
*Personnel Assessment
Monographs*

■

The Rating of Experience and Training:

A Review of the Literature and
Recommendations on the Use of
Alternative E & T Procedures

■

Robert T. Schneider

*Volume 3, Number 1
July, 1994*

IPMAAC

*A publication of the Assessment Council of the
International Personnel Management Association*

*Publication of this Monograph
was supported in part by
generous contributions from*
Personnel Decisions, Inc.
and
Bruce Davey & Associates

*Personnel Assessment
Monographs*

The Rating of Experience and Training:

A Review of the Literature and Recommendations
on the
Use of Alternative E & T Procedures

Robert E. Schneider

ipmaac

Copyright © July, 1994

International Personnel Management Association
Assessment Council
1617 Duke Street, Alexandria, VA 22314
(703) 549-7100

**The Rating of Experience and Training:
A Review of the Literature and Recommendations
on the Use of Alternative E & T Procedures**

Robert E. Schneider
Research Division,
Pennsylvania State Civil Service Commission

Personnel Assessment Monographs is published by the Assessment Council of the International Personnel Management Association (IPMAAC), 1617 Duke Street, Alexandria, Virginia 22314. A copy is provided without charge to members of the Assessment Council. Others may obtain copies by writing the Director of Assessment Services at the above address. The monographs are intended to provide the most current information available on topics of broad interest in the field of applied personnel assessment. Manuscripts and proposals are solicited by the editor. Most appropriate are reviews of research on assessment methods, and innovative applications and solutions to personnel assessment problems in employee selection, performance, appraisal, program evaluation, management analysis and related areas which are not generally available elsewhere. Manuscripts are reviewed by the editor and by consulting editors selected for their expertise in the area covered by a particular paper. They are selected for publication based on the technical and professional merit of the work, its applicability to current issues, and its utility and interest to the targeted readers. Inquiries, proposals, and manuscripts may be sent to the editor, Harold P. Brull, Personnel Decisions, Inc., 2000 Plaza VII Tower, 45 South Seventh Street, Minneapolis, Minnesota 55402.

Harold P. Brull, Editor

Editor's Preface

The Rating of Experience and Training: A Review of the Literature and Recommendations on the Use of Alternative E & T Procedures began as a project for the Research Division of the Pennsylvania State Civil Service Commission. The monograph should provide an excellent foundation for the conceptualization, design, and administration of experience and training ratings.

The editor wishes to thank Bob Schneider for his hard work and graciousness in accepting a barrage of editorial comments. Thanks also to members of the IMPAAC monograph committee for their helpful insights and Norma Williams, of my organization, for her thorough, professional copy editing.

Robert E. Schneider

Robert E. Schneider is a senior personnel analyst for the Research Division of the Pennsylvania State Civil Service Commission (PA SCSC) in Harrisburg, PA. While with the PA SCSC, he has developed numerous examination programs and participated in a variety of applied research projects, including development of a new classification plan and career advancement system for engineering and architectural job titles, development and validation of a transportable video-based assessment battery for Psychiatric/Mental Retardation Aides, development and implementation of a computerized adverse impact analysis system, and development and implementation of a program for testing accommodation for disabled persons. He is a coauthor of the IPMAAC Monograph on Testing Accommodation.

He received his BS from Temple University in Philadelphia and his MA in Governmental Administration from the Wharton School of the University of Pennsylvania. He is a member of the International Personnel Management Association Assessment Council (IPMAAC) and currently serves as Training Committee Chair for the Mid-Atlantic Personnel Assessment Consortium (MAPAC).

Acknowledgments

I would like to thank the many people who helped with the preparation of this monograph, while acknowledging that I am solely responsible for all errors of omission or commission. I am particularly indebted to Charles Sproule for his support, encouragement and guidance as this monograph was written. My thanks also go to Steven Berkley, Harold Brull, Linda Coover, Donna Denning, Daniel Masden, Brian Moritsch, Leroy Sheibley, Christina Valdez and Pasquale Zarro for their helpful recommendations. Finally, I want to convey my sincere appreciation to Drs. Nancy Abrams, Ronald Ash, James Johnson and Michael McDaniel for their perceptive reviews of the initial draft.

R.E.S.

December, 1993

Personnel Assessment Monograph

The Rating of Experience and Training: A Review of the Literature and Recommendations on the Use of Alternative E & T Procedures

CONTENTS	PAGE
Abstract	1
Preface	2
Introduction	4
Standards for Development and Use of E&T Examinations	5
E&T Rating Methods	8
Point Method	9
Grouping Method	12
Knowledge, Skill and Ability (KSA)-Based Method	15
Task-Based Method	17
Behavioral Consistency Method	19
Holistic Rating Method	22
Summary of Findings and Conclusions	22
Use of Specific Rating Elements	26
Educational Measures	26
Academic Achievement	26
Level of Education	29
Field of Study	30
Institutional Quality	31
Extra-curricular Activities	31
Work Experience	34
Additional Issues and Research Findings	37
Self-Reports/Self-Ratings	37
Validity of Self-Ratings	38
Reliability of Self-Ratings	39
Score Inflation and Faking	39
Automated E&T Testing	42
Scannable Forms	43
Direct Data Entry	46
References	49
Appendix A--Suggested Reading	55

Abstract

Ratings of Experience and Training (E&T) are frequently used as screening devices in the selection of employees. This monograph provides a taxonomy for the classification of E&T ratings by method and reviews recent literature on the validity and reliability of six different E&T rating methods, as well as commonly used specific measures of education and experience. The feasibility of automating E&T ratings is discussed, and information is provided on the use of scannable forms and various methods of direct data entry via personal computers.

The literature provides very little research support for E&Ts that use traditional Point Methods of rating or for the use of High School Class Rank as a predictor of future job performance. While there is limited empirical evidence supporting the use of E&Ts patterned after the KSA-based Method and Task-based Method, the empirical evidence supporting the Behavioral Consistency Method is more substantial. The literature supports the use of academic achievement, field of study, and, to a limited extent, level of education as predictors of job success. Too little information was available to draw conclusions regarding the validity and reliability of the Grouping Method and Holistic Method of rating, and no evidence was provided on the value of Participation in Extra-curricular Activities as a predictor of job success.

The support for the use of self-ratings by candidates is generally positive but inconclusive. Further research is needed to develop reliable and accurate self-assessment methods. Such methods can be used in conjunction with personal computers and scanning technology to produce highly efficient, automated assessment instruments.

Preface

This monograph is intended to serve as a resource guide for testing specialists and other practitioners who routinely develop and use Experience & Training (E&T) selection instruments. Its primary objective is to summarize what has been reported in the professional literature about E&T examining that has practical value for testing specialists. It explores both traditional and non-traditional methods of E&T examination. It will also attempt to address such questions as what types of E&T instruments are most likely to yield valid prediction of success in training and job performance and under what circumstances E&Ts should be considered for use as assessment instruments.

Where appropriate, the assertions made in this monograph are supported by citations from published and unpublished literature, as well as a number of personal communications between the author and those whose work forms the central core of this monograph. These citations are included to provide the reader with sources of more detailed information about the topics covered. However, this monograph should not be considered prescriptive. A number of questions about E&T examination strategies and techniques remain unanswered. Many of the studies cited in this monograph utilized meta-analytic techniques. While meta-analysis is an extremely useful and powerful statistical technique, a number of practitioners believe that "what you get is what you get" in a validation study, and the application of statistical corrections will not change those results. Others have suggested that the application of too many corrections may present a distorted picture of the results. For these reasons, the operational significance of the findings of the studies presented here is open to interpretation.

This monograph does not address Biodata assessments and Minimum Qualification screens. Admittedly, there is considerable overlap between these selection methods, and much of the same information used for E&T ratings is also used for assessments using Biodata and Minimum Qualifications. However, despite their congruence, these methods also have their differences. Minimum Qualifications are used as simple "pass/fail" screens (more than that

effectively results in an E&T rating), while Biodata instruments are linked to a variety of criteria through empirical methods.

Introduction

Ratings of Experience and Training (E&T) have been defined as "a method used to rate or select applicants for employment . . . in which information provided by applicants on application forms, questionnaires, resumes, and/or other documents is evaluated in accordance with a specific plan" (Porter, Levine & Flory, 1976). It is one of the most widely used methods of personnel assessment. The information used for E&T ratings includes previous work and volunteer experience, education or other training, and credentials. E&T rating plans differ from Biodata assessment methods (which are *not* covered by this monograph), in that the factors rated through E&T examinations and the internal weighting of those factors are derived through job analysis and subjective judgment rather than empirical evidence (Ash, 1983). In addition, Biodata methods frequently contain questions related to "interest" and "attitude," which are not typically measured in E&T instruments.

Several fundamental assumptions underpin the use of E&T examination techniques. The first of these is competency based: What a person has done in the past is a valid predictor of what that person is likely to do in the future (Arvey, McGowan & Horgan, 1981). The second is based on motivation: As individuals gain experience in an occupation, they demonstrate greater commitment to it, and are thus "more likely to wish to pursue, perform well in and gain advancement in the occupation . . ." (Porter, et al., 1976).

Many public jurisdictions rely heavily on E&T examination procedures. For example, during the last five years, more than half of all examination programs developed by the Pennsylvania State Civil Service Commission utilized some form of E&T rating technique. E&Ts are used for a broad spectrum of jobs ranging from entry-level through senior-level full performance job classifications. They are used for the assessment of rudimentary skills needed for very low-level jobs as well as complex dimensions needed to perform high-level professional work.

Standards for Development and Use of E&T Examinations

E&T procedures are examinations. As such, the legal and professional guidelines for test construction and use that apply to other selection instruments also apply to Ratings of Experience and Training. These include the American Psychological Association Standards and the *Uniform Guidelines On Employee Selection Procedures* (Guidelines). Section 14-C-6 of the Guidelines states:

A requirement for or evaluation of specific prior training or experience based on content validity, including a specification of level or amount of training or experience, should be justified on the basis of the relationship between the content of the training or experience and the content of the job for which the training or experience is to be required or evaluated.

The Guidelines further state:

The critical consideration is the resemblance between the specific behaviors, products, knowledges, skills, or abilities in the experience or training and the specific behaviors, products, knowledges, skills or abilities required on the job...

In addition to the requirement that E&Ts meet legal and professional standards, other more specific criteria for E&T construction have been written. Among them are the following recommendations, extracted from the course on "Evaluation of Experience and Training," developed by the Mid-Atlantic Personnel Assessment Consortium:

A properly constructed E&T should:

1. be based on job analysis.
2. concentrate on evaluating the most important entry-level work behaviors or job requirements that differentiate superior workers from adequate workers.
3. give all candidates an equal opportunity to demonstrate possession of the job requirements. (This usually means the use of a supplemental application form to collect information

in a format which is standardized and reduces the need for interpretation in the rating process).

4. use a structured and well defined rating procedure.
5. be supported by reviewable documentation.
6. be reliable and valid.

Each of these principles warrant brief discussion.

Item 1 requires that an E&T must be supported by job analysis that provides enough information to demonstrate the relationship between the test and the job. Generally, this has been interpreted to mean a listing of job tasks and their associated KSAs, as well as a rating of task/KSA importance to the job. The analysis must also provide the information needed to determine the important entry-level work behaviors/job requirements, to develop the rating procedure and supplemental application forms (if used), and to provide content validity documentation.

Item 2 conforms to the intent of the Uniform Guidelines. It is also based on a fundamental assumption underlying the use of any E&T—that past performance is a valid predictor of likely future performance. The major considerations are that the test instrument should assess the most important dimensions of the job, cover a representative sampling of dimensions relevant to the work, and include only those dimensions needed on entry. If candidates are to be ranked, the test should also focus on dimensions that differentiate superior performers from those who are merely adequate.

Item 3 requires that every candidate have an opportunity to present all the relevant information they can regarding their qualifications for the job. This implies the need, in most instances, for supplemental application forms that are designed to be understandable to the applicant group, insuring that candidates know what information to provide, and that allow the crediting of experience and training obtained in non-traditional ways. The use of such supplemental application forms is also likely to increase reliability of assessment.

Item 4 addresses fairness and uniformity of treatment of all candidates. This standard requires that E&T rating scales be carefully crafted to include well defined rating benchmarks and that all those who use the rating scales fully understand the scales and their application. This practice also serves to increase the reliability of assessment.

Item 5 simply means that enough information is provided about what was done and why it was done to allow an objective observer to determine whether the examination was properly constructed and appropriately used.

Item 6 is a professional responsibility as well as a legal mandate, which applies to any selection instrument developed. In an assessment setting, the reliability of an E&T is typically determined by examining the level of agreement between two or more raters who independently evaluate and score the same E&T materials. For a variety of reasons, the validity of an E&T measure is often based on content validity evidence. While this is adequate, criterion-related and construct validity evidence provide additional support for a test instrument. Some researchers contend that the validity of an E&T measure may also be inferred from the comparison of the E&T assessment approach with cumulative data on the validity of similar E&T methods.¹

¹ Personal communication from Dr. Michael McDaniel to the author, dated April 26, 1993. Also refer to (Johnson, Guffey & Perry, 1980).

E&T Rating Methods

The wide variation in the layout and content of individual E&T ratings and the absence of a clearly defined and generally accepted model for classifying E&T examination types is a significant problem. There is, at best, limited agreement among researchers on this issue. What one researcher classifies one way, another may classify quite differently. However, the published literature describes six different methods of E&T examining which provide a useful framework for this discussion (Ash, 1984):

- Point Method
- Grouping Method
- Knowledge, Skill and Ability Method
- Task Method
- Behavioral Consistency Method
- Holistic Method

Each of these E&T methods is discussed below. Although researchers differed in the labels they used to describe the E&T methods they studied, most provided enough descriptive information about their work to determine where their studies fit within the model chosen for this review. The information presented includes a description of each method, a summary of available information on the reliability and validity of the method, and recommendations on the use of the method. Additional information regarding issues such as cost of examination development and administration and applicant acceptance of the method are included if they have been addressed in the literature. Information regarding E&T examination programs recently issued by the Pennsylvania State Civil Service Commission appears in Appendix B. A list of suggested readings on various E&T topics is included as Appendix C.

A. Point Method

Point Method ratings are based on the presumption that there is a positive linear relationship between the duration of exposure to different kinds of experience and training deemed relevant to a given job, and job performance. Point Methods of rating are highly credentialistic, focusing on the amount of time spent in a job title or performing a specific group of functions.

As defined in the E&T training course developed by the International Personnel Management Association Assessment Council (IPMAAC), “Traditional” Point Method E&Ts are supported by very limited job analysis data.² Classification specifications or job descriptions alone provide the information used for their construction. While the information used to construct classification specifications and job descriptions is legitimately considered to be “job analysis” data of a sort, their utility for developing rating schedules is limited, owing to their lack of detail regarding essential job content.

More typically, “Improved” Point Method examinations are used.³ These are Point Method E&Ts that begin with the use of a semi-structured interview type of job analysis to find out from Subject Matter Experts (SMEs) the important KSAs in the target job and the *specific* things that a person would have done in the past to indicate that the person possesses these important KSAs. The next step involves incorporation of the *specific* things—which, if done by the candidate in the past, indicate possession of particular KSAs of interest—into an E&T scoring key.

Either technique results in a scoring key that includes benchmarks describing the types of education and/or experience required for different levels of performance. Various kinds of experience and training are usually differentially weighted; the closer the apparent relationship of the experience/training to the work to be performed, the greater the credit given for possession of that type of experience/training. These rating benchmarks are used to evaluate the experience and training individual candidates present on their application forms and occasionally in application supplements (Lyons, 1992). Normally,

² Personal communication from Dr. Nancy E. Abrams to the author, dated April 21, 1993.

³ “Improved Point Method” is a term first noted in Ash (1981).

candidates are assigned a base (passing) score if they meet the minimum qualification requirements. Additional points are awarded to candidates based on the duration of their participation in each type of experience and training identified beyond the minimum needed for qualification. The greater the amount of relevant experience/training a candidate possesses, and the closer the apparent match between the candidate's experience/training and the job, the higher that candidate's score will be.

E&Ts using Point Methods generally have very high reliability. For example, Schmidt, Caplan, et al. (1979) reported the results of four earlier studies⁴ that yielded inter-rater reliabilities⁵ ranging from .76 to 1.00; and intra-rater reliabilities⁶ in the range of .81 to .83. These findings are not surprising. Point Methods of rating involve counting the number of months or years of relevant experience and the number of years, courses, degrees, etc. of relevant training possessed by a candidate. Therefore, subjective interpretation is minimal, and the results achieved are typically very consistent across any number of raters using the same measurement instrument. Unfortunately, there is a big difference between an instrument's reliability and its validity. Just because something can be measured with a high degree of accuracy time after time does not imply that the measurement will be useful in predicting something else.

⁴ Wright, R. & Ulm, R. (1974); Reliability Study of Training and Education Ratings of Social Worker II Applications. Athens, GA: Georgia State Merit System, Examination & Recruitment Division. Cobb C.L., Spool, M.D. & Pollock, A.L. (1974), *Flexible and Comprehensive System for Rating Experience & Training*. Paper presented at a meeting of the Southeastern Psychological Association, Atlanta, GA. Levine, E.L. & Flory, A. (1974), *Conceptual Framework for the Evaluation of Job Applications and a Descriptive Study of the Inter-rater Reliability of a Categorical Approach to Screening Applicants*. Phoenix, AZ: Arizona Department of Administration. Burke, L.K. & Taylor, E.K. (1950), *Rating Training and Experience*. Journal of Applied Psychology, Vol. 34, 381-383.

⁵ Comparing the results of ratings of the same data by different raters.

⁶ Comparing successive ratings by the same rater of the same data.

Point Method E&T rating schemes are frequently supported on the basis of a content validation strategy. However, additional justification for using any employment test is its criterion-related validity—the extent to which scores on a test correlate with performance on the job. Content validation strategies should be supported by evidence that the E&T rating scheme produced an examination that possesses predictive (criterion-related) or construct validity. However, there are any number of reasons why it may not be possible to determine the criterion-related validity of an individual examination. Small candidate pools, limited hiring, few incumbents, test score range restriction among those hired, and a host of other legitimate constraints often preclude criterion-related validation studies.

The literature contains a number of empirical validation studies relating to various E&T examination methods. One such study (Schmidt, Caplan, et al., 1979) cited an earlier unpublished work that examined the validity of Point Method E&T examinations used for 52 different job titles in 18 classification series used by the U.S. Veterans Administration.⁷ Scores on the examinations were compared with supervisory ratings to obtain validity coefficients for each of the 52 job titles. The average of those validity coefficients was .10 (uncorrected). Schmidt, et al. (1979) proposed that factors such as range restriction artificially suppress validity coefficients. This view has some justification. As a rule, many more persons participate in examination programs than are hired. Because only the top scoring candidates are employed, differences in examination scores among candidates who are hired may be quite small. Hunter & Hunter (1984) recalculated the data presented in Schmidt, et al. (1979), using meta-analytic methods designed to correct for range restriction and other artifacts. They reported a corrected validity coefficient of .13.

Two additional studies were found which used meta-analytic methods and corrections for artifacts. One of these studies (McDaniel & Schmidt, 1985), analyzed 81 studies involving 6,468 subjects. They calculated a corrected validity coefficient of .15 for Point Method-based E&Ts. The second (McDaniel, Schmidt & Hunter, 1988a) yielded a corrected validity coefficient of .11.

⁷ Molyneaux, J.W. (1953), *An Evaluation of Unassembled Examinations*. Unpublished Master's Thesis, The George Washington University.

Although the numbers (81 studies and 6,362 subjects) are similar to those reported in McDaniel & Schmidt (1985), they result from a different group of studies. Clearly, a growing body of empirical evidence suggests that Point Method E&T examinations generally have very low predictive validity.

Typically, Point Method-based E&Ts are used to rank candidates.

Comments and Conclusions: There are still a number of unanswered questions regarding the criterion-related validity of Point Method-based E&T examinations, which can only be answered through further research. For example, different methods of crediting and scoring, such as those discussed in the section of this monograph on “Use of Specific Rating Elements,” may increase the criterion-related validity of Point Method E&Ts. Nonetheless, currently available empirical evidence suggests that Point Method-based E&T examinations generally have very low predictive validity and that there is very limited justification for their use. For these reasons, other assessment procedures should be considered first. However, Point Method E&Ts may be a necessary, if not desirable, approach in situations where the use of another testing strategy is not feasible. In those situations, the “Improved” Point Method would be the E&T Point Method best supported on a content validity basis.

B. Grouping Method

The Grouping Method is used to sort candidates into a small number of categories (e.g., well qualified, qualified, and not qualified), based on the simultaneous evaluation of all relevant training and experience. These global evaluations are obtained by evaluating all aspects of a candidate’s background in relation to predetermined objective benchmarks. McDaniel & Schmidt (1985) gave the following example of a rating benchmark used by Ash (1981) for Automotive Equipment Repair Foreman:

Three months of motorized vehicle repair schooling and five 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 motorized equipment (cars, trucks, gas and diesel). At least two years of the five must have been in a lead work or supervisory capacity. This experience must have occurred all within the last ten years. At some point in the work history there should be evidence of ordering parts and supplies, writing service orders, etc., to indicate the capacity to do paper work.

The Grouping Method of E&T examining and the technique of Score Grouping or Banding are sometimes confused. Therefore, the distinctions (and similarities) between them require some clarification. In Score Grouping or Banding, which can be applied to any examination method, candidates' raw scores are determined, but then the raw scores are grouped into a limited number of "bands," each of which represents a range of scores likely to represent approximately equal levels of proficiency. All candidates attaining raw scores that fall within a given band may be assigned the same Final Earned Rating or certified as a block for consideration for a job opening. Score Grouping or Banding can be used with any type of assessment instrument, not just E&T ratings.

To date, the Grouping E&T Method has received scant attention from researchers. There is very little published literature about this technique. One study (Levine & Flory, 1975) analyzed ratings by two raters for 15 jobs, each having an average candidate pool of 19. They reported uncorrected inter-rater reliabilities ranging from a low of .67 to a high of 1.00, with a median of .96. A second study (Ash, 1981)⁸ compared E&T rating scores obtained using Point and Grouping Methods for subjects in three job titles: Automotive Equipment Repair Foreman, Computer Operations Supervisor and Medical Disability Examiner Supervisor. Subsequent analysis of those data yielded uncorrected correlations between the scores achieved on the two methods of .60 (N=126), .71 (N=94) and .74 (N=63); and a mean corrected correlation coefficient between the two methods of .93. Inter-rater reliabilities of .48, .76, and .78 were reported for the Grouping Method E&Ts used.

Ash (1981) reported that the Point Method ratings yielded higher coefficients of variation than were produced through the Grouping Method. However, these differences appeared to be insignificant. In a

subsequent paper, Ash & Levine (1985) suggested that the greater score dispersions yielded by the Point Method as reported in Ash (1981) implied a greater precision in measurement than was warranted by the test instrument. Given the low correlation coefficients between predictor and job performance yielded in studies of Point Method E&Ts, this observation makes considerable sense.

Ash (1981) noted that some candidates may have trouble understanding the scoring mechanism used with the Grouping Method. To the extent this is so, reduced candidate acceptance of the scoring mechanism may be a problem.

By definition, Grouping Methods are *not* used to rank candidates.

Comments and Conclusions: There is no published research evidence on the criterion-related validity of the Grouping Method. Although the Point Method and the Grouping Method may seem to be quite dissimilar, both methods score candidates on the basis of credentialistic information. Therefore, the Grouping Method should be viewed as a Point Method variant. The cautions and limitations that apply to the Point Method should be considered to apply to the Grouping Method as well.

Under the Grouping Method, all candidates within each group are assigned the same score, on the presumption that they have approximately equal qualifications. Given the imprecision and low predictive power of credentialistic E&T rating methods as reported in the literature, this appears to be a prudent assumption. Grouping Methods appear to conform better than other Point Methods with the level of precision one can achieve using credentialistic E&T methods.

Because applicant scores are clustered in fairly broad blocks, the Grouping Method may serve as a means of attempting to minimize disparate effect.

⁸ Although not nearly as comprehensive as the dissertation cited here, a significant portion of the essence of the dissertation is reported in Ash, R.A. & Levine, E.L. (1985), *Job Applicant Training and Work Experience Evaluation: An Empirical Comparison of Four Methods*, Journal of Applied Psychology, Vol. 70, 572-576.

C. Knowledge, Skill and Ability (KSA)-Based Method

Most of the published analyses treat the KSA Method as a variant of the Point Method, primarily because scores are usually based on possession of credentials and exposure to certain kinds of work, rather than proficiency of performance. This is so because ratings are based on possession of credentials. However, it is treated here as a separate E&T method because of its focus on possession of the knowledges, skills and abilities that underlie job performance, regardless of source and irrespective of time spent in their acquisition.

Under the KSA-based Method, the knowledges, skills and abilities deemed most important on the job are linked to the test instrument through job analysis. Candidates are provided with a list of the KSAs considered essential on entry to the job, and then asked to rate the extent to which they possess each KSA and where they acquired the KSA. Candidates are credited with the performance of tasks, possession of diplomas, degrees, certificates and/or other academic awards; and completion of specific training courses (and sometimes the grades received in those courses) that indicate exposure to—and, by extension, possession of—the knowledges, skills and abilities that are required on the job.

One paper discussed the theoretical strength of the relationships among E&T instruments, KSAs, and job performance (Johnson, et al., 1980). This paper referenced earlier work done by staff of the U.S. Office of Personnel Management, which suggested that E&T ratings are unlikely to correlate better than .40 with KSAs, while KSAs are unlikely to yield correlations higher than .50 with actual job performance. Thus, the best possible correlation between test score and job performance that can be attained with E&T ratings based on KSAs will be on the order of $.40 \times .50$, or .20.

Two studies that addressed the criterion-related validity of KSA-based E&Ts were found⁹. The first of these (McDaniel, Schmidt &

⁹ A third study (McDaniel & Schmidt; 1985) was also found which reported corrected correlations of .20 and .23 for KSA-based E&Ts. However, these coefficients were derived from earlier studies involving the *Illinois Job Element Approach*. A primary feature of this approach is the use of multiple assessment instruments, only one of which is a KSA-based E&T rating. Although the focus of McDaniel & Schmidt is E&T assessment, it is unclear whether the coefficients reported were derived from the results of the KSA-based E&T instrument alone, or from the results of the entire selection battery.

Hunter, 1988a) used meta-analytic techniques, which yielded a mean corrected correlation of .20 (16 studies, 3,168 subjects). The second (Lyons, 1992) reported the results of a series of small scale studies, yielding validity coefficients of .05 (N=55) and .06 (N=73) for two entry-level jobs, and .21 (N=109 $p < .05$) and .41 (N=50 $p < .05$) for two full performance level jobs. However, it is once again worth noting that while KSA-based E&Ts yield low to moderate validity coefficients, even low positive correlations provide some predictive value. None of the studies reviewed contained useful estimates of reliability for KSA-based ratings.

The KSA-based Method may be used to rank or group candidates.

Comments and Conclusions: Based on the available evidence, KSA-based E&Ts appear to be somewhat better predictors of job performance than Point Method E&Ts. Because of the relative ease with which KSA-based E&Ts can be linked to the job through content validation, this E&T method may be considered for use in situations where an examination must be developed in a short time and possession of specific knowledges and skills is deemed important to successful performance. Because KSA-based E&Ts are relatively easy to administer and score, they are also useful for examination programs in which a substantial number of candidates are expected. Since KSAs can be acquired through formal and informal training and life experiences, as well as formal training and paid work experience, this method can be used for capturing and crediting a broad range of activities that may be important for many types of jobs. This characteristic makes KSA-based E&Ts useful as a possible means of reducing disparate effects, since credit can be given for non-traditional life experiences that may go unrecognized under other E&T types.

However, KSA-based E&Ts require a great deal of analytic ability on the part of candidates. Even when KSAs are described operationally, the candidate must comprehend the concept that underlies the description, then link that concept to a past activity or a possessed credential indicating possession of the KSA, and finally describe that experience or credential in a meaningful way on an application supplement. For this reason, KSA-based E&Ts are inappropriate as an assessment method for jobs that demand little or no analytic ability or require minimal reading and writing skills.

D. Task-Based Method

This method is based on job analysis, which is used to develop a list of tasks deemed critical to the job. Candidates are presented with the task list and a set of scales describing a range of exposure to each task, or competency in task performance. Candidates are asked to check the tasks they have performed and their level of exposure to each (e.g., "I have observed the task being done"; "I have helped others with this task"; "I have done the task myself"; "I have supervised others who did this task"). In some cases, candidates are asked to indicate their level of proficiency in performing a task (e.g., "Those who know my work would say that I perform this task [worse] [better] than anyone else"). Alternatively, candidates may be asked to indicate the level of supervision they received while performing a task (e.g., "I have performed this task with close supervision"; "I have performed this task independently"; "I have taught others to perform this task"), providing an indirect estimate of task proficiency.

McDaniel & Schmidt (1985) and van Rijn (1980) cite earlier studies conducted on Task-based E&T ratings, including (Blanchard, 1979), (Ash, 1980), (Lyons, 1980), and (Pennone, 1981). These studies compared Task-based E&T ratings with job performance and yielded correlations ranging from a low of .16 to a high of .42. The corrected correlation across these studies is .28 (6 correlations, 686 subjects). Two additional meta-analytic studies explored the validity of Task-based E&Ts. The first of these (McDaniel & Schmidt, 1985) found a corrected validity coefficient of .159 (6 studies, 641 cases).¹⁰ McDaniel, Schmidt & Hunter (1988a) and Ash, Johnson, Levine & McDaniel (1989) reported an analysis of ten studies involving 991 subjects in which Task-based E&T ratings were compared with supervisory ratings of job performance. They reported a corrected correlation of .15.

Reliability data are sparse. One paper (Hough, Keyes & Dunnette, 1983) cited an earlier work (Ash, 1981), which reported inter-rater reliabilities of .96 (N=111), .99 (N=93), and 1.00 (N=60); and reported

¹⁰ The original study reported a correlation of .32 (9 studies, 905 cases). However, three of the studies included in this meta-analysis used peer ratings of *expected* future job performance as the criterion measure. In essence, a predictor was correlated with another predictor. A recalculation of the 1985 data, dropping these three studies, resulted in the correlation coefficient cited above.

a median inter-rater reliability of .78 (N=329) for a study of attorney positions in a federal agency.

Questions have been raised by some researchers regarding the issues of accuracy and score inflation by candidates. One study examined the effect different kinds of rating scales may have on the accuracy and inflation of candidate responses (Mabe & West, 1982). They argue that the rating scales used to describe various competency levels are frequently emotionally loaded. That is, they imply value judgments about the individual that extend beyond the task in question (i.e. "I perform this task worse than most other people"). However, the findings of one study suggest that the kinds of rating scales used have negligible effects on response inflation or the accuracy of the responses received (Lin, Dobbins, et al., 1992). Some researchers have explored the effectiveness of "lying scales" in controlling for intentional score inflation. This issue and the findings of relevant studies are discussed in the section of this monograph on "Additional Issues and Research Findings."

The Task-based Method may be used to either rank or group candidates.

Comments and Conclusion: The empirical evidence suggests that Task-based E&Ts yield low to moderate prediction and appear to have high reliability.

If adequate job analysis information is available, the linkage between rated tasks and job content is relatively easy to establish. However, Task-based E&T instruments may take more time to develop than some other E&T methods, because detailed job analysis data is needed to construct adequate task lists and care must be taken to write task descriptions and benchmarks that can be understood by candidates.

Well written task descriptions are important because Task-based E&Ts can require a great deal of analytic ability on the part of candidates. If the task descriptions are overly complex or ambiguous, candidates may have difficulty comprehending them or relating them to a past activity or to a possessed credential indicating task mastery. This problem can be moderated to a degree by careful writing and designing the Task-based E&T in check-off format, which facilitates candidate

understanding and response. It is also worth noting that the rating scales used with Task-based E&Ts lend themselves to template scoring. This feature could result in significantly lower administrative costs than can be achieved with other rating methods.

If properly constructed, Task-based E&Ts are extremely useful for jobs that attract candidates with limited reading/writing skills. However, they may not be appropriate for entry-level jobs because most candidates will not yet have had the opportunity to perform many of the tasks on which the ratings are based.

Because candidates typically self-report the information used for their ratings, candidate acceptance of scores is fairly high, although, as will be discussed later in this monograph, inflation of candidate responses may be a problem. In many situations, ease in developing linkages and high candidate acceptance are very real advantages. However, more effort is required to develop Task-based E&T rating instruments than E&Ts using either Point or Grouping Methods, and the results achieved may be only incrementally better.

E. Behavioral Consistency Method

The Behavioral Consistency Method is unique among the methods for E&T examining. It is the only one that focuses exclusively on assessing the demonstrated quality of a candidate's previous training and experience. A supplemental application form is used to collect information from candidates regarding their most significant accomplishments relative to a list of critical job behaviors. Critical behaviors are defined by McDaniel & Schmidt (1985) as those for which "differences in level of performance clearly differentiate between superior performers and all others."

For most jobs, there are from five to ten such behaviors. Thorough job analysis is an essential component of this strategy. Job Analysis is used to identify the critical job behaviors, as well as establish the range of performance for each behavior to be defined by the benchmarks. It is vitally important that the behaviors are carefully defined. Broderick (1992) cautioned against including too many behaviors, holding that "few jobs have more than ten behaviors that truly differentiate between superior performers and others." He also stressed the need to guard

against the use of broad, generalized descriptions of work, which may sound impressive but provide nothing measurable. Critical behaviors must have specific content.

Behavioral Consistency E&Ts take more time to develop than most other E&T methods, and they are comparatively expensive to administer. This is because they typically employ trained Subject Matter Experts for rating and require, on average, twice as much time to score as other E&T methods (Broderick, 1992). Ash (1984) reported average scoring times ranging from 2.5 to 5.0 minutes per application for a multi-dimensional Behavioral Consistency Rating. Schmidt, Caplan, et al. (1979) reported an average scoring time of 12 minutes per applicant for a Behavioral Consistency Rating.

Usually, Subject Matter Experts review candidate responses for each behavior and compare them to a set of anchored rating scales, which incorporate three to five benchmarks. Each benchmark denotes a different level of performance for a single behavior. The closer the reported performance to that identified with superior performers, the greater the credit awarded. When candidate populations are large, an alternative procedure can be used. Subject Matter Experts can be used to rate a sample of the larger group to develop benchmark achievement rating scales, (similar to behaviorally anchored rating scales). The benchmarks may then be used by evaluators or even the candidates themselves to score job-related achievements (Ash, 1984) and (Hough, 1984).

Based on the literature, E&T ratings using the Behavioral Consistency Method yield relatively high validity coefficients. McDaniel & Schmidt (1985) used meta-analytic techniques to correlate scores achieved on Behavioral Consistency ratings with supervisory ratings of job performance. They reported an uncorrected validity coefficient of .24 and a corrected validity coefficient of .49 (10 studies, 614 subjects). McDaniel, Schmidt & Hunter (1988a), using essentially the same methods, reported a corrected coefficient of .45 (15 studies, 1,148 subjects).

Behavioral Consistency-based E&Ts appear to have acceptable reliability. One study reported rate-rater reliability at .94 and inter-rater reliability of .84 (Ash, 1984). Another study reported inter-rater

reliability of .79 (Broderick, 1992). Behavioral Consistency E&Ts appear to correlate poorly or not at all with other E&T methods (Broderick, 1992). The implication of this finding is that the dimensions measured by Behavioral Consistency E&Ts differ from those measured through other commonly used E&T methods. However, as will be discussed below, one researcher found that Behavioral Consistency Ratings correlated .36 with ratings generated through the Holistic Judgment Method (Ash, 1984).

The Behavioral Consistency Method is most frequently used to rank candidates.

Comments and Conclusions: Behavioral Consistency ratings appear to yield relatively high correlations with job performance. However, Behavioral Consistency ratings have several inherent disadvantages compared to other E&T methods. They are time consuming and expensive to administer; they require candidates to possess and use considerable analytic ability as well as writing skill to identify relevant past accomplishments and to provide narrative descriptions of these accomplishments on which the ratings are based. Several researchers report that the return rates for Behavioral Consistency questionnaires are low. This phenomenon may result from the comparatively heavy burden Behavioral Consistency E&Ts place on candidates. The low return rates may indicate that completing a Behavioral Consistency questionnaire requires more effort than candidates are willing to expend or that it simply requires skills which are beyond the candidates ability.

Some of these problems may be partially overcome by modification, although the modifications present their own set of difficulties. Questionnaires could be structured to include descriptions of different performance levels under each of the Behaviors to be measured. Candidates could then indicate which of the descriptions most closely approximate their personal achievements. The drawbacks to this approach include the difficulty of developing benchmarks that are unambiguous enough to be correctly interpreted by candidates with any consistency, yet broad enough to capture the intended range of candidate achievements, and the difficulty of providing candidates with the guidance they need to accurately extrapolate and generalize their

personal achievements to the appropriate descriptions. (Candidates will not have access to the extensive training typically provided to SMEs before they are entrusted with rating.) If these concerns are not addressed effectively, the results of the test may be seriously compromised.

Behavioral Consistency ratings are an E&T method of choice for full performance level jobs that require significant reading/writing skills, and for which sufficient resources can be allocated to developing and rating the examination. They should not be used for jobs that require low reading, writing or analytic skill.

F. Holistic Rating Method

This method is used quite heavily in the private sector but rarely in public service. Some combination of hiring line managers and senior personnel staff evaluate information derived from candidates' applications, résumés, interviews, and so forth, in comparison to the perceived requirements of the job to be filled. Judgments derived under this method are generally unstructured. Only one study was found which attempted to address this method empirically (Ash, 1984). That study compared candidate ratings achieved holistically with scores obtained using a Behavioral Consistency E&T. Interestingly, Ash's study found a moderate rank order correlation of $+0.36_{avg.}$ between the two methods, rate-rater reliability of $+0.77_{avg.}$ and inter-rater reliability of $+0.82_{avg.}$ for the Holistic method.

Comments and Conclusions: The most desirable feature of the Holistic Rating Method is its flexibility. However, the difficulty of documenting content validity is likely to be monumental at best and impossible at worst. The method does not appear to lend itself to "merit selection" as defined by statute, regulation, or commonly accepted professional practice. In the absence of empirical evidence supporting the validity of Holistic Ratings, this method should not be used.

G. Summary of Findings and Conclusions

The literature provides no significant information on the validity of the Grouping Method or the Holistic Method. All four of the other

methods appear to offer at least limited predictive value in at least some circumstances. Point Methods offer very low prediction, the KSA-based and Task-based Methods provide slightly better prediction, and the Behavioral Consistency Method appears to have substantial value as a means of predicting job success.

The “Improved” Point Method may be useful as an assessment approach when time constraints or legitimate resource limitations restrict the range of available alternatives. Tests developed under this method can be defended on a content validity basis. The KSA-based Method and Grouping Method should be considered Point Method variants. Compared to Point Method E&Ts, KSA-based E&Ts are supported by more job-specific content validity evidence, while Grouping Method E&Ts may reflect the actual predictive power of Point Method instruments more pragmatically. All three of these E&T methods credit exposure rather than addressing competency.

The Task Method and the Behavioral Consistency Method both appear to be better choices than the others, but for different reasons. Unlike Point, Grouping or KSA-based E&Ts, Task-based E&T instruments can be structured to measure competencies (albeit very indirectly). Task-based E&Ts require less reading, writing, and analytic skills than Behavioral Consistency ratings.

As suggested above, Task-based rating schemes should be constructed to focus on competencies rather than simply exposure. Although no empirical evidence was found that specifically addressed this issue, a logical case can be made supporting the assessment of competencies. Consider, for example, a situation in which two new employees are hired at the same time to perform the same work. Over a three-year period, Employee A consistently performs at a very high level, producing high-quality work and introducing many innovative improvements. Meanwhile, Employee B performs at a level just barely high enough to avoid being fired. If only exposure is measured in a subsequent E&T assessment program, both will get the same score. Thus, an experience score based solely on exposure cannot be expected to have much validity.

Because of its heavy reliance on narrative responses from candidates, the Behavioral Consistency Method does not lend itself to automation.

For the same reason, it is recommended for use *only* when candidates must have advanced reading, writing, and analytic skills. The development and administration of Behavioral Consistency ratings is both time consuming and expensive, when compared to other E&T methods. However, this method consistently yields the highest correlations with job performance of any E&T method. Thus, if validity is the primary goal, a Behavioral Consistency procedure is the E&T method of choice, provided the job requires high-level reading, writing, and, analytic skills.

Table 1 summarizes the criterion-related validity findings regarding each E&T method discussed above:

Table 1 - Reported Validity and Reliability

METHOD	VALIDITY CORRECTED META-ANALYSIS	INTER-RATER RELIABILITY
Point	.11 to .15	.76 to 1.00
Grouping	No Data	.48 to 1.00
KSA Based	.20	No Data
Task-based	.15 to .28	.78 to 1.00
Behavioral Consistency	.45 to .49	.79 to .84
Holistic	No Data	.82

Relatively little information is available regarding the comparative costs associated with development and administration of E&T examinations using different methods. According to Dye & Reck (1988), developmental costs are largely a function of the amount of data required to develop the test instrument. The Point and Grouping methods cost the least to develop because much of the data needed may be readily extracted from class specifications and job descriptions.

In contrast, the development costs for "Improved" Point, Task-based, KSA-based, and Behavioral Consistency Method E&Ts require extensive job analysis data, as well as assistance from Subject Matter Experts. A summary of estimated relative costs for each E&T method is shown below in Table 2. Some of these estimates were reported in (McDaniel & Schmidt, 1985); others are based on the author's judgment.

Table 2 - Estimated Costs of E&T Development and Administration

METHOD	ESTIMATED COST OF	
	DEVELOPMENT	ADMINISTRATION
Point	Low	Moderate
Grouping	Moderate	Moderate
KSA-based	Moderate	Moderate
Task-based	Moderate	Low
Behavioral Consistency	High	High
Holistic	Low	Moderate

In operational use, administrative costs depend primarily upon the resources needed to derive candidate scores. Task-based ratings can be administered very inexpensively through template scoring, because candidates do most of the work of assessment. The costs associated with all the other methods are higher, because raters must read considerable material and then extract and evaluate the relevant portions of that material prior to making a rating. Rating time will vary considerably from examination to examination, depending on the complexity of the rating plan and the method used.

Use of Specific Rating Elements

The E&T Rating Methods section of this monograph focused on the validity of E&T instruments as a whole. Typically, the research summarized in that section examined correlations between total score on an E&T Rating and job performance. However, several researchers have also examined specific education and experience measures used as predictors of job performance. These studies deserve attention because they suggest ways in which the criterion-related validity of E&T ratings may be improved.

A. Educational Measures

One study (Dye & Reck, 1988) identified five types of educational measures used for selection. These measures include:

- Academic Achievement
- Level of Education
- Field of Study
- Academic Achievement
- Institutional Quality
- Participation in Extra-curricular Activities

Each of these measures and the findings regarding their validity will be discussed separately.

Academic Achievement. School grades and class rank are by far the measures of education most commonly used for assessment. This is probably so because grades are popularly viewed as a common and seemingly objective measure for comparing candidates. Underlying the use of academic achievement as a predictor is the presumption that there is a linear relationship between academic achievement and job performance: higher levels of academic achievement are associated with higher levels of job performance. One researcher (Baird, 1982) provides three persuasive arguments supporting the existence of a strong relationship between academic achievement and job performance. First,

the cognitive skills needed to attain good grades are the same ones needed to perform well on the job. Second, the level of motivation demonstrated in the achievement of grades will carry over to job performance. Third, grades are indicative of the level of knowledge and understanding gained from a particular field of study.

However, (Nelson, 1975) and others also comment on this relationship. First, the strength of the relationship between academic achievement and job performance is dependent on the similarity of the cognitive skills needed to attain good grades and those needed to perform well on the job. Second, academic achievement has only limited value in predicting the range of skills needed on the job. For example, some job competencies such as physical ability or interpersonal skill require little that can be addressed through academic achievement. Third, life experiences and/or other influences that are unrelated to academic achievement may effect job performance. Differences in grading standards across institutions, mean differences in student ability levels, grade inflation, differences in course difficulty, self-selection of classes, and differences in how GPAs are calculated all contribute to the unreliability of academic achievement measures.

Three meta-analytic studies addressed the validity of academic achievement as a predictor of job success and found a positive relationship. The first (O'Leary, 1980) reported on 31 studies involving 6,782 subjects. This study resulted in a corrected correlation of .17. The second study (Dye & Reck, 1988) reported a corrected correlation of .18 for 50 studies involving 7,518 subjects. The third (Lyons, 1989) found positive correlations between overall grade point average (GPA) and job performance of .20 for nine studies covering 7,453 subjects.

However, other researchers have come to the opposite conclusion. In one meta-analytic study involving 209 correlations, Sampson, Graue, Weinstein & Walberg (1984) reported, "Academic grades and academic test scores are nearly useless in predicting occupational effectiveness." In another study, Cohen (1984) found "little relationship between grade averages and job success." Bretz (1989) used much of the same data as the other studies cited above but applied slightly different meta-analytic techniques. This study used salary growth rather than supervisory ratings as the criterion measure of job success. This study found

academic achievement to have no predictive validity. (Interestingly, Bretz (1989) reported that the number of hours worked while in school was a far better predictor of salary growth than was academic achievement.)

Still other researchers have examined the relationship between GPA and success in training. Harris (1988) reported an average correlation coefficient of .31 for an unspecified number of studies covering 2,418 subjects, conducted by the United States Office of Personnel Management (U.S.O.P.M.).

Many sub-elements of GPA information, such as GPA in the major field, are routinely used as predictors of job success. Dye & Reck (1988) examined the relationships among job success, success in training and six sub-sets of academic achievement. The findings reported in their paper are summarized below in Table 3.

Table 3 - Validity and Reliability of Academic Achievement Measures

MEASURE OF ACHIEVEMENT	— JOB PERFORMANCE —			— SUCCESS IN TRAINING —		
	CORRECTED CORRELATION	N OF STUDIES	N OF SUBJECTS	CORRECTED CORRELATION	N OF STUDIES	N OF SUBJECTS
High School						
Class Rank	-.02	12	721	No Data	-	-
G.P.A.	.19	13	2445	No Data	-	-
College						
Overall G.P.A.	.18	50	7518	.29	14	1652
G.P.A. (first 2 years)	.34	13	862	No Data	-	-
G.P.A. (last 2 years)	.39	14	1133	.39	4	635
G.P.A. (Major Field)	.35	31	3411	.35	11	1133

However, Dye & Reck (1988) and others caution that sub-elements of academic achievement may be sensitive to the effects of moderator variables, such as the type of job to be filled and the field of study for which GPA is reported. Dye & Reck (1988) reported significantly higher

validities for overall GPA as a predictor of success for jobs in business, nursing, and military/government careers than for jobs in teaching, engineering, and medicine. This study also reported higher validities for GPA in the major field when used as a predictor of job success in secondary education and nursing than when used to predict job success in the fields of business, engineering and medicine. O'Leary (1980) reported an average correlation coefficient of .26 for four studies in which overall GPA was used as a predictor of job success for 2,723 managers. He also reported a correlation coefficient of .08 for seven studies in which overall GPA was used to predict the performance of 693 engineers. The second study (Lyons, 1989) examined the predictive power of various GPA sub-sets. This study found a correlation of .35 for GPA in the major field (14 studies, 1,113 subjects) and .31 for GPA during the junior and senior years (31 studies, 3,411 subjects).

Level of Education. This indicator refers to the number of years of formal education successfully completed as well as the attainment of degrees. One rationale for using Level of Education as a predictor is that those who possess educational credentials have demonstrated a certain level of intellectual ability and/or acquired the minimum competency needed to function on the job (Merritt-Haston & Wexley, 1983).

The use of Level of Education as a predictor has been attacked on the grounds that it fails to ensure equal or prerequisite learning. First, it fails to take into account that knowledges and skills can be acquired in ways other than formal education (Anastasi, 1988). Second, the use of this measure presumes that all those who attain a certain Level of Education have acquired essentially identical job skills (Schmidt, et al., 1979). Finally, Dye & Reck (1988) cite earlier work done by Lamont & Lundstrom (1977), which provides empirical evidence that education beyond a certain point may not yield correlative improvements in performance.

Other researchers have speculated that the "true" validity of Level of Education may be consistently underestimated. The pivotal issue is the effect screening for Minimum Qualifications may have on the data. If Minimum Qualifications are set too high, strict screening tends to create a more or less homogenous applicant pool. In effect, overly strict

Minimum Qualifications capture essentially the same variance as the subsequent E&T rating procedure. The effects of homogeneity are most pronounced in relation to measures like Level of Education.

Two meta-analytic studies examined Level of Education as a predictor of job success. The first of these (Dye & Reck, 1988) calculated an average correlation between Level of Education and job performance of .13 for 592 studies including 67,809 subjects. A second study (Lyons, 1989) calculated an average correlation coefficient of -.04 for eight studies involving 6,882 subjects.

One meta-analytic study addressed the relationship between Level of Education and success in training (Harris, 1988). This study reported an average correlation coefficient of .29 for an unspecified number of studies involving 17,196 subjects. This study also reported correlation values between -.01 and .29 for an unspecified number of studies conducted by U.S.O.P.M. The average correlation coefficient across these studies was reported to be .28

Field of Study. This is an indicator of exposure to various courses and subject matter. The most commonly used variant of this predictor is college major. However, the number of courses studied in a particular subject and the hours of instruction received in a particular course of study are also used. Field of Study is used as a predictor on the assumption that individuals differ in the range of knowledges and skills they possess and that these differences are at least partially dependent on the previous training an individual has received (Dye & Reck, 1988). Presumably, candidates who have received formal training in a particular area of specialization will require less training on the job.

Based on the logical underpinning of Field of Study, this indicator should be viewed as highly time dependent. It should be assumed that work performed during the first few years in a career field relies most heavily upon the application of knowledges and skills acquired through preparatory training. The literature supports this view.¹¹ If so, then Field of Study will have more impact on job performance during the first year or so after hire than it will have after an individual gains more on-the-job experience.

¹¹ Bisconti, A.S. (1980), *College and Other Stepping Stones*. Bethlehem, PA: The CPC Foundation.

Criterion-related validity data for Field of Study as a predictor is practically non-existent. Dye & Reck (1988), suggest two reasons for this. The first is that Field of Study requirements are most often supported through content validity. The second is that requirements for possession of degrees or similar educational credentials are occupationally specific. The limited validity evidence that does exist for Field of Study is mixed. Dye & Reck (1988) reported two earlier studies,¹² which found no relationship between Field of Study and job performance. They reported one study,¹³ which found Field of Study to be predictive. At this point, insufficient data is available to draw any firm conclusions regarding the predictive value of Field of Study.

Institutional Quality. This indicator has been used as a predictor of job performance on the assumption that institutions of "higher quality" maintain higher selection and retention standards than do other institutions, and provide better education. Therefore, the graduates of such institutions should have better credentials, which should translate into better work performance. According to one recent source (Dye & Reck, 1988), the literature contains no evidence supporting the validity of Institutional Quality as a performance indicator.

Extra-curricular Activities. This indicator addresses activities performed outside the classroom, such as participation in student government, sports, or accomplishments in art, science, literature, drama, and the like. Frequently, crediting distinctions are made between participatory and leadership roles within such activities. Strictly speaking, this indicator is not really a measure of academic achievement. However, it is generally treated as one on the premise that "education" is an interrelated set of experiences, gained in and out of the classroom.¹⁴ Participation in Extra-curricular Activity is presumed to indicate leadership and well-roundedness (both of which are constructs) and/

¹² Calhoun, R.P. & Reddy, A.C. (1968), *The Frantic Search for Predictors of Success*. Journal of College Placement, Vol. 28, 55-66. Sampson, G.E., Graue, M.E. Weinstein, T., & Walberg, H.J. (1984), *Academic and Occupational Performance: A Quantitative Synthesis*. American Educational Research Journal, Vol. 21, 311-321.

¹³ Howard, A. (1986), *College Experiences and Managerial Performance*. Journal of Applied Psychology, Vol 71, 530-552.

¹⁴ Bisconti, A.S. (1980), *ibid*.

or special talents that may be defended on the basis of content. This indicator may also be a way of assessing "teamwork," a KSA of growing importance. Further discussion regarding the rating of experience follows in Part B of this section.

No meta-analyses of participation in Extra-curricular Activity as a predictor of job performance were found. One study (Dye & Reck, 1988) cited two earlier works¹⁵ that found positive relationships of unspecified strength between the amount of Extra-curricular Activity and various measures of job success. One additional paper (Lyons, 1989) cited an earlier unspecified study that reported a correlation of .28 between participation in Extra-curricular Activity and job success. In the absence of more substantial empirical data, the validity of participation in Extra-curricular Activity remains an open question. If this indicator is to be used in E&T ratings, it should be done only to the extent that there is a demonstrable content link between the Extra-curricular Activity being credited and a job activity.

Comments and Conclusions: There seems to be general agreement that High School Class Rank has no value as a predictor of job success. The evidence for and against all of the other measures of academic achievement was quite mixed. The majority of the meta-analytic studies reported that all measures of academic achievement except High School Class Rank have some value for predicting job success in at least a few circumstances.

While the relative merits of different criterion measures are certainly debatable, it is likely that differences in the criterion measure selected accounted for at least some of the difference in findings. Given the existing economic climate and the emphasis on "doing more with less," validation studies linking E&T scores to supervisory ratings of job performance may be of more immediate interest than those correlating E&T scores with salary growth, rate of promotion, or self-reports of job satisfaction, all of which are legitimate alternative measures of job success. If so, then the use of academic achievement in E&Ts is supported by the literature, though not unreservedly.

¹⁵ Calhoun, R.P. & Reddy, A.C. (1968), *op. cit.*
Howard, A. (1986), *ibid.*

Grade Point Average (last 2 years) and Grade Point Average (major field) seem to be the most generally useful predictors of success in training and job performance. However, it is extremely difficult to establish a credible link between a college GPA in a field that is unrelated to the job. For this reason, GPA should be used as a predictor only in those instances where a clear and direct link can be shown to exist between the degree program in which the GPA was earned and the requirements of the job to be filled.

Level of Education is not a preferred predictor of job performance in E&T ratings. However, the literature does support the use of Level of Education to predict success in training.

Field of Study should be assessed through evaluation of Minimum Qualifications. If it can be demonstrated that possession of a particular academic qualification is required to perform the job—as is presumed to be the case with medicine, law, and most other regulated professions—then Field of Study can be defended on the basis of content validity. If Field of Study is used as a scored item on an E&T rating, it should be restricted to crediting fields of study that are directly related to the job.

There is no support in the literature for the validity of Institutional Quality as a performance indicator. In the absence of such support and given the difficulty of establishing and applying criteria for the ranking of institutions, this indicator should not be used.

The validity of participation in Extra-curricular Activity remains an open question. If this indicator is to be used in E&T ratings, it should be done only to the extent that there is a demonstrable content link between the Extra-curricular Activity being credited and a job activity.

Table 4 summarizes the preceding discussion regarding the use of educational measures as predictors.

Table 4 - Recommended Uses of Educational Measures

EDUCATIONAL MEASURE	VALIDITY	— USE TO PREDICT —	
		TRAINING SUCCESS	JOB PERFORMANCE
Academic Achievement			
High School Class Rank	No	No	No
High School G.P.A.	Low	No	Yes
College G.P.A. (cumulative)	Low	Yes	Yes
College G.P.A. (first 2 years)	Moderate	No	Yes
College G.P.A. (last 2 years)	Moderate	Yes	Yes
College G.P.A. (major field)	Moderate	Yes	Yes
Level of Education	Low	Yes	Yes
Field of Study	Unknown	Situationally Specific	No Data
Institutional Quality	Unknown	No	No
Extra-curricular Activities	Unknown	Situationally Specific	Situationally Specific

B. Work Experience

Regardless of the E&T method used, all rating schemes use prior work experience as a predictor of future performance. Historically, prediction is based on some combination of three indicators, these being the type, quality, and duration of prior experience. The underlying assumption is that the prior performance of work related to the job will predict future job performance.

Three meta-analytic studies were found which explored the validity of crediting prior experience. The first of these (Hunter & Hunter, 1984) correlated the length of experience in the current job for the current employer with supervisory ratings of job performance. This study

yielded a mean validity coefficient of .18, corrected for unreliability of the criterion measure. The second study (Schmidt, Hunter & Outerbridge, 1986) was of similar design and yielded a corrected mean validity coefficient of .33.

Schmidt, Hunter & Outerbridge (1986) argued that it is the *relative*, rather than the *absolute*, differences in experience that produce individual differences in job knowledge and performance. If the relationship between length of experience and job performance is non-linear, then using length or duration of experience as a predictor in Point Method E&Ts may partially account for the low validities generally reported in the literature. They illustrate their point with the following example:

Assume 100 persons are hired at the beginning of each year for 4 years. At the end of the 4th year, the least experienced employee has only 25% as much experience as the most experienced employee (1 year vs. 4 years). However, after 20 years of continuous employment, the least experienced employee has 85% as much experience as the most experienced employee (17 years vs. 20 years).

One implication of this hypothesis is that up to a certain length, experience may very well predict future job performance. However, beyond that length, prediction breaks down.

In the third study (McDaniel, Schmidt & Hunter, 1988b), total length of relevant job experience with all employers was correlated with supervisory ratings in the current job. This study yielded a corrected validity coefficient of .27. This same study also investigated the moderating effect of length of experience on the validity coefficients obtained. The validity coefficients were highest (.49) for samples where the mean length of job experience was less than 3 years. Validity coefficients declined as the mean length of experience increased. The lowest validity coefficient reported was .15 for the sample having more than 12 years of job experience. When corrected for unreliability, small sample bias, and range restriction and range enhancement, the validity coefficients increased for samples having little experience and decreased for samples having more. This study offers some empirical support for the view that after the first few years of work, the value of length of experience as an E&T measure decreases dramatically.

One related point deserves mention. Some E&T rating plans limit creditable experience and training to that which was gained within a certain period of time (typically 3 to 5 years) immediately preceding the rating. No analytic studies were found to address the desirability of this practice. However, Lyons (1989) reported two unpublished studies done by staff of the U.S. Office of Personnel Management, which found "limited support" for the validity of recency of education but cited no specific evidence on this matter.

Comments and Conclusions: In the absence of strong empirical evidence to the contrary, it is prudent to assume that work experience is probably a situationally specific predictor. If so, then prior work experience should be credited only in the same occupational area as that in which performance is to be predicted.

On the basis of very limited data, more work experience does not necessarily mean a better worker. Credit for length or duration of work experience should be limited to a few years and should be considered to be most useful for predicting performance in jobs for which candidates will typically have relatively little experience.

Insufficient data are available to firmly establish the legitimacy of using "recency" as a decision rule in crediting experience or training. However, many examples can be cited of job fields in which rapid technological advances have brought about significant changes in the work and the training/experience needed by employees to carry out that work. It can be argued that in such fields, recent experience and training provides exposure to current knowledge and technology while dated experience does not. If recency is to be used as a decision rule for awarding credit, it should be justified on a rational case-by-case basis.

Additional Issues and Research Findings

So far, this monograph has focused on various ways in which E&T ratings are structured as a whole and specific elements rated in those structures. However, this discussion would not be complete without some consideration of the means used to derive candidates' scores. E&T examinations are commonly rated and scored by personnelists or SMEs who are specifically trained to conduct those ratings. However, a number of E&T strategies employ self-assessment by candidates. KSA-based ratings are sometimes structured to utilize candidate self-assessments, while Task-based ratings are often structured in this manner. Coupled with automation, self-assessment E&T ratings may be very cost effective, provided the concerns self-assessment engender among selection specialists can be successfully addressed.

This section will explore the accuracy and reliability of candidate self-assessment and the effect different types of rating scales may have on the results achieved through self-assessment. It will also discuss alternative strategies through which E&Ts can be automated.

Self-Assessments/Self-Reports/Self-Ratings

In personnel selection, the term *self-assessment* has two distinct meanings (MacLane, 1979). It can refer to self-reports: candidates reporting information about their past experiences and accomplishments that demonstrate their ability to perform certain jobs. Candidates have the opportunity to determine which of their past experiences have the most relevance to the qualifications needed to perform a particular job and then highlight those experiences. Self-reported achievements are commonly used as an integral part of all E&Ts. However, trained evaluators, rather than the candidates themselves, usually make the final judgment regarding the scores candidates receive. *Self-assessment* can also, as used in this monograph, refer to candidates making their own evaluations of their achievements and capabilities. It is this aspect of self-evaluation which generates the most concern and interest among personnel psychologists. The term *self-ratings* will be used to differentiate this second definition of self-assessment from self-reports.

Self-ratings have considerable appeal to selection specialists for several reasons. First, because candidates know more about themselves than anyone else, they should be in a better position to evaluate their capabilities and limitations. Second, the cost of administering and scoring self-ratings should be considerably lower than the cost of examinations requiring evaluation and interpretation by examiners.

Self-ratings employed in E&Ts used for selection purposes continue to be viewed with some suspicion. This is largely due to belief that self-ratings (a) typically result in overestimates of performance, (b) are neither reliable nor valid, and (c) lack discriminability (van Rijn, 1980). Until recently, available studies tended to support the view that self-ratings were inherently flawed because of response bias problems such as self-enhancement of ratings and lack of objectivity by candidates (DeNisi & Shaw, 1977) and (Anderson, Warner & Spencer, 1982). Broderick (1992) correlated E&T scores with self-ratings of job performance (1,137 studies, 22,491 subjects). He reported a corrected correlation coefficient of .42. Although this view is moderated somewhat by the findings of other studies (Mabe & West, 1982), much of the research on self-rating was conducted in an academic setting or under conditions in which the result of self-ratings had negligible effects, economic or otherwise, on the subjects who provided them. The absence of potential economic ramifications effectively eliminated the normal incentives that typically tempt candidates to enhance or inflate their capabilities and experiences.

Validity of Self-Ratings. Self-reports are an inherent part of any E&T rating scheme, while self-ratings are an option in many rating situations. In recent years, research has provided evidence that E&T instruments involving candidate ratings of their own abilities and/or achievements may have greater validity and reliability than previously believed (Thornton, 1980) and (Ashford, 1989). One early study (Levine, et al., 1977) compared self-ratings of typing ability with the scores achieved on objective typing tests. This study yielded correlations ranging between .03 and .58, with a median of .36. A later meta-analytic study (Mabe & West, 1982) examined data from 55 studies, involving 14,811 subjects and 267 correlation coefficients. This study produced a

weighted mean validity coefficient of .29 for self-ratings in relation to a number of different criterion variables. A third study reported higher validities for examinations that used self-ratings of academic achievement than for examinations using objective records such as transcripts (Dye & Reck, 1988).

Reliability of Self-Ratings. Consistent with the limited interest shown in self-ratings generally, very little research has been conducted that examines their reliability. In those studies of self-ratings which have been done, practical constraints have operated to limit the availability of information on reliability. Estimates of internal consistency are not calculated, because most ratings are single ratings of each of a number of totally different dimensions. Test-retest reliability estimates are not appropriate, since it is difficult to assure that the candidates didn't change over time. Finally, the third type of reliability estimate, inter-rater agreement, is by definition meaningless for self-ratings.

One researcher (van Rijn, 1980) calculated internal consistency reliability estimates for a number of specific dimensions rated in an earlier study. These estimates ranged from .47 to .74. Using data from another study, van Rijn (1980), calculated a composite test-retest reliability estimate of .74.

Score Inflation and Faking. Score inflation by candidates remains a significant issue. One group of researchers examined the effect of inflation bias in self-rating based E&T examinations involving 351 subjects applying for 13 job titles (Anderson, et al. 1982). Their study incorporated fictional tasks into the test instruments as a means of identifying inflation bias. In this study 45% of the subjects evidenced some degree of inflation. The study also yielded a mean correlation of .61 between examination score and inflation bias. A meta-analytic study compared 36 independent sets of self-ratings of work performance and supervisory performance ratings (Harris & Schaubroeck, 1988). They found that self-ratings of job performance consistently yielded scores about one-half a standard deviation higher than the supervisory ratings of job performance.

Some researchers suggest that differences such as those reported by Anderson, et al. (1982), and Harris & Schaubroeck (1988) may result from something other than inflation. One plausible explanation for these findings is that supervisors may not fully understand the work being done by those they are rating. For example, those who supervise technical and/or medical personnel often have difficulty in providing accurate performance appraisals of their professional subordinates. Others suggest that the validity and reliability of self-ratings are directly related to the extent to which candidates understand the rating dimensions (Levine, et. al, 1977) and (Primoff, 1980). Generally, the validities are higher when the rating dimensions are made explicit and are clearly understood by the candidates (van Rijn, 1980).

These same studies show that in addition to understanding the dimensions being measured, candidates must also know where to rate themselves in terms of the rating anchors and in relation to the other candidates. Several researchers suggest that "relative rating scales" (e.g., "I have more ability or knowledge than 10% of those doing this work" or "I know much less about this topic than the average person") provide candidates with the frame of reference they need for reliable self-rating (van Rijn, 1980) and (Lin, Dobbins et al., 1992).

Mabe & West (1982) suggest a number of ways in which the accuracy of self-reported information can be enhanced including:

1. Correspondence between the self-rating measure and the performance to be evaluated on the job.

Although the need for such a correspondence may seem obvious in principle, poor matches between predictors and their associated criterion measures may not always be so obvious in practice. Mabe & West (1982) cite the example of a study in which self-reports of "Accuracy of Detail" were correlated with supervisory ratings on the "Ability to Type Statistics."

2. The focus of self-ratings should be on past and future performance rather than on the possession of abilities.

Self-ratings of ability involve an interpretation of the relative contribution of that ability in explaining a hypothetical performance outcome. In contrast, self-ratings of performance allow a candidate to match their experiences with the performance required on the job.

3. Self-rating measures that are too complex to be measured in absolute terms should be phrased in ways that allow candidates to make comparative evaluations of their performance in relation to others in the rating pool.

The primary function of most E&T ratings is to order candidates on the basis of their expected job performance. When quantifiable measures, such as typing speed, are measured, it is sufficient to list a range of performance and ask candidates to rate themselves against these objective standards. However, when self-ratings are used to assess complex requirements, such as building stair risers, it may not be possible to develop objective standards. In those instances, most candidates will attempt to evaluate their performance in relation to others (e.g. better than 90% of those who perform this function, better than average, etc.), rather than in an absolute way (e.g. good, excellent, etc.). Rating scales used for complex measures should reflect this approach.

4. Candidates should have the expectation that their responses will be reviewed or verified.

Several researchers cited earlier works which suggest that the accuracy and validity of self-ratings may be improved if candidates expect verification of the information they report. One such mechanism involves the comparison of self-ratings with external criterion measures. Under this approach, evaluators review the self-ratings. If the ratings appear to be consistent with the candidate's background, credit is given for the response; if not, then credit is withheld (Mabe & West, 1982). Another approach is to require candidates to list sources of verification, such as previous employers, who may be

contacted as the final step in the selection process (van Rijn, 1980).

Comments and Conclusions: The use of candidate self-rating is a potentially valuable assessment strategy, but one that is rife with unanswered questions. Too little research has been done to provide a consensus of opinion regarding the validity and reliability of self-rating strategies. Researchers offer a number of suggestions for controlling potential sources of unreliability and improving validity:

1. Require candidates to provide information needed to verify their claims.
2. Use one of the suggested mechanisms to detect and control candidate faking (e.g. lie scales used in conjunction with a regression equation to factor out inflated scores).
3. Take care in constructing the descriptions of the rated dimensions and the scales candidates will use to rate themselves to insure that the candidates understand them.

However, it must be noted that the effectiveness of these suggestions has not been empirically demonstrated. There is no absolute means of assuring that self-ratings will be accurate. If a decision is made to use self-ratings, it should be with the clear understanding that the efficacy of mechanisms used to control bias-induced inflation, faking, and other potential sources of inaccuracy have not been fully proven. Provided the risks associated with self-ratings are considered acceptable, such instruments can be cost-effective selection tools.

B. Automated E&T Testing

Electronic technology has been used in support of selection activities for some time. For example, the use of scannable written test answer sheets and computer scoring of such answer sheets is a common practice. In recent years, the introduction of low-cost computers, the availability of sophisticated software, and the increasing computer

literacy of testing specialists have led to the rapid expansion of selection strategies that use advanced technology.

With the exception of Behavioral Consistency ratings, any of the E&T methods described in this monograph are readily adaptable to automated administration and scoring. Even Behavioral Consistency ratings may, with some difficulty, be adapted to automation through the use of benchmark achievement indicators¹⁶ (Smith & Kendall, 1963).

Mechanisms for E&T automation include the use of scannable forms as well as direct computer entry of data by candidates. The use of templates to score E&T ratings may also be considered. E&T scoring templates are used in a manner similar to that through which the accuracy of machine scored written test answer sheets are verified by hand. While template scoring falls short of true automation, it provides some of the same advantages and could be viewed as a logical intermediate step toward automation.

Scannable Forms. The use of scannable forms for E&T ratings is an adaptation of the generally well known technology currently used for scoring written tests. In brief, the information which serves as the basis for E&T scores is elicited from candidates using questionnaires and response forms designed for machine scoring. The text and layout of the questionnaires and response forms can vary according to the job classification and the information needed for scoring. Any system based on the use of scannable forms requires an Optical Mark Reader (OMR) capable of reading and processing data from various forms, and forms which can be read by the OMR.

Previous generations of OMR equipment were sometimes "cranky" and had very limited capabilities. Sensitivity to extraneous marks on documents was often a problem. Documents that had been folded for mailing frequently jammed the feed mechanism. The instructions used to define where the OMR was to look for marks on a document, and how to treat the marks it found, required "hard wiring" of program boards that were inserted into the OMR.

¹⁶ Scoring keys which are similar in structure to Behaviorally Anchored Rating Scales (BARS).

Newer scanning equipment typically has adjustable sensitivity, allowing for greater reading reliability and accuracy than was previously possible. Many scanners can now process folded documents, allowing the user to scan mailed forms. Finally, most newer scanners are software programmable. That is, scoring instructions can be entered on a personal computer attached to the OMR. The coded instructions are "fed" to the OMR through a "driver program" loaded on the personal computer. (The principle is the same as that used in personal computers to control an attached printer.) This feature eliminates the need for the "hard wired" program boards used with many older OMRs. In short, OMRs are now versatile machines that are readily adaptable to scoring E&Ts.

All scannable forms must contain timing marks that orient the document for the OMR. The placement of these marks is critical. If the timing marks are placed inaccurately or are not clearly identifiable as such by the OMR, the form will not be readable or responses will be read incorrectly. Placement of the "bubbles" read by the OMR is also critical. To be read properly, the bubbles must be positioned accurately in relation to the timing marks. There are two ways in which OMR readable forms can be obtained: they may be purchased from vendors or generated in-house, using software designed for that purpose.

Forms produced by vendors have to be laid out in advance, and purchased in quantity, if they are to be economical. Scannable E&T forms would contain spaces for candidate identification, some basic instructions for completing the form, and one or more columns of scannable bubbles with appropriate headings. The forms would also contain question numbers or blank space in which the text of E&T questions could be overprinted. One advantage of this approach is the time saved during examination development. The use of a standard format for questionnaires and rating scales eliminates the need for selection specialists to develop them from scratch for each examination. However, it should be noted that forcing the text of E&T questions and explanatory notes into the constraints of a pre-existing form may not be easy or desirable. The most significant limitation of this approach is that all E&Ts would have to conform to one of a limited number of pre-

established standard models and standard rating scales. Flexibility would be minimal.

The production of in-house forms requires special OMR forms design software, a 386 or 486 personal computer, and a laser printer having PostScript capability. The capability of generating forms in-house provides considerable flexibility of design and layout. E&T text, instructions to candidates, and rating scales, tailored to almost any situation, can be readily transferred to an OMR readable form.

If hardware and software need to be purchased, the initial start-up costs for in-house forms development would be about double those associated with vendor purchased forms. However, the use of in-house forms generation provides greater flexibility and greater control over the product. Since forms can be printed on standard 24-pound paper stock, the capability to develop in-house forms could be very cost-effective over the long term.

Regardless of how the forms are generated, a mechanism must exist to tell the OMR how to read and score information. As indicated previously, on newer OMR equipment, this is the function of a "driver program." Such programs are available with a range of capabilities. Basic OMR driver programs allow "hit or miss" scoring only. A mark in the correct location is treated as a "hit," which accumulates a point to the score counter of the OMR. A mark in any other location (or no mark at all) is treated as a "miss", for which no credit is given. This limited capability is adequate for scoring multiple-choice written tests and some Point Method variant E&Ts. Items used in these instruments have only one "correct" answer.

While it is possible to structure E&T selection procedures as a battery of discrete "yes/no" questions to accommodate "hit or miss" scoring, this approach has a number of drawbacks. Except for the least complex jobs, the number of questions needed to cover all creditable combinations of experience and training can be huge. Second, it is often quite difficult to structure questions in a manner that allows full credit for all relevant training and experience while still avoiding overlap or redundancy. Third, no matter the care taken to develop E&T questions in this format, there is a likelihood that something important will be

missed. Finally, it is difficult to properly weight items according to their importance or criticality.

“Hit or miss” scoring is inadequate for use with typical E&T selection procedures. Typically, E&T questions cover training and experience which vary in significance to the work. Moreover, it is rarely meaningful to evaluate the responses to individual E&T questions in terms of being either correct or incorrect. Rather, responses usually fall along a continuum of desirability, ranging from highly undesirable through neutral to highly desirable. Some software packages used to create in-house forms provide enhanced OMR drivers, which permit differential scoring of responses. Using such a driver program, it is possible to assign different values to each of five or more possible responses to a question, making the OMR much more amenable to scoring E&T instruments. However, such driver programs are not available with all software packages or for all OMRs.

Comments and Conclusions: For many reasons, the use of scannable forms is an attractive alternative for automating E&T tests. With the possible exception of learning to apply new conventions for document layout and formatting, selection specialists should need no new skills to develop or use automated E&T test instruments. Second, given a high enough volume of ratings, the use of scannable forms would result in cost savings relative to manual scoring. Finally, scannable forms provide convenience for candidates, since forms can be sent to candidates to complete and return on their own. A parallel manual data entry capability must be maintained in addition to the automated E&T system. Inevitably, at least some of the E&T forms received will be so badly defaced that they will have to be scored or keyed in manually. Finally, the issues raised above regarding the accuracy of self-ratings need to be considered and steps taken to address those issues.

Direct Data Entry. Direct data entry involves the administration of E&Ts to candidates via personal computer (PC). PC-based systems permit selection specialists to develop E&T instruments and the scoring instructions to be used with those instruments via PCs, the completion of specific E&T instruments by candidates via PC, and the electronic

scoring of those instruments. More comprehensive systems, such as one being pilot tested by the U.S. Office of Personnel Management, include more support and development features, such as the entry of job analysis data, development of E&T instruments using the job analysis data, and vacancy information tracking.

The number of commercial software packages designed for use in developing and administering PC-based E&T ratings is, at present, quite limited. This is unfortunate, because such software minimizes the training needed by test developers and administrators who want to use PC-based E&T ratings. Commercial software packages rely heavily on the use of prompts, dialog boxes and menus to invisibly perform much of the work involved in E&T instrument development and layout, administration of the E&T to candidates, and E&T rating.

However, viable alternatives do exist. Tests and scoring procedures can be programmed in PC-based languages such as Basic, C++, and Pilot, as well as in more difficult "mainframe" languages such as TurboPascal and Cobol. Many commercial database programs are available in "developer editions," which allow a user to write stand-alone test administration and scoring routines that can be distributed to remote sites for independent use. Finally, a few commercial software programs have ready-to-run E&T testing modules. At this point, several public jurisdictions have initiated automation projects involving E&T ratings.¹⁷

Comments and Conclusions: There is no technological reason why PC-based E&T testing cannot become a practical reality for many who now develop and administer such programs manually. However, the initial start-up costs of PC-based E&T assessment using any of these options appear to be fairly high—high enough that it is difficult to see how conversion to PC-based E&T assessment could be justified on economic grounds alone. None of the software packages mentioned above are inexpensive. PCs need to be available or purchased to run the software. Selection specialists or other staff must learn to use the software or learn programming in the selected format. The testing and scoring modules may need to be developed from scratch and debugged.

¹⁷ Schneider, R.E. (1993), *The Rating of Experience & Training: Alternative Approaches for Automation*. Harrisburg, PA: PA State Civil Service Commission, Research Division.

Finally, links must be established between the programs developed for automated E&T testing and existing candidate tracking systems.

Some candidates may be computer illiterate and reluctant to admit it. Others may be extremely nervous about using PCs. Either of these circumstances may lead to inaccurate E&T scores. While it is conceivable that some day E&Ts will be available on demand to anyone with a PC and a modem at any location, the current reality requires candidates for PC-based E&T tests to report to designated test sites where the E&Ts can be administered. It is likely that the problems associated with computer illiteracy or aversion, and the increased candidate traffic generated by PC-based testing at these sites, may result in the need to assign additional personnel at these sites to serve as monitors and facilitators. The cost of these personnel could offset any savings that might accrue from PC-based E&T testing.

The truly critical unanswered questions concerning automated E&T testing actually have little to do with technology. Rather, they are administrative and practical. The answers to those questions depend to a great extent on the managerial philosophy of the organization, its willingness to commit resources to obtaining and using the necessary technology, and the likely cost-benefit of such an investment for a particular organization.

References

- Anastasi, A. (1988), *Psychological Testing (6th ed.)*. New York: McMillian
- 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*, Vol. 69, No. 4, 574-580
- Arnold, D.W. & Thiemann A.J. (1993), *Test Scoring Under the Civil Rights Act of 1991*. *Industrial-Organizational Psychologist*, Vol. 30, No. 3, 65-67
- Arvey, R.D., McGowan, S. & Horgan D. (1981), *The Use of Experience Requirements in Selecting Employees*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Denver, CO
- Ash, P. (1974), *Test Development and Validation Manual*. Chicago: University Civil Service System of Illinois
- Ash, R.A. (1980), *Self-assessments of Five Types of Typing Ability*. *Personnel Psychology*, Vol.33, 273-282
- Ash, R.A. (1981), *Comparison of Four Approaches to the Evaluation of Job Applicant Training and Work Experience*. Dissertation Abstracts International, 42, 4606B. (University Microfilms No. DE082-07909)
- Ash, R.A. (1984), *A Comparative Study of Behavioral Consistency and Holistic Judgement Methods of Job Applicant Training and Work Experience Evaluation*. *Public Personnel Management*, Vol. 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*, Vol. 15, No. 3, 325-343
- Ash, R.A., Johnson, J.C., Levine, E.L., & McDaniel, M.A. (1989), *Job Applicant Training and Work Experience Evaluation in Personnel Selection*. In G.R. Ferris & K.M. Rowland (eds.), Research in Personnel and Human Resources Management, Vol. 7, 183-226. Greenwich, CT: JAI Press, Inc.

- Ash, R.A. & Levine, E.L. (1983), *Job Applicant Training and Work Experience Evaluation: An Empirical Comparison of Four Methods*. Public Personnel Management, Vol. 13, 157-172
- Ashford, S.J. (1989), *Self-Assessment in Organizations: A Literature Review and Integrative Model*. Shaw, B. & Cummings, L. (eds.) in Research in Organizational Behavior. New York, N.Y.: JAI Press
- Baird, L.L. (1983), *The Role of Academic Ability in High-level Accomplishment and General Success*. College Board Report No. 82-6. New York: College Entrance Examination Board
- Blanchard, P.N., Camp, R.R. & Bass A.R. (1979) *Self-Assessments vs Objective Tests: A Question of Incremental Validities*. Paper presented at the annual conference of the American Psychological Association, New York, N.Y.
- Bretz, R.D. (1989), *College Grade Point Average as a Predictor of Adult Success: A Meta-Analytic Review and Some Additional Evidence*. Public Personnel Management, Vol. 18, No. 1, 11-22
- Broderick, W.N. (1992), *A Ten Year Meta-Analysis of Validation Studies of Job Element Based Behavioral Consistency Questionnaires Using Self-Ratings as a Criteria*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Baltimore, MD
- Beardsley, V.A. (1976), *A Study of the Rating of Education and as an Examination Method in the Pennsylvania State Civil Service Commission*. Harrisburg, PA: PA State Civil Service Commission, Bureau of Personnel Assessment
- Cohen, P. (1984) *College Grades and Adult Achievement*. Research in Higher Education, Vol. 20, 281-293
- Dye, D.A. & Reck, M. (1988), *A Literature Review and Meta-Analysis of Education as a Predictor of Job Performance*. Washington: Office of Personnel Research and Development, U.S. Office of Personnel Management
- DeNisi, A.S. & Shaw, J.B. (1977) *Investigations of the Use of Self-Reports of Abilities*. Journal of Applied Psychology, Vol. 62, 641-644

- Farh, J.L. & Dobbins, G.H. (1989), *Effects of Comparative Performance Information on the Accuracy of Self-Ratings*. *Journal of Applied Psychology*, Vol. 74, 606-610
- Farrell, B.M. (1979), *Task Performance Self-Evaluation: An Alternative Selection Procedure to Traditional Experience and Training Ratings*. Minneapolis, MN: Minnesota Department of Personnel, Selection Research Unit
- Gion, R.M. & Gibson, W. (1988), *Personnel Selection and Placement*. *Annual Review of Psychology*, Vol. 39, 349-374
- Goldstein, G. & Hersen, M. (eds.) (1990), *Handbook of Psychological Assessment (2nd. Edition)*. New York: Pergamon Press
- Grove, D.A. (1981), *A Behavioral Consistency Approach to Decision Making in Employment Selection*. *Personnel Psychology*, Vol. 34, No. 1 55-64
- Harris, P.A. (1988), *The Effectiveness of Grade Point Average and Level of Education in Predicting Performance in the Workplace: A Lay Summary of Findings*. (OPRD Report 88-1). Washington: U.S. Office of Personnel Management, Office of Personnel Research and Development
- Harris, M.M. & Schaubroeck, J. (1988), *A Meta-analysis of Self-Supervisor, Self-Peer, and Peer-Supervisor Ratings*. *Personnel Psychology*, Vol. 41, 43-62
- Hough, L.M. (1984), *Development and Evaluation of the Accomplishment Record Method of Selecting and Promoting Professionals*. *Journal of Applied Psychology*, Vol. 69, 135-146
- Hough, L.M., Keyes, M.A. & Dunnette, M.D. (1983), *An Evaluation of Three "Alternative" Selection Procedures*. *Personnel Psychology*, Vol. 36, 261-276
- Hunter, J.E., & Hunter R.F. (1983), *The Validity and Utility of Alternative Predictors of Job Performance*. (OPRD Report 83-4). Washington: U.S. Office of Personnel Management, Office of Personnel Research & Development
- Johnson, J.C., Guffey, W.L. & Perry, R.A. (1980), *When is a T and E Rating Valid?* Paper presented at the annual conference of the International

Personnel Management Association Assessment Council, Boston, MA

Lawton, J.R. (1982), *A Study of Return Rates and Applicant Qualifications in the Behavioral Consistency Model*. Madison, WI: Department of Industry, Labor and Human Relations

Levine, E.L. & Flory, A. (1975), *Evaluation of Job Applications: A Conceptual Framework*. Public Personnel Management, Vol. 4, 378-385

Levine, E.L., Flory, A. & Ash, R.A. (1977), *Self-Assessment in Personnel Selection*. Journal of Applied Psychology, Vol. 62, 428-435

Lin, T.R., Dobbins, G.H., Farh, J.L., Doyle, T.F., & Spaulding, L. (1992), *Rating Anchors and Self-Assessment of Ability: A Field Study*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Baltimore, MD

Lyons, T.J. (1989), *Validity of Education and Experience Measures in Traditional Rating Schedule Procedures: A Review of the Literature*. (OPRD Report 89-2) Washington: U.S. Office of Personnel Management, Office of Personnel Research and Development

Lyons, T.J. (1992), *Validity of Three E & E Rating Methods for Entry and Full Performance Level Jobs*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Baltimore, MD

McDaniel, M.A. & Schmidt, F.L. (1985), *A Meta-analysis of the Validity of Training and Experience Ratings in Personnel Selection*. Washington: U.S. Office of Personnel Management, Office of Staffing Policy

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, Vol. 41, No. 2, 283-314

McDaniel, M.A., Schmidt, F.L., & Hunter, J.E. (1988b), *Job Experience Correlates of Job Performance*. Journal of Applied Psychology, Vol. 73, 327-330

Mabe, P.A. & West, S.G. (1982), *Validity of Self-Evaluation of Ability: A Review and Meta-Analysis*. Journal of Applied Psychology, Vol. 67, 280-296

- Merritt-Haston, R. & Wexley, K.N. (1983), *Educational Requirements: Legality and Validity*. Personnel Psychology, Vol. 36, 743-753
- Muchinsky, P.M. (1986), "Personnel Selection Methods". In *International Review of Industrial & Organizational Psychology*, Cooper, C.L. & Robertson, I.T. (eds.). New York: John Wiley & Sons
- Nelson, A.M. (1975) *Undergraduate Academic Achievement in College as an Indicator of Occupational Success*. (OPRD Report 75-5). Washington, DC: U.S. Office of Personnel Management, Personnel Research and Development Center
- Ocasio, B.P., & Hillary, J.M. (1983), *An Experimental Analysis of Self-Assessment for Employee Selection*. Washington: U.S. Postal Service, Office of Test Development and Validation
- O'Leary, B.S. (1980), *College Grade Point Average as an Indicator of Occupational Success: An Update* (OPRD Report 80-23). Washington: U.S. Office of Personnel Management, Personnel Research & Development Center
- Pannone, R.D. (1982), *Predicting Test performance: A Content Valid Approach to Screening Applicants*. New York: Port Authority of New York & New Jersey
- Porter, W.R., Levine, E.L., & Flory, A. III (1976), *Training and Experience Evaluation*. Tempe, AZ: Personnel Services Organization
- Primoff, E.S. (1975), *How to Prepare and Conduct Job Element Examinations*. (Technical Study 75-1). Washington: U.S. Civil Service Commission; (U.S. Government Printing Office)
- Primoff, E.S. (1980), *The Use of Self-Assessments in Examining*. Personnel Psychology, Vol. 33, 283-290
- Reilly, R.R. & Chao, G.T. (1982), *Validity and Fairness of Some Alternative Employee Selection Procedures*. Personnel Psychology, Vol. 35, 1-62
- Sampson, G.E., Graue, M.E., Weinstein, T. & Walberg, H.J. (1984), *Academic and Occupational Performance: A Quantitative Synthesis*. American Educational Research Journal, Vol. 21, 311-321

Schmidt, F.L., Caplan, J.R., Bemis, S.E., Decuir, R., Dunn, L., & Antone, L. (1979), *The Behavioral Consistency Method of Unassembled Examining*. Washington: U.S. Office of Personnel Management Personnel, 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*, Vol. 71, 430-439

Smith, P.C. & Kendall, L.H. (1963), *Retranslation of Expectations: An Approach to the Construction of Unambiguous Rating Scales*. *Journal of Applied Psychology*, Vol. 47, 149-155

Sproule, C.F. (1990), *Recent Innovations in Public Sector Assessment*. International Personnel Management Association Assessment Council, Personnel Assessment Monographs, Vol. 2, No. 2

Tenopyr, M. (1977), "Procedures for Job Analysis and Federal Examination Planning". In *Examination Preparation Manual*, Washington: U.S. Civil Service Commission

Thornton, G.C. (1980), *Psychometric Properties of Self Appraisals on Job Performance*. *Journal of Applied Psychology*, Vol. 62, 641-644

van Rijn, P. (1980), *Self-Assessment for Personnel Examining: An Overview*. (OPRD Report 80-14). Washington: U.S. Office of Personnel Management, Office of Personnel Research and Development

Appendix A

Suggested Reading

General

Ash, R.A., Johnson, J.C., Levine, E.L., & McDaniel, M.A. (1989), *Job Applicant Training and Work Experience Evaluation in Personnel Selection*. In G.R. Ferris & K.M. Rowland (eds.), *Research in Personnel and Human Resources Management*, Vol. 7, 183-226. Greenwich, CT: JAI Press, Inc.

Point Method

Porter, W.R., Levine, E.L., & Flory, A. III (1976), *Training and Experience Evaluation*. Tempe, AZ: Personnel Services Organization

KSA-based Method

Ash, P. (1974), *Test Development and Validation Manual*. Chicago: University Civil Service System of Illinois

Primoff, E.S. (1975), *How to Prepare and Conduct Job Element Examinations*. (Technical Study 75-1). Washington: U.S. Civil Service Commission; (U.S. Government Printing Office)

Task-based Method

Farrell, B.M. (1979), *Task Performance Self-Evaluation: An Alternative Selection Procedure to Traditional Experience and Training Ratings*. Minneapolis, MN: Minnesota Department of Personnel, Selection Research Unit

Behavioral Consistency Method

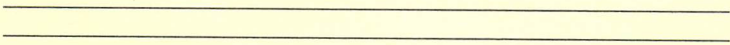
Schmidt, F.L., Caplan, J.R., Bemis, S.E., Decuir, R., Dunn, L., & Antone, L. (1979), *The Behavioral Consistency Method of Unassembled Examining*. Washington: U.S. Office of Personnel Management Personnel, Personnel Research and Development Center

Academic Achievement

Lyons, T.J. (1989), *Validity of Education and Experience Measures in Traditional Rating Schedule Procedures: A Review of the Literature*. (OPRD Report 89-2) Washington: U.S. Office of Personnel Management, Office of Personnel Research and Development

Self-assessment/Self-rating

Lin, T.R., Dobbins, G.H., Farh, J.L., Doyle, T.F., & Spaulding, L. (1992), *Rating Anchors and Self-Assessment of Ability: A Field Study*. Paper presented at the annual conference of the International Personnel Management Association Assessment Council, Baltimore, MD



IPMAAC
July, 1994

