* stringent requirements it may be possible to attract a higher proportion of applicants from the higher portion of the ability distribution among those who meet the less stringent requirements.

Attributes Assessed

The importance of determining the validity of paper and pencil tests and similar quantitative selection techniques is widely accepted, as generally are the methods for doing so. Establishing the validity of MQs, however, has posed

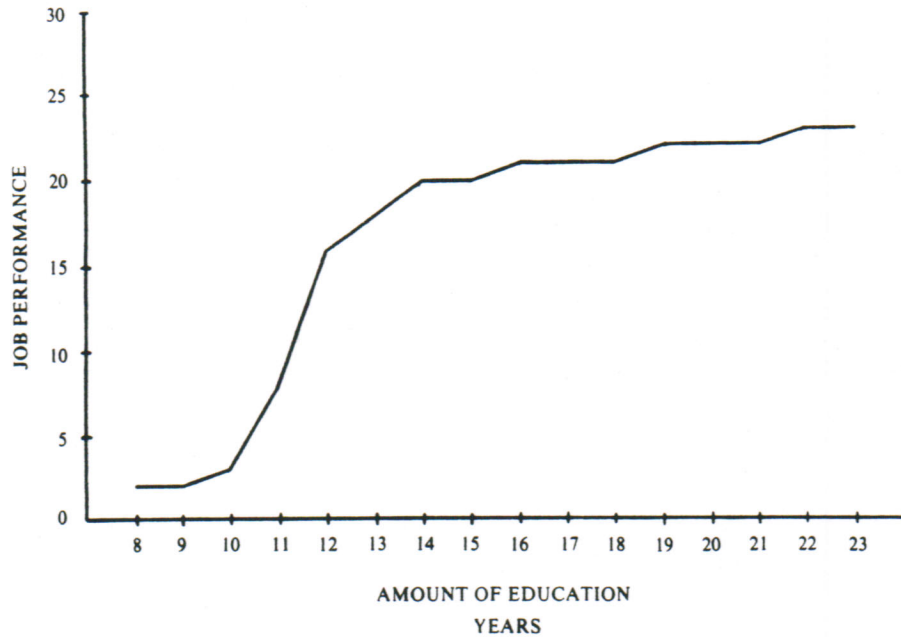


Figure 1a. Ideal relationship between amount of education and job performance.

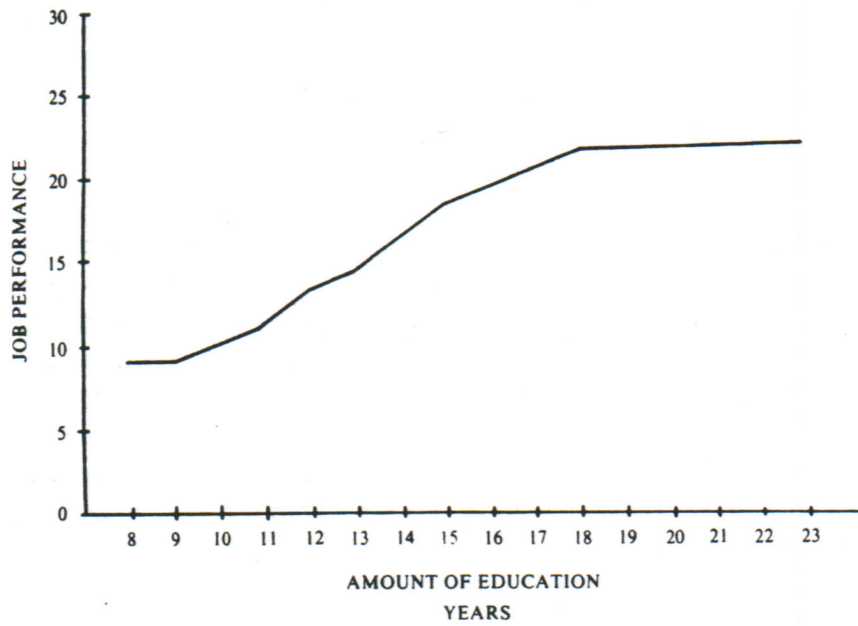


Figure 1b. Realistic hypothetical relationship between education and job performance.

a problem: virtually no discussion of research findings or appropriate methodologies exists in either personnel selection texts or in professional journals. Effective use of a minimum education and experience requirement demands linkage of such requirements to abilities, knowledge, or other attributes associated with job performance. Gibson and Prien (1977) demonstrated criterion-related validity of minimum qualifications which were determined following the conduct of job analyses designed specifically to identify such attributes. Previous requirements in use by the employer (not derived from job analyses) failed to differentiate more successful from less successful employees. Minimum qualifications are commonly determined, however, without such an analysis, and thus may not differentiate appropriately among job applicants. As a result, efforts to devise post hoc content-oriented linkages, when such requirements are legally challenged, have often been unsuccessful, particularly for jobs which are not professional in nature (Dichter, 1985).

Use in Job Evaluation

Particularly in public organizations, minimum qualifications have traditionally been established by classification and compensation specialists, using methods of job evaluation that rely heavily on amounts of education and experience necessary for "satisfactory" job performance. The judgmental processes involved in making such determinations are ambiguous, but probably focus on educational and work history backgrounds which are typical of employees in the job rather than qualifications that can accurately be described as minimum.

A Problem of Passing Points

Also common is the assumption that there is one best education and experience requirement appropriate for a job. The assumption presupposes a relationship between amount of education or experience and relevant job knowledge or abilities (or job performance), as illustrated in Figure 1.

Figure 1a depicts an ideal hypothetical relationship which, if typical, would readily permit establishment of a minimum qualification requirement (Johnson, Guffey, & Perry, 1980). Unfortunately, as will be discussed in greater detail shortly, such relationships are probably uncommon. More likely are relationships similar to that depicted in Figure 1b, in which the relationship is smaller in magnitude and accelerates less rapidly.

Determination of an appropriate minimum qualification requirement is analogous, in many respects, to the setting of a passing point for a paper-and-pencil test. It is unlikely that one specific amount of education or experience clearly distinguishes "qualified" from "unqualified" candidates. It is, instead,

a problem of deciding which among a range of several options is most appropriate, given other facts about the job, the employment process, and the setting (cf. Cascio, Alexander, & Barrett, 1988).

Equity

Perceptions of fairness also play a role in many settings. To employ or promote an individual possessing less training and experience than another is often viewed as inequitable and unfair; the person with more training, experience, or seniority, for example, "deserves" to be hired or promoted. Similarly, it is "unfair" to require, say, two years of experience as a minimum requirement to be an "Accountant 1" in an organization, and impose the *same* two-year requirement to be considered as an "Accountant 2;" many organizations, particularly in the public sector, would require more than two years' experience for the higher level job simply as a matter of organizational policy.

Tradition

In many professional occupations, particularly the traditional professions such as law and medicine, specific patterns of education and experience have evolved through which most persons gain entry to the occupation. Guion (1974), in a different context, alluded to such a pattern in considering the validation evidence needed to justify an educational requirement for a heavy equipment designer: is a criterion validity study necessary, or even desirable, to determine that a degree in mechanical engineering should be a minimum requirement for employment in such a position? From a legal perspective, the pressure on employers to "validate" their requirements for such occupations has been much more limited than has been true for most jobs (Dichter, 1985).

The practices described here and the problems they sometimes create probably stem, at least in part, from two characteristics of education and experience credentials: (1) they are at best indirect *indicants*, rather than *direct measures* of competencies (Johnson et al., 1980); and (2) they are visible, readily verified, and widely presumed to be highly associated with job competence (Gottfredson, 1986).

Effective Use of Minimum Qualifications in the Employment Process

In principle, employers could eliminate all use of minimum education and experience requirements from their employment processes, and assess prospective candidates using other, more direct measures of relevant knowledge, skills, and abilities. Such a radical shift in employment procedures, however, is not likely or even desirable. The preferred practice would be careful

use of such requirements based on a better definition of the attributes to be assessed, and consideration of the utility of an education and experience requirement to assess such attributes in comparison to the utilities of alternative assessment methods.

General Education Requirements

Amount of general education is clearly associated with many other variables, including basic language and quantitative abilities. The key issues are (1) whether to assess general or basic abilities through a general education requirement, and if so, (2) the amount of education to require.

In the general population, an estimated correlation between amount of education and general intelligence is about .60, a fact which may account in part for the widespread use of general education as an employment requirement (Gottfredson, 1986). A relationship of this magnitude is unlikely, however, within a pool of applicants for a *particular* job, depending on the nature of the job and the salary offered; in fact, the relationship may approach zero or even be inversely associated with job performance under some conditions. In a concurrent validity study, for example, high ability and well educated employees are more likely to secure promotions or other employment; the obtained correlation between education and job performance is therefore reduced (or becomes negative) as a consequence. Similar distortions from population correlations occur in applicant samples, particularly when pay or other inducements are not competitive (Johnson & Waldron, in preparation). Paradoxically, then, an employer might wisely decide to assess basic abilities using an alternative method, such as paper and pencil tests, and adopt a very minimal general education requirement, or none at all, if the salary offered is not highly competitive.

If a general education requirement is considered a sound alternative because of the numbers of candidates and the cognitive demands of the job, a job analysis to specify *clearly* the nature and level of basic abilities associated with job performance should be carried out. Johnson (1982), Johnson and Waldron (in preparation), and Prien and Hughes (under review) have described studies based on the use of the General Education Development (GED) Scales (U.S. Department of Labor, 1972) and suggested procedures for doing so. Job analysts or incumbents rate job tasks (or the entire job) on the 6-point GED scales for mathematical demands and for language demands—attributes associated with level of general education. Prien and Hughes (under review) have also devised a method of determining the educational equivalence (in years) of each point on the GED scales, an important advance in their use. Having educational subject matter experts (teachers) utilize a mixed standard scale format, a linkage between the GED scales and years of general education was established in one setting and replicated successfully in a second setting.

These and other uses of job analysis information to identify general educational requirements, however, cannot result in precise statements of what the requirements should be; they can only provide a range of potential "passing points" from which to choose and a basis for making the choice.

Specific Education or Training Requirements

The requirement of a specific kind of education or training implies the assessment of specific knowledge or skills. Such competencies might be relatively narrow and simple, such as knowledge of trigonometric functions or the proper administrative procedures for a group employment test, or broad and complex, such as knowledge of gastroenterology or methods of analyzing consumer spending patterns. Three questions arise concerning the use of such requirements: (1) the extent to which the knowledge or skill should be required among applicants rather than trained by the employer; (2) the extent to which the credential specified by the requirement excludes others who have also acquired a comparable level of knowledge or skill; and (3) the amount of training or education appropriately required.

The first question—the extent to which the knowledge should be required among applicants rather than trained by the employer—requires consideration of the availability of persons possessing the training in the labor market, and the costs of providing the necessary training. The second can be more complex; if a pattern of education is not the *exclusive* means of acquiring the knowledge and abilities sought, requiring that pattern necessarily reduces the job-relatedness and utility of the requirement. For many occupations, defining such patterns is increasingly complex. Course and job titles, for example, though common in public sector statements of requirements, are increasingly subject to challenges by candidates because they are not the exclusive means of acquiring or demonstrating the competencies sought. A knowledge of elementary statistics, for example, may be acquired through a variety of different educational experiences.

The amount of training to require has been particularly troublesome for more demanding jobs, especially those requiring a breadth of knowledge not readily defined through linkage of specific job knowledge with the content of specific courses. This problem has been particularly difficult in determining whether to establish or defend on the basis of content the requirement of an academic major, such as "a Bachelor's Degree with a major in Accounting." Such requirements are appropriate if the job analysis supports one or more of the following conclusions:

1. *The job requires a breadth of knowledge comparable to that demanded by the "traditional" professions.* Examples are law, medicine, economics, meteorology, engineering, psychology, and microbiology.

Such jobs typically require exercise of considerable discretion by the employee, and the application of a complex body of knowledge which can be acquired only through an extensive sequence of educational courses.

2. *The knowledge and ability requirements are difficult to assess by other means.* Many knowledge domains are sufficiently vast that it would be impractical for an employer to devise an appropriate paper and pencil test, list of relevant courses, or job simulation to assess candidates.
3. *The consequences of not requiring an academic major are severe.* For example, in order to minimize the likelihood of errors affecting the health and safety of others as a consequence of insufficient knowledge of relevant principles, facts, or procedures, a major in microbiology might well be a reasonable minimum requirement for a job in a health laboratory, even if four or five specific courses could be identified as essential to the knowledge requirement.
4. *The academic major under consideration is the exclusive means of acquiring required knowledge.* As noted earlier, failure to consider all educational patterns resulting in acquisition of the knowledge sought inappropriately excludes some candidates from consideration.

Experience Requirements

The recommendations concerning use of experience requirements are logically analogous to those described above for education and training. An exception is use of the Specific Vocational Preparation (SVP) scale developed by the U.S. Department of Labor (1972). The SVP scale is used to rate the amount of specific training or experience, beyond general education, demanded by a particular job. In contrast to the GED scales discussed earlier, the SVP scale, though seemingly useful as an aid in determining experience requirements, lacks job-relevant anchors permitting a content-oriented linkage to job tasks or associated knowledge or skill demands.

As the review of relevant research suggests, empirical evidence exists confirming relationships between the *relative* amount of experience and job performance. Much of this evidence, however, is based on experience in the *same* job, rather than experience in *different* prior jobs. The content relevance of prior work experience, and the appropriate amount of such experience to require, are critical decisions.

Conclusions

While use of a minimum amount and kind of training and experience is often an appropriate requirement, there are alternatives which may be advantageous both to employers and prospective candidates. Among these are

use of *preferred* qualifications (sometimes used in addition to minimum qualifications), and explicit expressions in recruitment efforts of the *competencies* sought by the employer. These alternatives permit applicants the opportunity to exercise self-selection, and provide to applicants in advance at least part of the rationale for potential rejection. Because minimum qualification requirements are analogous to passing points in other assessment procedures, it is also possible to devise methods for assessing relevant training and experience to produce scores which can rank order candidates; a passing point in such a procedure can be adopted as a "minimum qualification." Before further discussion of these options, descriptions of various T&E evaluation methods are provided.

METHODS OF T&E EVALUATION

Holistic Judgment

The method used most frequently to evaluate applicant training and work experience is *not* a formally scored T&E evaluation method, but the rather cursory review of the application or resume. An employment manager, line manager, or some other individual reads one or more completed job application forms or resumes and makes a rather general judgment about the relative suitability of the applicant(s) for employment in the target job. Depending on the stage of the employment process at which this review takes place, the judgment may be something like one of the following: "not well suited—screen out," "may be well suited—screen further and seek additional information," "strong in some areas but weak in others—discuss with other people involved in the hiring decision," "well suited, best of the current crop of candidates—make offer," and so forth. This informal approach is addressed here *not* because it is a practice to be endorsed, but simply in deference to its pervasive use.

The key feature of this *holistic judgment* approach to evaluating job applicant training and work experience is its *unstructured* nature. That is, the specific bits of information used to make judgments about applicants and how those bits are combined to arrive at the judgments are determined by the individual T&E evaluator. In terms of the Levine and Flory (1975) conceptual framework discussed earlier, the holistic judgment method could be classified in any of the four "judgmental" cells (high or low job relatedness, surface characteristics, or inferred traits), depending on what the evaluator uses as a basis for making the evaluation. This method is strikingly similar to the unstructured employment interview. Note that, while any form of T&E evaluation can take place independently of the employment interview, virtually every employment interview involves the holistic judgment method of T&E evaluation in that all interviewers have the application or resume of the

interviewee at hand (cf. Dipboye, Fontenelle, & Garner, 1984; Tucker & Rowe, 1979).

In practice, evaluating the validity and reliability of the holistic judgment method is problematic. Note that this method is not covered in the next section on research on T&E evaluation methods due to the paucity of published data. Both the criteria and the process for making judgments about applicants from resume data exist only in the minds of individual evaluators, and hence, are both unrecorded and unstandardized. In addition, many T&E holistic judgments are inextricably confounded with information gained during employment interviews or other social interactions (e.g., dinner with the job applicant).

Traditional Point Method

The most prevalent formal T&E evaluation method used by public sector jurisdictions is the traditional point method (Cook, 1980). As described by Porter et al. (1976), the point method most often used in public personnel agencies consists of a mechanical formula set out in a formal schedule. Points are credited for the number of months or years of different kinds of relevant training, education, and experience. Most typically, applicants are assigned a base (passing) score of 70 if they meet minimum qualification requirements. Additional points are then added for months or years of specified types of training and experience. Different types of training and/or experience are assigned differential point values. (Hopefully, these differential values are determined through consultation with subject matter experts during a thorough job analysis.) These are typically multiplied by the respective number of months or years associated with each applicant's background, then summed to arrive at the applicant's score. Each score that is derived tends to be unique, and the collection of scores tends to be distributed continuously over a 30 point range (from 70 to 100).

By way of illustration, inspect the rating schedule shown in Table 1. The first paragraph describes minimum qualifications for the job while the second and third paragraphs cover substitutions and equivalences. Next follows the general schedule of points, illustrating the different point values for different types (levels) of experience, education, and technical training. This is followed by specific education and experience schedules. These define what constitutes "A" level, "B" level, and "C" level education and experience for the target job. Last is the scoring procedure and guidelines. In short, the T&E evaluator first sees if the applicant meets MQs. If so, the evaluator then proceeds to score the applicant's training and experience according to the guidelines and various schedules.

Table 1. Phantom County Traditional Point Method for Automotive Equipment Repair Foreman (AERF)

Minimum Qualifications

Graduation from a standard high school and four years of experience as an automotive mechanic.

Full-time paid automotive mechanic experience may be substituted for the required high school education on a year-by-year basis.

An equivalency diploma issued by a state department of education or by the United States Armed Forces Institute may be substituted for high school graduation.

General Schedule of Points

Points are to be awarded as follows:

	<i>Experience</i>		<i>Advanced Education</i>	<i>Technical Training</i>
	<i>Per Year</i>	<i>Per Month</i>	<i>Per 30 Semester Hours or 45 Quarter Hours</i>	<i>Per 720 Classroom Hours</i>
A level	3.0	0.25	3.0	3.0
B level	1.5	0.13	1.5	1.5
C level	0.75	0.06	0.75	0.75

Educational Schedule

- A. Automotive or mechanical trades course work
- B. Building construction or maintenance trades course work

Experience Schedule

- A. Highly skilled and supervisory mechanical work in the maintenance and repair of a wide variety of automotive, construction and heavy equipment in an automotive shop.

Example: Automotive Equipment Mechanic II

- B. Skilled mechanical work in the maintenance and repair of automotive and related equipment.

Example: Automotive Equipment Mechanic I

- C. Routine skilled work on the maintenance of automotive and related equipment.

Example: Automotive Serviceman

Scoring Procedure and Guidelines

1. a. Review the minimum qualification requirements and determine whether or not the applicant meets the minimum qualifications.
- b. If MQs are met, assign a base score of 70 and proceed to step 2.
- c. If MQs are not met, designate the applicant as "NOT QUALIFIED" and go on to the next applicant.
2. a. Review experience and education *above* the MQs. For each determine whether it (education or experience) is A level, B level or C level according to the attached schedules. Then add points as specified to the base score of 70.

(continued)

Table 1 (continued)

-
- b. For experience, do *not* count the first month if exact dates are not given (e.g., "July, 1979-August 1980" is equal to 13 months).
 - c. No credit is given for part of a month.
 - d. Majors in college are credited, not minors.
 - e. Technical training receives credit only if exact dates and classroom hours are listed.
 - f. The applicant may receive one year's credit for each 2 years of college if credit hours are not specified on the application. Points are awarded only for 30 semester hours credit chunks (or 45 quarter hours credit), *not for fractions of 30 (or 45) hours.*
-

In a later section, it is noted that the validity of the traditional point method T&E procedures is quite low, and not significantly different from zero in a nontrivial number of instances. Reflecting upon these procedures for a moment, one can understand why they generally lack validity. The apparent precision of the traditional point method is certainly more specious than real. Even in the very best of situations where highly detailed applicant information is available, the effect of different or even identical training and work situations on different individuals cannot be measured with high accuracy (Ash & Levine, 1985; Porter et al., 1976). One may question the argument that an applicant who scores 87 is more qualified than the applicant who scores 86 or 85 for the target job in Table 1. How does one know what additional KSAs the applicant with the score of 87 gained, if any, in those additional four months of "A" level work experience that the applicant with the score of 86 does not have? Two different people taking the same course at the same time and earning the same grade may not gain identical amounts of knowledge from the course. Similarly, two different people working in the same job for the same organization may not gain the same KSAs by virtue of this similar experience. Such measurement error in traditional point method scores is even more severe when one considers that the quality and content of knowledge gained at different educational institutions and employers can be very different. For example, the knowledge gained from earning a bachelors degree in engineering will vary with the demands of the schools' engineering curricula.

In addition to the potential problems of lack of job-related and specious precision, the traditional point method may often result in adverse impact against females (Ash & Levine, 1981), minorities, and young applicants (Porter et al., 1976). The traditional point method generally results in higher scores for applicants with the greatest amounts of work experience and education. To the extent that only a portion of the work experience or training is related to successful performance of the target job, minority, female, and young applicants may be well suited for the job but score lower because they have not had the time to accumulate as many experience or training points as white, male, and older applicants.

Table 2 Basic Grouping Training and Experience Evaluation Guide

 Job Title: *Automotive Equipment Repair Foreman* Classification Code: 2805

 Evaluation Characteristics of Applicants Meriting the Indicated Evaluation

95	Three months of motorized vehicle repair schooling, and 5 years of experience in equipment repair on all types of systems (transmissions, brakes, etc.), which must have included heavy equipment repair (e.g., draglines, bulldozers, caterpillars) in addition to servicing a variety of motor equipment (cars, trucks, gas and diesel). At least two years of the five must have been in a lead worker or supervisory capacity. This experience must have occurred all within the last ten years. At some point in the work history there should also be evidence of ordering parts and supplies, writing service orders, etc. to indicate the capacity to do paperwork.
90	No schooling but experience in 95 category.
80	Four years of experience as in the 95 group without evidence of recency or lead worker/supervisory experience or paperwork handling.
70	Four years of automotive repair experience which consisted of one make and one type of vehicle or specializing in only one function (e.g., brakes, transmissions).
Not Qualified	MQs not scored. All other applicants should be placed in this category (e.g., automotive serviceman). MQs: Four years of experience as an automotive mechanic.

The Grouping Method

In the basic grouping method (as described by Porter et al., 1976), applicants are usually divided into a small number of groups on the basis of *simultaneous* consideration of training and experience. Rather than being ranked vertically over a continuous range, applicants assigned to each respective group are assigned the same score. For example, if the applicants are divided into three groups, those in the high group might all be assigned a score of 90, those in the middle group a score of 80, and those in the low group a score of 70. The basic idea of the grouping approach is to identify the well-qualified and well-suited candidates for position vacancies, and to the extent feasible, allow all of them to be considered by the hiring authority. The number of groups, the score differences, and the nature and complexity of each grouping T&E evaluation plan are determined by such factors as the outcomes of a job analysis, the probable number of applicants, and the probable turnover in the job classification, among others.

By way of illustration, inspect the basic grouping T&E evaluation guide shown in Table 2. A T&E evaluator using this rating guide would examine applicant data from a standard application form or resume, and then determine which of the five categories described in the guide matched best the background of the applicant. The applicant would then be assigned the score for that category. In the particular example shown in Table 2, the *Not Qualified* category contains the formally stated minimum qualifications (MQs), "Four

years of experience as an automotive mechanic," and the notation "MQs not scored." In this case, the formally stated MQs are too general to permit treating them within a single scoring category. Depending on the specific nature of the experience, four years of experience as an automotive mechanic may be scored as 70 or 80.

As compared to the traditional point method, grouping applicants may make the rating procedure conform better to the actual level of precision attained in T&E evaluation and measurement based on data obtained from traditional application forms and resumes (cf. Ash & Levine, 1985; Porter et al., 1976), and thereby conform more closely to a decision-making model of selection where individual applicants are categorized into more and less suitable groups from which a hiring supervisor or manager can choose (cf. Cronbach & Gleser, 1965). Another important aspect of the basic grouping approach is that applicants receive no credit for experience and education beyond optimal levels specified in the rating plan. On the other hand, grouping methods are similar to point methods in that types and amounts of both training and experience are used as indicants of competencies which, in theory, are related to job performance criteria.

The Behavioral Consistency Method

Perhaps more than any other set of T&E evaluation methods, the behavioral consistency methods are based on the assumption that past behavior is the best predictor of future behavior. In addition to *behavioral consistency* (Schmidt, Caplan, Bemis, Decir, Dunn, & Antone, 1979), this particular approach to T&E evaluation has also been referred to as the *achievement history questionnaire* method (State of Wisconsin, 1979) and the *accomplishment record* method (Hough, 1984). Although there are some subtle differences among these specific variations, the goal of behavioral consistency type methods is to rank order applicants on the basis of the kind of achievement behaviors that are required for superior performance in the target job.

In general, the method begins with the development of major achievement dimensions, usually from five to ten. In the Schmidt et al. (1979) procedure, these dimensions are derived by combining KSAs generated by subject matter experts. In the Hough (1984) approach, the dimensions are derived by grouping critical incidents (Flanagan, 1954) generated by subject matter experts. For each major achievement dimension, applicants are encouraged to describe in detail at least two past achievements which best demonstrate their capabilities. Applicants are asked to include the following information for each achievement: (1) what the problem or objective was; (2) what he/she actually did and when; (3) what the outcome or result was; (4) the percentage of credit he/she claims for the outcome; and (5) the name/address/phone number of someone who can verify the achievement.

Once achievements have been collected from an applicant or present employee group, a sample of the achievements is subjected to a scaling process similar to that used in deriving behaviorally anchored rating scales (e.g., Campbell, Dunnette, Arvey, & Hellervik, 1973; Smith & Kendall, 1963). First, the achievements are classified into the achievement dimensions independently by subject matter experts. Only achievements which are assigned consistently to respective dimensions are retained for further consideration as scale anchors. Next, the surviving achievements are rated by subject matter experts in terms of the level of achievement each represents. On the basis of these ratings, achievements which reliably represent the continuum from low to high accomplishment for each respective dimension are selected as anchors (benchmark achievements) for the achievement rating scales.

These rating scales are used by T&E evaluators to score the achievements supplied by each applicant. Typically, achievement scores are combined across dimensions to derive a single score for each applicant. Theoretically, this score indicates the applicant's relative standing in terms of his or her past level of *behavioral achievement* ("samples" in the terminology of Wernimont & Campbell, 1968) in areas directly related to performance in the target job.

Self-Rating T&E Evaluation Methods¹

The self-rating approach to T&E evaluation is actually a family of methods. What distinguishes them from other T&E evaluation methods is that they require applicants to make judgments or evaluations of their backgrounds in terms of tasks they have performed or KSAs they have acquired. The self-rating method of T&E evaluation seems to have been developed in the 1950s by Ernest Primoff in his "job element" examining for blue collar trades for the federal government (Primoff, 1975). Three variations of the self-assessment approach are described here. In general, self-rating T&E evaluation methods are time consuming to develop, *only* in that they require a detailed job analysis with specific task or KSA statements which serve as the basis of the self-rating. They are efficient to score in that they are easily adapted so that they can be machine scored using an optical scanner. The most significant problem with the self-assessment methods of T&E evaluation is that they are highly subject to inflation bias on the part of applicants (Anderson, Warner, & Spencer, 1984).

Task-Based Methods

The task-based approach relies on the premise that adequate validity can be achieved by obtaining detailed information on specific tasks that an applicant has performed in the past, regardless of the job in which the task was performed. In this approach, then, job tasks serve as indicants of past performance which are presumed to predict future performance. In its simplest

form, the task-based approach is operationalized by means of a supplemental application form consisting of a task inventory composed of nontrivial tasks (as determined through job analysis) in the target job. Applicants are asked to check the tasks they have performed in the past. In more complex forms of this approach, applicants may be asked to rate the tasks using a more in-depth scale such as: "0—I have not performed this task; 1—I have performed this task under close supervision; 2—I have performed this task independently, seeking advice in only the most difficult situations; 3—I am considered an expert in this task by others."

Applicants also may be asked to specify the setting or job in which they performed each task for verification purposes. Some task-based approaches even incorporate "inflation" or "frankness" scales by including bogus tasks in the task listing (Anderson et al., 1984; Farrell, 1979). Applicants who report performing nonexistent tasks are judged to be falsifying their responses. An applicant's score is a weighted function of the ratings provided.

KSA-Based Methods

These methods are similar to the task-based methods except that applicants check or rate KSAs they have acquired, rather than tasks they have performed. KSAs important for consideration in selection are identified by means of a job analysis (e.g., Primoff, 1975) and arranged in a format that requires the applicant to first indicate the level of each KSA he/she possesses (e.g., on a 5-point scale ranging from "0—I know little or nothing about this" to "4—I possess this KSA at a superior level and am called on to do unusually difficult jobs requiring it"). Then the applicant must specify the jobs or training courses through which he/she acquired the nonzero level of each KSA. As in the task-based method, an applicant's score is a weighted function of the ratings provided. Johnson et al. (1980) suggested a simple checklist on which the applicant simply checks whether or not he/she possesses each KSA as a useful way to implement the KSA-based method.

The Illinois Job Element Method

P. Ash, Taylor, and Hoel (1973) proposed an approach to job element examining which is used in the University Civil Service System of Illinois. The Ash et al. approach allows a variety of measurement methods to assess an applicant's job-related KSAs. The Illinois civil service examinations include T&E evaluations, written tests, interviews, and performance tests. Although the point method is sometimes used in the Illinois examinations, the primary T&E evaluation method is a scored "biographic element." McDaniel, Schmidt, and Hunter (1988a) labeled this T&E evaluation method the Illinois job element approach. It is described separately here because McDaniel et al. located and

Table 3. Outline of the Primary Job Analysis Questions/Instructions for the Improved Point Method

-
1. Define the major tasks performed to carry out the responsibilities of the job.
 2. What specific characteristics (skills, knowledge, abilities) are necessary to perform these tasks well?
 3. What specific experience, training, or other factors in a persons background would show that a person is able to do this job?
 4. What are the working conditions (e.g., use of uniforms, travel, working outdoors, high production speed, interpersonal stress, etc.)?
 5. Describe the critical situations that occur on this job. What situations clearly separate the best from the worst incumbents in terms of how they handle the situations?
 6. What specific characteristics in the workers (their skills, knowledge, abilities) make the difference between handling the situation well or poorly?
 7. Think of the best performers of this job you have known. What characteristics (skills, knowledge, abilities) make them superior workers? What things in their background (experience, training, etc.) made them able to perform this job so well?
 8. Think of some poor performers of this job. What characteristics do they lack? What is lacking in their backgrounds?
 9. Categorize all the above tasks and characteristics. Try for four to eight categories.
 10. Determine weightings for categories. (All else being equal, is characteristic X more, less, or equally desirable as characteristic Y?) If possible, try to group and arrange the specifics so the categories will be of equal weight.
-

reported the results of 16 criterion-related validity studies conducted on this particular method. In this approach, applicants describe their experience and then provide a self-assessment on each KSA dimension. A T&E evaluator reviews the job experience description to determine if it supports the self-rating on the respective dimensions. If so, the self-rating is the applicant's score on the particular biographic element; if not, the applicant is given no credit on that element.

Improved Point Method²

A rather detailed method of T&E evaluation has been developed and implemented by the State of Washington Personnel Department (Swander & Schultz, undated)—the *improved point method*. It begins with the use of a semi-structured interview type of job analysis to find out from subject matter experts the important KSAs in the target job, and the *specific things* that a person would have done in the past which indicate that the person possesses these important KSAs. Table 3 shows an outline for this job analysis approach. A key aspect of these job analysis interviews involves getting information stated in terms of KSAs. For example, if a subject matter expert says that "someone must have a college education to do the job," the job analyst should ask, "What

is it that is gained in college that would make a person successful on this job?" Answers like "ability to learn a complicated body of information," or "knowledge of engineering procedures," are examples of more specific KSAs that are desired.

The next step in the improved point method involves incorporation of *specific things*—which, if done by an applicant in the past, indicate that the applicant possesses particular KSAs of interest—into a T&E scoring key. Particular emphasis is placed on developing the scoring key in such a manner that all applicants who possess the qualities needed on the job have a fair chance to get credit for them. Thus, it is important that the T&E developer do a reasonable job of providing for ways to credit the qualities in the scoring key so that the majority of persons who possess them will have them detected in the T&E evaluation. As an example, consider the job of Mental Health Community Programs Chief. An applicant's work experience is scored for KSAs relating to six areas: (1) administration; (2) planning and budget; (3) coordination; (4) laws, regulations, and judicial system; (5) management or supervision of service delivery; and (6) professional involvement. For the fourth area—laws, regulations, and judicial system—an applicant will receive one point credit *each* if he/she has done the following *specific things* in a job held for 12 months or more:

- wrote administrative policy;
- wrote legislation;
- made presentations to legislature;
- worked with legal cases related to commitment;
- prepared legal defense by organizing testimony (evidence);
- contracted for legal services; and/or
- prosecuted or defended clients in court as attorney or employer of an attorney.

Similarly, the T&E evaluator will review the application materials of the applicant checking for and crediting other *specific* types of experience under each of the five remaining areas.

Because of the relatively detailed information required from applicants to score their backgrounds correctly using this method, it is necessary to tell them how to describe relevant experience (and occasionally education) in enough detail. Instructions include a listing of the KSA areas being assessed, plus up to two pages of general and specific instructions. The specific instructions for the job Right-of-Way Agent include the following:

In addition to the completed standard application, we need to know the details of your experience in the areas listed below. Prepare your responses on additional sheets of paper ...

1. Negotiating land use leases, agreements, acquisitions, sales, and easements with (a) landowners, (b) governmental agencies, (c) corporations.
2. Property management and appraisal (specify if this was public property).
3. Oral presentations to legislative and other state committees (e.g., the Interagency Committee on Outdoor Recreation).
4. Writing reports.

The improved point method appears to be somewhat more promising than the traditional point method or the holistic judgment method in that it focuses to some extent on scoring behaviors underlying the KSAs judged to be required for successful performance in the target job. Note this method's similarities in some respects to the behavioral consistency and task-based self-assessment T&E evaluation methods discussed previously. Unfortunately, the authors are not aware of any studies assessing the criterion-related validity or reliability of the improved point method.

General Summary and Comparison of T&E Evaluation Methods

The most pervasively used approach for determining the suitability of applicants for employment in particular jobs through assessment of education, training, and work history is the unsystematic and unstandardized *holistic judgment* method. The formal approach to training and work history assessment that is predominant in public sector organizations is the rather simplistic *traditional point* method. Both of these approaches to T&E evaluation have been criticized for serious conceptual and measurement problems (e.g., Ash, 1984; Ash & Levine, 1981; Johnson et al., 1980).

The *grouping* method appears to offer some improvement over the traditional point method in that its level of precision is more appropriate to the nature of the data used as the basis for T&E scores. However, the grouping method is similar to the traditional point method in that it uses *credentials*—indirect and fallible indicants of applicant competencies.

The *improved point* method, although relying on credentials to a limited extent, focuses more on *specific* job behaviors or duties as indicants of job-related KSAs. It appears to constitute a less arbitrary and more acceptable approach to allocating rather specific point scores to applicants on the basis of their respective backgrounds. As yet, however, this method has not been subjected to the rigors of empirical reliability and validity research.

The *self-assessment* methods, both *task-based* and *KSA-based*, share the problem of potential inflation bias, which appears to be a significant factor in limiting their application. Yet, these methods appear to offer more direct indicants of job competencies than the traditional point and grouping methods.

The task-based approaches, in particular, seem quite justifiable on the basis of content validity.

The *behavioral consistency* methods appear relatively sound in many respects. The behavioral emphasis in the job analysis and rating scale development phases works in favor of the content validity of selection decisions based on these methods. On the other hand, behavioral consistency scores are affected by applicant memory, verbal ability, and writing skill. Thus, this method retains content validity only when applied to jobs in which these three KSAs are of nontrivial importance (Ash, 1983). Also, applicant completion rates for the behavioral consistency method application supplements are often substantially lower than for standard application forms or self-assessment supplements (cf. Ash, 1983; 1986; Schmidt et al., 1979).

In examining various methods of T&E evaluation, it is important to emphasize that these methods apparently measure different things in applicant background data, and the rank order of applicants can be affected substantially by using one approach instead of another. Correlations between behavioral consistency and traditional point method scores are consistently zero or slightly negative (Ash & Levine, 1981; Johnson et al., 1980). Correlations between behavioral consistency and task-based scores are consistently moderate, ranging from .36 to .54 (Ash & Levine, 1981; Johnson et al., 1980). Correlations between behavioral consistency and grouping scores range from zero to .29 (Ash & Levine, 1981; Schmidt et al., 1979). The only correlation found between behavioral consistency and KSA-based scores was essentially zero, but the correlation between grouping and KSA-based scores from the same study was .47 (Schmidt et al., 1979). Correlations between grouping and task-based scores are moderate, ranging from .28 to .53, while correlations between grouping and traditional point method scores are high, ranging from .60 to .74 (Ash & Levine, 1981). Correlations between task-based and traditional point scores range from .10 to .49 (Ash & Levine, 1981; Johnson et al., 1980). The only correlation found between behavioral consistency and holistic judgment scores was .36 (Ash, 1984).

Next the reliability and validity of the various T&E evaluation methods are examined.

RESEARCH ON RELIABILITY AND VALIDITY OF T&E EVALUATION METHODS

Reliability of T&E Evaluation Methods

The reported reliabilities of T&E measures are usually high. Ash and Levine (1985) provided 12 estimates of the reliability of T&E procedures, revealing a mean inter-rater reliability of .83. Their reliability for the task method was

near 1.0, probably due to the lack of rater judgment needed for the scoring process. The review by Schmidt et al. (1979) found inter-rater reliabilities in the .80s to be typical of point system T&E evaluation methods. However, the Ash and Levine (1985) and Schmidt et al. (1979) reliabilities reflect the correlation between evaluators scoring the same applicant responses (i.e., conspect reliability, see Cattell, 1971). A more appropriate (and probably lower) estimate would be obtained if applicants retook the T&E evaluation instrument. Such a reliability estimate would include an assessment of error variance due to intra-applicant variability. However, such reliability estimates are not available.

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 et al., 1979; State of Connecticut, 1978). There are only a few literature reviews of the criterion validity of T&E examinations. Schmidt 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 et al. (1979), which involved traditional T&E evaluation methods, and offered .13 as the average validity coefficient after correction for unreliability in supervisors' ratings.

Ash and Levine (1985) examined the criterion-related validity of four T&E evaluation methods (i.e., traditional point, grouping, task-based, and behavioral consistency) across three occupations. They termed their validities "quasi validities" since their criterion was a peer nomination score derived from statements of who would make the best supervisors if promoted. Only the grouping method produced significant quasi-validity coefficients ($r = .21$ and $r = .30$) and for only two of the three jobs under study.

The most comprehensive summary of the validity of T&E evaluation methods for personnel selection is presented in McDaniel, Schmidt, and Hunter (1988a). Using meta-analysis methods (Hunter, Schmidt & Jackson, 1982), McDaniel et al. summarized the validity evidence for four T&E evaluation methods (traditional point, Illinois job element, task-based, and behavioral consistency) based on 132 validity coefficients with a total sample size of 12,048.

McDaniel et al. (1988a) noted that meta-analyses of formal T&E evaluation method validities was different from meta-analyses of ability constructs (e.g., verbal ability) in four major ways, and those differences are important for the interpretation of meta-analytic results. The first three of the four major differences were that the McDaniel et al. meta-analysis had less control over

three sources of variance: heterogeneity of (a) constructs measured, (b) measurement process, and (c) occupational categories. The fourth difference between the McDaniel et al. meta-analysis and past validity generalization studies was the number of validity coefficients available for analysis. While these four differences were detailed in McDaniel et al., they are presented here in brief form because of their importance in interpreting the validity of T&E measures.

Heterogeneity of Measured Constructs

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. T&E evaluations, like paper and pencil tests, may measure different constructs (e.g., cognitive ability, interpersonal skills). Unfortunately, researchers seldom report detailed information on the constructs assessed by the T&E evaluation method under study. Although researchers may partition T&E evaluations by method, such a categorization can only permit a gross content division, because within each method category heterogeneous groups of constructs are measured. Thus, a meta-analysis of existing validity evidence for T&E measures cannot provide much information on the validity of specific constructs measured by them, although it may provide 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.

Heterogeneity of Measurement Process

The McDaniel et al. (1988a) meta-analysis also differs from most previous validity generalization research on ability measures because there is more variability in how T&E data are collected than there is in how ability data are collected. Although different paper and pencil measures of a given ability may vary slightly in the measurement process (e.g., different item types may be used), 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 resumes, others from traditional job application forms, and still others from structured supplemental application forms. While the T&E evaluation method categories used in McDaniel et al. (1988a) were meaningful ones, the categories were not perfectly homogeneous. For example, among task method evaluations, some scoring

schemes focused on time spent performing the task, while others focused on self-assessments of task skill.

Heterogeneity of Occupations

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) or job attributes, such as the level of cognitive demands placed on an employee (Gutenberg, Arvey, Osburn, & Jeanneret, 1983; Hunter, 1980). McDaniel et al. (1988a) noted that too few validity studies have been conducted on T&E measures to permit separate meta-analyses by job attributes or job content category.

Number of Validity Coefficients

Although McDaniel et al. (1988a) assembled a substantial number of coefficients (132), the number is small relative to most validity generalization research. Furthermore, when McDaniel et al. (1988a) divided the coefficients into formal T&E evaluation method categories, they were left with fewer coefficients to analyze. This leaves their results open to distortion caused by second-order sampling error (Schmidt, Hunter, Pearlman, & Hirsch, 1985, Q&A No. 25).

The three sources of uncontrolled variance made 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 nonartifactual variance (i.e., moderators). In the McDaniel et al. (1988a) study, the variance remaining after correcting for statistical artifacts could be due to moderators or to differences among studies in (a) the constructs measured by the T&E evaluation, (b) the T&E evaluation measurement process, and (c) uncontrolled job attribute and job content differences. The first three of the four major differences between the McDaniel et al. and past validity generalization studies also affected how one could interpret the validity generalization results. McDaniel et al. have defined a measure to show 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 three uncontrolled sources of variance in the McDaniel et al. study probably caused an overestimation of true variance and a resultant underestimation of the lower bound value of the true validity generalization. This probable underestimation of the lower confidence value

Table 4. Validity of T&E Evaluation Methods for Job Performance Criteria

<i>T&E method distribution</i>	<i>Total N</i>	<i>No. r's</i>	<i>Mean r</i>	<i>Obs. SD</i>	$\hat{\rho}$	$SD_{\hat{\rho}}$	<i>90% C.V.</i>
All cases	12,048	132	.09	.16	.17	.22	-.12
Point	6,741	91	.06	.17	.11	.24	-.20
Illinois job element	3,168	16	.11	.07	.20	.00	.20
Task	991	10	.08	.18	.15	.27	-.19
Behavioral consistency	1,148	15	.25	.12	.45	.10	.33

Note: The mean true validity ($\hat{\rho}$) was corrected for range restriction and unreliability in the criterion. The variance of the true validity distribution ($SD_{\hat{\rho}}$) was corrected for sampling error and for differences among studies in predictor reliability, criterion reliability, and range restriction.

Source: Adapted from McDaniel et al. (1988a).

would make the McDaniel et al. conclusions concerning validity generalization conservative.

Table 4 summarizes the McDaniel et al. (1988a) results. The first column of the table identifies the T&E distribution analyzed. The next four columns of data show the sample size, the number of validity coefficients in each distribution, and the mean and standard deviation of the observed distribution. Columns 6, 7, and 8 present the estimated mean ($\hat{\rho}$), standard deviation ($SD_{\hat{\rho}}$), and 90% credibility value for the distribution of true validities. The mean true validity is corrected for range restriction and unreliability in the criterion. 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.

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 indicated that T&E ratings as a whole have only moderate mean validity which cannot be generalized across situations, since the value at the 10th percentile of the distribution is negative (-.12). An inspection of the results for specific T&E evaluation methods revealed that some of the variation in the distribution of all coefficients results from the type of T&E evaluation method employed.

The point method studies, which compose 69% of all known validity coefficients for T&E ratings, showed a mean true validity of .11 and a relatively large standard deviation of .24. The 90% credibility values for the point distribution was -.20. Thus, the point method has a low mean validity and lacks generalizability.

The Illinois job element 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. The 90% credibility value (.33) supported a conclusion of validity generalization for the behavioral consistency method. McDaniel et al. (1988a) concluded 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.

McDaniel et al. (1988a) cited recent theoretical and causal modeling work by Schmidt, Hunter, and Outerbridge (1986) that provides a clue to the source of the large variance in the point and task distributions. Schmidt et al. (1986) argued that it is *relative* individual differences in job experience which cause individual differences in job performance. They proposed that these relative individual differences in job experience decrease as the mean level of job experience in a sample increases. In brief, their 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 the decay in the validity of job experience validities is reported by McDaniel (1986) and McDaniel, Schmidt, and Hunter (1988b). 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, scores are expected to be at least moderately correlated with the length of job experience. Ash and Levine (1981) reported correlations between point and task-based T&E scores of .21, .22, and .49. The longer one works in an occupation the more opportunity one will have to perform and gain skill at various tasks. Thus, length of job experience may moderate both task and point method validities.

An additional analysis presented by McDaniel et al. (1988a) permitted a partial test of the job experience moderator hypothesis. A study by Molyneaux (1953) provided the grade level of samples used in 51 point method validity studies. 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. McDaniel et al. (1988a) showed that the mean validity was substantially higher for the samples from the lower grade levels (.29) than in the middle grade levels (.10) and in the highest grade levels (-.03).

McDaniel et al. (1988a) argued that, if the validity of point and task methods decays with increasing levels of job experience, their conclusions about the lack of validity generalization for these two methods could be too broad. The validity of the point method may be generalizable for situations where the mean experience of the applicant population is low, but relative individual differences in amount of job experience exist. They proposed that future research should examine whether point and task methods of T&E evaluations show higher validities for applicant pools with low mean levels of job experience. They also called for additional validity studies on all T&E evaluation methods so that meta-analyses could be rerun with larger distributions, and firmer conclusions drawn regarding mean validity and generalizability. In closing, McDaniel et al. sounded an appropriate cautionary note to the effect that their results should best be viewed as tentative, given the relatively small number of primary studies and their study's lack of control over three sources of variance: heterogeneity of (a) constructs measured, (b) measurement process, and (c) occupational categories.

Summary of Reliability and Validity Evidence for T&E Measures

1. Reliability estimates for T&E measures are typically in the .80s as measured by inter-rater agreement. T&E measures that require little rater judgment (e.g., some applications of the task method) can yield higher reliabilities. Inter-rater reliability does not measure the error variance due to intra-applicant response variability and thus probably overestimates the reliability of T&E measures.
2. The validity of T&E measures has traditionally been judged on the basis of content validity.
3. In criterion validity studies of T&E measures, validity may be a function of both measurement method and the content or constructs being measured. Meta-analytic summaries of the criterion validity of T&E measures face several problems, including the relatively few primary validity studies and the lack of control over validity variance due to heterogeneity of (a) constructs measured, (b) measurement methods, and (c) occupations.
4. Based on the relatively small existing data base of criterion-related validity studies, several conclusions are offered. The validity of T&E measures varies with the type of method. The behavioral consistency and Illinois job element methods show moderate validity in the studies that have been compiled, and are potentially generalizable. The point and task-based methods show lower validity that may not be generalizable. Length of job experience may moderate the validity of point and task-based methods, such that the methods show useful levels of generalizable validity for applicant pools with low mean levels of job experience. Little can be said about the criterion-related validity of the

grouping method, because only three quasi-validity coefficients are available for it. Even less can be said regarding the criterion-related validity of the *improved* point method, since virtually no validity coefficients are currently available for it. However, the facts that grouping method scores may conform better than traditional point method scores to the actual level of precision attained in T&E evaluation, and that *improved* point method scores are based on *extremely specific* job-related experiences relative to traditional point method scores, argue for more research on the grouping and *improved* point methods.

5. Many more primary validity studies are needed on the various methods of T&E evaluation before firm conclusions are drawn regarding their effectiveness. Greater attention needs to be focused on the distinction between the content and the method of T&E evaluations.

THE FUTURE OF T&E EVALUATION: RECOMMENDATIONS FOR PRACTICE AND RESEARCH

A Revised Conceptualization of T&E Evaluation

The assessment techniques described in this paper, though broadly characterized as “T&E evaluation” methods, are quite diverse, differing in the kinds of information sought from job applicants, and how that information is used to derive scores. Those methods most widely understood as literally evaluating the training and experience of applicants are based on *credentials*. These form the usual basis for assessing whether candidates meet minimum qualifications, a hurdle which must be passed in order for the candidates to be considered further in the employment process. Credentials also serve as the basis for “scores” assigned in the traditional point and grouping methods of T&E evaluation. The behavioral consistency and self-rating methods (and to some extent the improved point methods) are *not* based literally on past training and experience, but on self-described actions, results or demonstrations of achievements, or acquired KSAs—presumed *outcomes* of prior training and experience.

In contrast to *direct* measures of competencies, such as paper and pencil and performance tests, all T&E evaluation methods rely on self-reports of applicants. They differ, however, both in the nature of the information sought, and the extent to which the information can be confirmed as accurate. The credential-based methods provide information which can most readily be verified through former employers and educational or training institutions referenced by applicants. In principle, so too can the judgments and assertions of candidates in self-report and behavioral consistency methods, but with

greater difficulty stemming largely from the greater subjectivity of such judgments.

Johnson et al. (1980) proposed models of these methods based partly on other distinctions about the nature of information sought from applicants. They argued that differences in validities and attributes appropriately assessed by various methods are probable, and depend on the type of data obtained from applicants (e.g., credentials, KSAs, tasks, or behaviors), the nature of the attribute to be assessed, the method of scoring, and certain characteristics of the applicant population.

Among task-based, self-report methods, applicants are asked both to indicate whether they have performed particular tasks in the past, and usually, to rate their competencies with respect to performance of those tasks. These data about candidates can be predictive of future performance to the extent that *only* those candidates who have performed the tasks listed in the questionnaire have acquired the competencies associated with effective job performance. Such data are likely to be valid indicants of future performance, then, only for some KSAs and in some applicant populations. Further, the response options, the judgments required of applicants, the criticalities and difficulties of the tasks defined in the supplemental application, and the scoring method itself are also likely to affect the validities of scores in predicting performance on particular jobs.

The KSA-based methods are similar in many respects, but they rest on an opposing assumption—that the attributes assessed can be acquired or demonstrated through a variety of *different* experiences or “task performances” in prior work, training, or educational settings, rather than exclusively through past experience with particular, previously defined tasks. The behavioral consistency methods also rely on the premise that possession of the competencies to be assessed can be determined by asking each candidate to describe (in addition to actions and achievements) settings and occasions in which the competency has been demonstrated. This approach shares with the KSA-based methods the assumption that the competencies assessed can be manifest in a variety of settings and in the performance of a variety of tasks. It is also similar to the task-based methods, however, in its requirement that information about task performance be provided by applicants.

Thus, information to be scored can be placed on a continuum of sorts ranging from demonstration of behavioral skills or abilities to documentation of exposures. At one extreme are assessment methods in which candidates are required to do something (paper and pencil tests, performance tests, certain types of interview, etc.), and scores can be obtained by more or less objective evaluations of how well the tasks presented are performed. The behavioral consistency methods are similar, except that candidates are asked to describe relevant tasks they have performed in the past and to provide information about how well the tasks were performed. This information is then scored,

again more or less objectively. Further along the continuum are the task-based methods, which present relevant tasks and require candidates to rate how well they can perform them; and then the KSA-based methods, in which candidates are asked to assess not how well they have performed particular tasks, but the extent of their competencies presumed to underlie performance of relevant tasks. Finally, at the opposite end of the continuum, are the credential-based methods, which require from candidates, and provide for scoring, data which often contain little information about how well the individual has performed tasks in the past or has developed task-relevant competencies.

Recommendations for the Use of T&E Evaluations

In Establishing Minimum Qualification Requirements

The alternative methods and research evidence described in this paper may therefore provide excellent options to consider as alternatives to traditional minimum qualification requirements. Earlier, three practices common among employers were discussed: requirement of a minimum amount and kind of training and experience, preference for a particular background, and expressions of the competencies sought by the employer. Both the research evidence and the experience of many employers strongly support reductions or even elimination of credential requirements in the employment process for many jobs. For recruitment purposes, conveying to prospective candidates explicit information about the attributes sought, whether cast as experience in the performance of particular tasks, or as knowledge, skills, or other attributes sought, may well be more effective and efficient than presenting descriptions of credentials that must be possessed before a candidate will be considered further in the employment process. Such approaches may serve both to provide access to candidates who would otherwise not be considered, even though fully competent to perform the job, and increase the numbers of those who, though properly "credentialed," may choose not to apply because they correctly conclude that they lack the competencies or personal characteristics needed for successful performance in the target job. Moreover, while legitimate concerns have been expressed concerning the honesty and accuracy of candidate self-assessment (e.g., Anderson et al., 1984), use of these methods as a preliminary screening mechanism is likely to increase the numbers of competent candidates screened into the employment process. Whether these methods lessen or increase the number of incompetent candidates screened in at an early stage may depend largely on factors in the external environment, such as the unemployment rate, which is likely to affect the number and type of alternative employment sources. Basing final hiring decisions on other means of assessment can minimize or eliminate risks that may exist in sole use of self-assessment approaches.

In Rank-Ordering Candidates for Employment Consideration

In addition to serving as the basis for minimum qualification requirements, T&E evaluation methods also serve either alone or in combination with other selection techniques as mechanisms for establishing the order in which applicants will be employed or considered for employment in many organizations. The validity evidence reviewed here suggests that the use of the traditional point method for such purposes should be limited to situations in which the mean job-related experience level of the applicant population is very low (i.e., two to three years). Given the measurement problems with this method discussed previously, justification for its use on the basis of content validity alone is difficult. It is suggested that organizations move in the direction of using T&E evaluation methods that focus directly on applicant past job-related achievements, accomplishments, and behaviors. Methods which do this are now available (e.g., behavioral consistency, task-based, and KSA-based self-assessment methods), are often justifiable on the basis of content validity, and show promise for moderate criterion-related validity in studies compiled to date. Until more research evidence has been accumulated, improved point and grouping methods should be viewed with healthy skepticism and interim alternatives to the holistic and traditional point methods as organizations work toward the transition to achievement- and behaviorally-oriented measures.

In practice, selection of assessment methods and strategies for their use should depend on the nature of the attributes to be assessed, as indicated by results from a job analysis, and other pertinent facts about the employment setting, such as base rates, selection ratios, dollar utility estimates, affirmative action concerns, and so forth. If methods based on the approaches described here are carefully developed to insure clear and explicit instructions to applicants and to devise appropriate scoring schemes, they may well permit assessment of many attributes that are currently not assessed by private or public employers. To maximize the predictive value of an assessment battery, it is necessary both to assess all relevant KSAs among applicant, and to do so reliably. Assessment methods vary, however, both in bandwidth and fidelity (Cronbach & Gleser, 1965). It is increasingly evident that T&E evaluation methods, despite their limitations, can contribute through the assessment of attributes not easily tapped by other methods. In effect, their greater "bandwidth" may outweigh their lesser "fidelity" in assessing many important attributes of job applicants. When combined with other methods, such as objective knowledge and ability tests, well-designed semi-structured oral interviews, and performance tests, it is reasonable to anticipate that greater overall predictive accuracy may be achieved.

Private sector employers are urged to abandon the unsystematic, unstructured, and *largely unstudyable* holistic judgment method, particularly in the initial screening that takes place in personnel departments. Replace this method with any of the formal methods of T&E evaluation described in this paper, and begin conducting (and reporting the results from) criterion-related validity studies.

Recommendations for Future Research on T&E Evaluation

The primary recommendation for future research on T&E evaluation is a call for more primary criterion-related validity studies on every formal method described in this paper. Given the paucity of criterion-related validity studies on these various approaches, about the only thing one can assert with relative certainty is that the traditional point method will generally *not* yield valid predictions of job performance except in instances where the relative experience level of the applicant population is quite low. The lack of criterion-related validity studies on T&E evaluation methods is inconsistent with the pervasive use of these methods in personnel selection, particularly in the public sector.

Few organizations use multiple methods of T&E evaluation. Yet, in order to *understand* what is important in the training and experience backgrounds of individuals for accurate prediction of performance in various occupational areas, data on the interrelationships among T&E evaluation methods of varying content, in addition to validity data, are needed. Systematic knowledge of such interrelationships should be a substantial boon to the growing field of career development and career planning. A large-scale, long-term, multi-T&E evaluation method, multi-occupation, multi-criterion predictive validity study on the order of the AT&T assessment center validity study (Bray & Grant, 1966) would be most informative.

For *complete* understanding and optimal use of T&E evaluation in *conjunction* with other components of personnel assessment systems, data on the relationships among various T&E evaluation method scores and scores on other components of selection procedures are required. For example, Ash (1984) reported correlations of .08 and .11 between *Wonderlic Personnel Test* scores and holistic judgment and behavioral consistency scores, respectively. The question as to whether or not various methods of T&E evaluation make any unique contribution to the prediction of criteria beyond the variance accounted for by other personnel assessment variables needs to be addressed empirically.

SUMMARY AND CONCLUSIONS

The purpose of this paper has been to acquaint the reader with the state of the art of a set of methods and procedures used extensively in personnel

selection—the assessment of job applicant training and work experience. The general practice of T&E evaluation as implemented in many American organizations appears to be sorely wanting. The two predominant methods which serve both as the basis for establishing minimum qualifications and for ranking applicants in terms of suitability for employment—*holistic judgment* and *traditional point* methods—seem to result in arbitrary decisions which generally lack validity. Potentially promising T&E evaluation methods exist. These include the *behavioral consistency* method, both *KSA-based* and *task-based* self-assessment methods, the *improved point* method, and the *grouping* method. In order for these methods to be more fully evaluated, organizations need to adopt them in place of holistic judgment and traditional point methods. While most of the more promising methods cost more to develop and implement than existing procedures due to increased job analysis data requirements, organizations are likely to gain through lower legal liability due to better documentation for content validity, and through higher levels of job performance, on average, from work groups selected by means of the more valid T&E evaluation methods.

It must be emphasized that these “potentially promising” T&E evaluation methods are exactly that—*potentially promising*. The potential promise needs more thorough evaluation. The set of potentially promising methods may be reduced, or it may be found that different T&E evaluation methods work better than others under certain circumstances. The volume of validity research for these T&E evaluation methods is simply insufficient at this point in time to permit one to draw firm conclusions about their merits relative to each other and to other personnel selection techniques. More criterion-related validity research is needed.

NOTES

1. The authors are grateful to Nancy Abrams for her contribution to the description and evaluation of the self-assessment T&E evaluation methods.
2. The authors are grateful to Charles B. Schultz for supplying information and examples related to the improved point method of T&E evaluation.

REFERENCES

- Adams, J.A. (1987). Historical review and appraisal on research on learning, retention and transfer of human motor skills. *Psychological Bulletin*, 101, 41-74.
- Anderson, C.D., Warner, J.L., & Spencer, C.C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, 69, 574-580.
- Arvey, R.D., McGowen, S., & Horgan, D. (May, 1981). *The use of experience requirements in selecting employees*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Denver.

- 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, R.A. (1983). The behavioral consistency method of training and experience evaluation: Content validity issues and completion rate problems. *Public Personnel Management*, 12, 115-127.
- Ash, R.A. (1984). A comparative study of the behavioral consistency and holistic judgment methods of job applicant training and work experience evaluation. *Public Personnel Management*, 13, 157-172.
- Ash, R.A. (1986). The activity/achievement indicator: A possible alternative to the behavioral consistency method of training and experience evaluation. *Public Personnel Management*, 15, 325-343.
- Ash, R.A., & Levine, E.L. (December, 1981). *An investigation of career service unassembled examinations*. Final report for STAR Grant No. 80-011. Tampa: Center for Evaluation Research, University of South Florida.
- Ash, R.A., & Levine, E.L. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology*, 70, 572-576.
- Asher, J.J. (1972). The biographical item: Can it be improved? *Personnel Psychology*, 25, 251-269.
- Beardsley, V.A. (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.
- Bray, D.W., & Grant, D.L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs*, 80, Whole No. 625.
- Callender, J.C., & Osburn, H.G. (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.
- Campbell, J.P., Dunnette, M.D., Arvey, R.D., & Hellervik, L.V. (1973). The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, 57, 15-22.
- Cascio, W.F. (1987). *Applied psychology in personnel management*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.
- Cascio, W.F., Alexander, R.A., & Barnett, G.V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1-24.
- Cattell, R.B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Chute, M.G. (1949). *Shakespeare of London*. New York: Dutton.
- Cobb, C.L., Spool, M.D., & Pollock, A.L. (1974). *A flexible and comprehensive system for rating training and experience*. Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, GA.
- Cook, C.L. (July, 1980). *Rating education, training, and experience in the public sector*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Boston.
- Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.
- Dichter, M.S. (Editor-in-Chief). (1985). *1983-84 Cumulative Supplement to Schlei & Grossman's Employment Discrimination Law*. Washington, DC: Bureau of National Affairs.
- Dipboye, R.L., Fontenelle, G.A., & Garner, K. (1984). Effects of previewing the application on interview process and outcomes. *Journal of Applied Psychology*, 69, 118-128.
- Dipboye, R.L., Fromkin, H.L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness and scholastic standing in evaluation of job applicant resumes. *Journal of Applied Psychology*, 60, 39-43.

- Dreher, G.F., & Sackett, P.R. (1983). *Perspectives on employee staffing and selection*. Homewood, IL: Irwin.
- England, G.W. (1971). *Development and use of weighted application blanks* (rev. ed.). Minneapolis: University of Minnesota Industrial Relations Center.
- Farrell, B.M. (December, 1979). *Task performance self-evaluations: An alternative selection procedure to traditional experience and training ratings*. Minneapolis, MN: State of Minnesota, Department of Personnel, Selection Research Unit.
- Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Gibson, J.W., & Prien, E.P. (1977). Validation of minimum qualifications. *Public Personnel Management*, 6, 447-451, 456.
- Gottfredson, L.S. (1986). Societal consequences of the g factor in employment. *Journal of Vocational Behavior*, 29, 379-410.
- Guion, R.M. (1974). Open a new window: Validities and values in psychological measurement. *American Psychologist*, 29, 287-296.
- Gutenberg, R.L., Arvey, R.D., Osburn, H.G., & Jeanneret, P.R. (1983). Moderating effects of decision-making/information-processing job dimensions on test validities. *Journal of Applied Psychology*, 68, 602-608.
- Hough, L.M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology*, 69, 135-146.
- Hunter, J.E. (1980). *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery [GATB]*. Division of Counseling and Test Development, Employment and Training Administration, U.S. Department of Labor, Washington, DC.
- Hunter, J.E., & Hunter, R.F. (1984). The validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage Publications.
- Johnson, J.C. (1980). *Expert testimony of James C. Johnson in Wallace Duncan v. State of Tennessee et al.* Nashville, TN: Technical Report of the Research Division, Department of Personnel, State of Tennessee.
- Johnson, J.C., Guffey, W.L., & Perry, R.A. (July, 1980). *When is a T&E rating valid?* Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Boston.
- Johnson, J.C., & Waldron, W.J. (in preparation). *The use of education and experience in job evaluation and as minimum qualifications for employment*. Nashville, TN: State of Tennessee, Department of Personnel, Research Division.
- Langdale, J.A., & Weitz, J. (1973). Estimating the influence of job information on interviewer agreement. *Journal of Applied Psychology*, 57, 23-27.
- Levine, E.L. (1983). *Everything you always wanted to know about job analysis*. Tampa, FL: Mariner.
- Levine, E.L., & Flory, A., III (1975). Evaluation of job applications: A conceptual framework. *Public Personnel Management*, 4, 378-385.
- MacLane, C.N. (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).
- Maslow, A.P. (1968). The unassembled examination. In J.J. Donovan (Ed.), *Recruitment and selection in the public service*. Chicago: Public Personnel Association.
- McDaniel, M.A. (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, M.A., Schmidt, F.L., & Hunter, J.E. (1988a). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology*, *41*, 283-314.
- McDaniel, M.A., Schmidt, F.L., & Hunter, J.E. (1988b). Job experience correlates of job performance. *Journal of Applied Psychology*, *73*, 327-330.
- Molyneaux, J.W. (1953). *An evaluation of unassembled examinations*. Unpublished master's thesis, The George Washington University, Washington, DC.
- Owens, W.A., & Schoenfeldt, L.F. (1979). Toward a classification of persons. *Journal of Applied Psychology*, *65*, 569-607.
- Porter, W.R., Levine, E.L., & Flory, A. III (1976). *Training and experience evaluation: A practical handbook for evaluating job applications, resumes, and other applicant data*. Tempe, AZ: Personnel Services Organization.
- Prien, E.P., & Hughes, G.L. (under review). *A content oriented approach to setting minimum qualifications*.
- Primoff, E.S. (1975). *How to prepare and conduct job element examinations (TS-75-1)*. Washington, DC: U.S. Government Printing Office.
- Sage, D.R., 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.
- Saso, C.D., & Tanis, E.P. (1974). *Selection and certification of eligibles: A survey of policies and practices*. Washington, DC: International Personnel Management Association (Personnel Report No. 743).
- Schmidt, F.L., Caplan, J.R., Bemis, S.E., 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.
- Schmidt, F.L., Hunter, J.E., & Outerbridge, A.N. (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, F.L., Hunter, J.E., Pearlman, K., & Hirsch, H.R. (1985). Questions and answers about validity generalization and meta-analysis. *Personnel Psychology*, *38*, 697-798.
- Schneider, B., & Schmitt, N. (1986). *Staffing organizations* (2nd ed.). Glenview, IL: Scott, Foresman and Company.
- Scollay, R.W. (1956). Validation of personal history items against a salary increase criterion. *Personnel Psychology*, *9*, 325-336.
- Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, *47*, 149-155.
- State of Connecticut, Department of Administrative Service, Personnel Division, Psychometric Unit. (1978). *Development and administration of experience/training evaluations [factor method]*. Hartford, CT: Author.
- State of Wisconsin, Department of Employment Relations. (April, 1979). *Sample of an actual achievement history questionnaire and cover letter*. Madison, WI: Author.
- Swander, C., & Schultz, C. (undated). *Selection for small job classes including practical procedures for E&T*. Olympia, WA: Washington State Department of Personnel.
- Tucker, D.H., & Rowe, P.M. (1979). Relationship between expectancy, causal attributions, and final hiring decisions in the employment interview. *Journal of Applied Psychology*, *64*, 27-34.
- U.S. Department of Labor (1972). *Handbook of analyzing jobs*. Washington, DC: U.S. Government Printing Office.