Operational Considerations in Personnel Selection

This section of the *Principles* describes operational issues associated with the development or choice of a selection procedure, the conduct or accumulation of research to support the validity inferences made, documentation of the research effort in technical reports and administration manuals, and subsequent implementation and use. The need for sound professional judgment based on the extant scientific literature and the researcher's own experience will be required at every step of the process. In addition, all aspects of the research described in the *Principles* should be performed in compliance with the ethical standards of the American Psychological Association as endorsed by SIOP.

Topics are introduced in an order that generally corresponds to the temporal progression of the validation effort. For example, the section on understanding work and worker requirements precedes decisions regarding the selection procedure. In other cases, the placement is based on the logical relationship among the topics. Therefore, the order in which steps are taken in practice is ultimately a matter of professional and scientific judgment based on the given situation. It is recognized that in some instances a selection procedure may be implemented at the same time the validation process is underway.

Initiating a Validation Effort

The researcher works collaboratively with representatives of the organization to define its needs and objectives, identify organizational constraints, plan the research, and communicate with major stakeholders regarding aspects of the process that will involve or influence them.

Defining the Organization's Needs, Objectives, and Constraints

Researchers use their expertise and experience to assist the organization in refining its goals and objectives. Different departments of the organization may have different and sometimes competing and conflicting objectives. For instance, one department may prefer rigorous selection standards even though they create hardships for the staffing department responsible for recruiting qualified applicants. As another example, one organization may need a large number of entry-level workers where there is minimal opportunity to move upward. In another organization, the focus may be on hiring a few individuals with the capacity to move upward in a relatively short period of time. In all situations, the researcher and the organization's representatives should factor in the desires of the various stakeholders and determine the relative weights to be given to each point of view.

The researcher provides accurate information regarding the benefits and limitations of various strategies in meeting the organization's goals based on

past experience and the extant scientific research. The researcher is encouraged to work with all departments (e.g., human resources, labor relations, legal) that may have an effect on or be affected by the selection procedure and stakeholders (e.g., internal or external individuals and groups such as labor organizations, advocacy groups, customers).

Climate and culture. Researchers face the challenge of ensuring high quality selection procedures in the context of the organization's history and current environment regarding employment-related strategies and practices as well as the cultural setting in which it operates. Organizations operate in complex environments that sometimes place extreme and conflicting pressures on the management team. Researchers must consider the attitudes and commitments of organization leaders and employees who are faced with intense competition, mergers, and other corporate events that may influence the relative importance of selection research in their view. Researchers also may need to take into account the legal and labor environment when deciding on validation approaches or selection instruments. In addition, global selection systems should take into consideration locally accepted practices.

Workforce size and availability. The number of individuals who currently perform the work and their similarity to the applicant population can be important considerations when designing the validation strategy. The number of workers may shape the validation strategy pursued (e.g., validity generalization, content-oriented strategy) as well as affect the feasibility and method for pilot testing procedures.

Even when the number of workers is sufficient, their availability and willingness to participate in a validation study may be limited. For example, organizational needs may require that a core group of workers be present on the job at all times; labor organizations may influence the number and type of persons willing to participate in the research; and workers who have experienced organizational restructuring may be skeptical about the purpose of the research and its effect on their own positions.

Large discrepancies in the capabilities of incumbents and the available applicant pool also present challenges, particularly in establishing norms and setting cutoff scores. For example, organizations that have a more capable work force than applicant pool may find cutoff scores based on the performance of incumbents on the selection procedure inappropriate for applicants. Similarly, organizations seeking to upgrade the skills of their current workforce may need other sources of information for setting cutoff scores.

Sources of information. Sources of information needed for the validation and implementation efforts include, but are not limited to, the workers themselves, managers, supervisors, trainers, customers, archival records, databases, and research reports internal and external to the organization. Based on the complexity of the work, the climate, and organizational constraints, some sources of information may be preferred over others. In some

situations, the preferred source of information may not be available. Depending on the organizational constraints, alternatives to the researcher's preferred source of information may be required. Alternative sources also may be used to supplement information gathered from the preferred source.

Acceptability of selection procedures. Most organizations want selection procedures that are predictive, easy to use, cost effective, and legally defensible. However, there are often additional considerations. For example, an organization's past experiences with respect to certain types of selection procedures may influence its decisions. Selection procedures that have been challenged in the past may not be acceptable to organizations, particularly if the organization was not successful in defending them. In addition, selection procedures that are viewed as controversial by individuals, labor organizations, or other stakeholders may not be acceptable.

Some organizations find certain types of selection procedure questions unacceptable. For example, some biodata and personality inventory items (e.g., childhood experiences, personal interests) may be viewed as an invasion of privacy, even if they can be shown to be related to the criterion measures or the requirements of the job.

Some organizations prefer selection procedures that provide information regarding the strengths and developmental needs of the test taker. Procedures that measure constructs that can be learned (e.g., keyboarding or word processing) may be preferred over procedures that elicit information concerning previous life experiences or stable personality traits. Procedures that appear more relevant or face valid to the organization may be more acceptable to the stakeholders than other procedures that relate to a less obvious construct regardless of any empirical evidence of validity. However, face validity is not an acceptable substitute for other forms of validity evidence as treated in the *Principles*. Although acceptability is important, it is just one of many factors to consider when selecting or designing a selection procedure. Nevertheless, the researcher should explain to decision makers issues underlying selection procedure acceptability as part of the initial planning effort.

Communicating the Validation Plan

Both management and workers need to understand in general terms the purpose of the research, the plan for conducting the research, and their respective roles in the development and validation of the selection procedure. The researcher must use professional judgment in determining the appropriate information to provide and the communication format and style that will be most effective.

Researchers encourage organizations to consider the effects of participation in the validation effort on employees, departments, and business units. Typically, organizations decide that data from validation studies will be kept confidential and not used for subsequent employment-related decisions.

Understanding Work and Worker Requirements

Historically, selection procedures were developed for specific jobs or job families. This remains the case in many situations. However, industries that experience rapid technological development or institute other strategies for accomplishing work may find that traditional jobs no longer exist. In such cases, considering important job requirements for a wider range or type of work activity may be more appropriate.

Strategies for Analyzing the Work Domain and Defining Worker Requirements

The approach, method, and analyses used in a specific study of work is a function of the nature of the work itself, those who perform the work, and the organizational setting in which the work is accomplished. There is no single strategy that must be carried out, and multiple strategies may be appropriate.

There are situations where the importance or relevance of a construct is self-evident and does not require extensive work analysis. For example, absenteeism and turnover and their underlying constructs may be relevant to all work activities in an organization. Therefore, demonstration of their relevance is not typically necessary.

Considerations in Specifying the Sampling Plan

The sampling plan for data collection should take into account the number of workers and their work locations, their characteristics (e.g., amount of experience, training, proficiency, etc.), shift or other work cycles, and other variables that might influence the work analysis.

Documentation of the Results

The methodology, data collection methods, analyses, results, and implications for the validation effort should be documented. Frequently, this documentation will include a description of the major work activities, important worker requirements and their relationships to selection procedure content, and scoring when appropriate. The documentation should provide sufficient detail for another researcher to replicate the work analysis process. The documentation should also help the researcher understand the role of the work analysis as the foundation for any validation efforts.

Selecting Assessment Procedures for the Validation Effort

The researcher exercises professional judgment to determine those selection procedures that should be included in the validation effort. This judgment takes into consideration the organizational needs as well as the issues discussed in this section.

Review of Research Literature

Researchers should become familiar with research related to the organization's objectives. The research literature is used to inform choices about selection procedures and the validation strategy to be employed.

Psychometric Considerations

When selecting one or more predictors, a number of psychometric characteristics of each instrument should be considered. Some of the more important psychometric considerations include reliability, evidence supporting the validity of the intended inferences, and differences among subgroups.

When choosing components of a selection battery, the researcher should consider the overall contribution of each component, its relative contribution, and potential construct redundancy and decide how much construct redundancy is desirable given the instruments and the situation.

Administration and Scoring Considerations

There are practical considerations regarding the consistent administration and scoring of a selection procedure. For example, the researcher must ensure that administration and scoring tasks can be completed consistently across all locations and administrators. To the extent that the selection procedure (e.g., work samples) requires subjective judgments in scoring, issues of rater training and inter-rater reliability become especially important. If standardized conditions are violated in the administration or scoring of a selection procedure, the generalizability of findings may be compromised.

Format and Medium

Format refers to the design of response requirements for selection procedure items (e.g., multiple-choice, essay). The choice of format may be influenced by the resources available to administer and score the selection procedure. For example, objectively scored items with established correct responses may be administered and scored in less time than selection procedures that require the individual to respond in more complex ways or that use individually judged responses.

Medium refers to the method of delivery of the selection procedure content. For example, a measure of cognitive ability could be presented via paper-and-pencil, computer, video, or orally.

There are advantages and disadvantages in selecting or adapting existing selection procedures from one medium to another. Computer-administered procedures may reduce the demands on administrators and enhance standardization. Computer-administered tests also provide opportunities to measure constructs that do not lend themselves to testing by paper-and-pencil (e.g., use of spreadsheets and database management**).** Research has found that carefully developed computerized versions of cognitive ability

power tests assess the same construct as the paper-and-pencil versions (Mead & Drasgow, 1993).

Changing the medium may also change the construct being measured. For example, converting a paper-and-pencil situational judgment test to a video where the situations will be acted out will reduce the reading component of the test. Also, adapting speeded tests of cognitive ability to a computerized version has been found to alter the construct being measured (Mead & Drasgow, 1993).

A number of considerations are important when evaluating the format and medium options. Cost and efficiency of operation may be the primary concern to the organization. In addition, security, standardization of testing conditions, candidate authentication, and accessibility of testing opportunities are all important considerations. Developers of selection systems should be cognizant that format and medium can affect mean score differences among subgroups (Hough, Oswald, & Ployhart, 2001).

Acceptability to the Candidate

In addition to the organization's needs and objectives, researchers also need to consider the acceptability of the selection procedure to candidates. A number of factors influence candidates' reactions to a selection procedure, including individual characteristics (e.g., work experiences, demographics, and cultural backgrounds), the role of the individual (e.g., applicant, incumbent, manager), the extent to which the content of the selection procedure resembles the work, the individual's capability with respect to the constructs measured, and the perceived passing or selection rate. Generally, the greater the similarity between the selection procedure and the work performed, the greater the acceptability to candidates, management, and other stakeholders. However, selection procedures that closely resemble the work may be perceived as obsolete when the work changes.

Some selection procedures may appear less face valid than other procedures. For example, the value of information collected on biodata forms and personality inventories in predicting job performance may not be obvious to many. Communications regarding the selection procedure, the constructs measured, and the role of incumbents and managers in developing the procedure may improve understanding and acceptance of a selection procedure.

There are situations where some candidates refuse to participate in certain types of selection procedures. It may be useful to consider whether desirable candidates remove themselves from consideration because of the selection procedure. In addition, recruiters sometimes resist or attempt to circumvent the use of selection procedures because it increases the need for additional candidates. Therefore, researchers should consider approaches designed to minimize any negative perception of a selection procedure.

Alternate Forms

Alternate forms of a selection procedure may be needed to reduce practice effects and enhance security. Alternate forms may allow the organization to continue assessment after a security breach; however, researchers may provide information to organizations to help them balance these advantages with the increased costs for development and validation of alternate forms. If alternate forms (including adaptive tests) are developed, care must be taken to ensure that candidate scores are comparable across forms. If alternate forms are used, establishing the equivalence of scores on the different forms is usually necessary. The statistical procedures used in equating studies typically take into account the size and relevant characteristics of the samples, the use of an anchor test or linking test items, and the feasibility of determining equating functions within subgroups.

Selecting the Validation Strategy

Once researchers have worked with the organization to define its objectives for developing a selection procedure, understand the requirements of the work, and reach agreement on the type of selection procedure, researchers must decide what validation strategy or strategies will be pursued to accumulate evidence to support the intended inference. In addition, the strategy selected must be feasible in the organizational context and meet the project goals within the constraints imposed by the situation.

Fit to Objectives, Constraints, and Selection Procedures

In choosing an initial validation strategy, the researcher should consider the fit of the strategy to the organization's objectives and constraints, as well as its fit to the selection procedures planned and the criterion measures. Three examples are provided below to describe possible ways in which validation strategies may be matched to organizational objectives and constraints. In the first scenario, an organization wanting to assemble validity evidence for a small population position may rely upon a validity generalization strategy because extensive cumulative evidence exists for the predictor-criterion relationship in similar situations. In contrast, another organization facing a similar problem that wants to extend a selection procedure from one business unit to another may use a transportability study to establish the validity of the employee selection procedure in another business unit with the same job. Neither option may be available when a position is unique to the organization. Thus, in the third situation, the organization may rely on a content-based validity strategy.

Individual Assessments

Individual assessment refers to one-on-one evaluations on the basis of a wide range of cognitive and noncognitive measures that are integrated by the assessor, often resulting in a recommendation rather than a selection decision or prediction of a specific level of job performance (Jeanneret & Silzer, 1998). The assessor should have a rationale for the determination and use of the selection procedures. In such instances, the validity of the assessor's clinical judgments is most important to the evaluation of the assessment process. If there are multiple assessors, the consistency of their assessment findings can be valuable to understanding validity and making accurate judgments about the relevant KSAOs. Validation research studies of clinical judgments are clearly an exception rather than the rule (Ryan & Sackett, 1998). However, both validity generalization and content-oriented validation strategies may be appropriate. For example, there may be a wide range of generalizable evidence that has been accumulated by a test publisher or the assessing psychologist demonstrating that a personality scale (e.g., conscientiousness) is predictive of successful managerial performance. Therefore, such a selection procedure would be appropriate for use in an executive assessment protocol. An example of a content-oriented validation approach would be demonstrating the relationship of an in-basket selection procedure that measures planning capability to the planning requirements of an executive position.

Selecting Criterion Measures

When the source of validity evidence is based on the relationships between predictor scores and other variables (criteria), the nature of the criteria is determined by the proposed uses of the selection procedures and outcomes from the analysis of work and worker requirements. Professional judgment should be exercised in selecting the most appropriate criteria given known organizational constraints and climate.

Performance-Oriented Criteria

Criteria that are representative of work activities, behaviors, or outcomes usually focus on the job performance of incumbents. Supervisory ratings are the most frequently used criteria, and often they are designed specifically for use in the research study. Other performance information also may be useful (e.g., training program scores, sales, error rates, and productivity indices). Consideration should be given to psychometric factors for all criteria whenever feasible.

Other Indices

Depending on the objective of the validation effort, indices other than those directly related to task performance may be most appropriate. Exam-

ples include absenteeism, turnover, and other organizational citizenship behaviors. Again, the researcher should be cautious about deficiencies or contaminating factors in such indices.

Relevance and Psychometric Considerations

Criteria are typically expected to represent some construct (often work performance), and the quality of that representation should be established. For example, the fidelity of a work sample used as a criterion should be documented on the basis of the work analysis. Supervisory ratings should be defined and scaled in terms of relevant work activities or situations. All criteria should be representative of important work behaviors, outcomes, or relevant organizational expectations regarding individual employee behavior or team performance.

Criteria should be reliable, and the determination of that reliability may be influenced by the study parameters and organizational constraints. For example, while it may be desirable to have two raters independently evaluate the performance of an employee to determine the inter-rater reliability of the ratings, the work situation and supervisory relationships may preclude such an effort (e.g., there may not be two supervisors knowledgeable about an employee's work). In any circumstance, the researcher should determine what reliability estimates are calculated, how they are obtained, and what levels of reliability are acceptable.

Data Collection

Collection of both predictor and criterion data in a validation study requires careful planning and organizing to ensure complete and accurate data. The standardized conditions under which the validation research is conducted are normally replicated to the extent possible during actual use of the selection procedure. In order to collect accurate and complete information, the test user should consider the following activities.

Communications

Relevant information about the data collection effort should be communicated to all those affected by the effort including management, those who take the test for research purposes, persons who provide criterion data, and those who will use the test. Appropriate communications will facilitate the data collection and encourage all involved to provide accurate and complete information. The kind of information shared depends on the needs of the organization and the individuals involved. For example, participants in the validation research will want to know how their test results will be used and who will have access to the results. Supervisors who provide criterion ratings and others who provide archival criterion data will want to know the logistics of data collection, ultimate use, and provisions for confidentiality.

End users, such as the staffing organization or the client organization employing individuals who were screened with the selection procedures, should have an overview of the study. When feasible, anticipated uses of job analysis, test, and criterion data should be shared with those who generated it.

Pilot Testing

The researcher should determine the extent to which pilot testing is necessary or useful. Previous experience with specific selection procedures may reduce or eliminate this need. Availability of test takers and opportunities to conduct pilot testing may be influenced by various organizational constraints.

Match Between Data Collection and Implementation Expectations

Selection procedures should be administered in the same way that they will be administered in actual use. For example, if interviewers are provided face-to-face training in the validation study, similar training should be provided in actual use. Instructions and answers to candidate questions should be as similar as possible during validation and implementation.

Confidentiality

Confidentiality is an ethical responsibility of the researcher. It is also a major concern to all those involved in the research. Those who provide this information, performance ratings, or content validity linkages may be more willing to provide accurate information if they are assured of the confidentiality of their individual contributions. Participants in validation research studies should be given confidentiality unless there are persuasive reasons to proceed otherwise.

The researcher should carefully decide what level of anonymity or confidentiality can be established and maintain it throughout the study. The researcher provides the maximum confidentiality feasible in the collection and storage of data, recognizing that identifying information of some type is often required to link data collected at different times or by different individuals. Web-based data collection presents additional confidentiality challenges.

Quality Control and Security

The test user should employ data collection techniques that are designed to enhance the accuracy and security of the data and test materials. Public disclosure of the content and scoring of most selection procedures should be recognized as a potentially serious threat to their reliability, validity, and subsequent use. All data should be retained at a level of security that permits access only for those with a need to know.

Data Analyses

A wide variety of data may be collected and analyzed throughout the course of a validation study. The responsibilities and supervision of the people who conduct data analyses should be commensurate with their capabilities and relevant experience.

Data Accuracy

All data should be checked for accuracy. Checks for data accuracy typically include verification that scores are within the possible ranges and that no apparent falsification of responses occurred.

Missing Data/Outliers

Often, one or more data points are missing, and/or outliers exist in the data set. Because these circumstances are typically unique to the validation effort underway, establishing hard-and-fast rules is not possible. Instead, the researcher must examine each situation on its own merits and follow a strategy based on professional judgment. Researchers should document the rationale for treating missing data and/or outliers so their work can be replicated. If imputation techniques are used to estimate missing data, such techniques and their underlying assumptions should be documented clearly.

Descriptive Statistics

Most data analyses will begin with descriptive statistics that present analyses of frequencies, central tendencies, and variances. Such descriptions should be provided for the total group and for relevant subgroups if they are large enough to yield reasonably reliable statistics.

Appropriate Analyses

Data analyses should be appropriate for the method or strategy undertaken. Data are frequently collected as part of the analysis of work and during the validation effort itself. Data analytic methods used also should be appropriate for the nature of the data (e.g., nominal, ordinal, interval, ratio), sample sizes, and other considerations that will lead to correct inferences regarding the data sets.

Differential Prediction

Organizations vary in their goals, and competing interests within the organization are not unusual. Efforts to reduce differences for one subgroup may increase differences for another. Given the difficulty in reconciling different interests in the case of substantial over- or underprediction, researchers oftentimes consider the effects of the prediction errors and their relationship to organizational goals.

A finding of predictive bias does not necessarily prevent the operational use of a selection procedure. For example, if the study is based upon an extremely large sample, a finding of a small but statistically significant differential prediction may have little practical effect. In general, the finding of concern would be evidence of substantial underprediction of performance in the subgroup of interest. Such a finding would generally preclude operational use of the predictor and would likely lead to additional research and considerations of modifying or replacing the selection procedure.

Absent a finding of substantial underprediction, a reasonable course of action for some organizations would be to recommend uniform operational use of the predictor for all groups. However, a large amount of overprediction may also lead to a consideration of alternate selection procedures.

Combining Selection Procedures Into an Assessment Battery

The researcher must exercise professional judgment regarding the outcomes of the validation effort to determine those predictors that should be included in the final selection procedure and the method of combination (including predictor weighting) that will meet the goals of the organization. The algorithm for combining the selection procedures and the rationale for the algorithm should be described. When combining predictors, the validity of the inferences resulting from the composite is of primary importance.

Multiple Hurdles Versus Compensatory Models

Taking into account the purpose of the assessment and the outcomes of the validity study, the researcher must decide whether candidates are required to score above a specific level on each of several assessments (multiple hurdles) or achieve a specific total score across all assessments (compensatory model). There are no absolutes regarding which model should be implemented, and at times a hurdle may be most appropriate for one predictor, while a compensatory model may be best for other predictors within the overall selection procedure. The rationale and supporting evidence should be presented for the model recommended for assessment scoring and interpretation. Researchers should be aware that the method of combining test scores might affect the overall reliability of the entire selection process and the subgroup passing rates (Sackett & Roth, 1996).

Cutoff Scores Versus Rank Orders

Two frequently implemented selection decision strategies are selection based on a cutoff score or selection of candidates in a top-down order. There is no single method for establishing cutoff scores. If based on valid predictors demonstrating linearity or monotonicity throughout the range of prediction, cutoff scores may be set as high or as low as needed to meet the requirements of the organization. Additionally, given monotonicity, selecting the

top scorers in top-down order maximizes estimated performance on the criterion measure if there is an appropriate amount of variance in the predictor. Where there is an indication of nonmonotonicity, this finding should be taken into consideration in determining how to use selection procedure scores.

Given the unitary concept of validity and the underlying premise (based on empirical evidence) that inferences regarding predictors of a cognitive nature and performance criteria are linear (Coward & Sackett, 1990), cognitively based selection techniques developed by content-oriented procedures and differentiating adequately within the range of interest can usually be assumed to have a linear relationship to job behavior. Such content-oriented procedures support rank ordering and setting the cutoff score as high or as low as necessary. Research has not yet established whether this same set of premises holds true for other types of predictors (e.g., personality inventories, interest inventories, indices of values).

Professional judgment is necessary in setting any cutoff score and typically is based on a rationale that may include such factors as estimated cost-benefit ratio, number of vacancies and selection ratio, expectancy of success versus failure, the consequences of failure on the job, performance and diversity goals of the organization, or judgments as to the knowledge, skill, ability, or other characteristics required by the work. When cutoff scores are used as a basis for rejecting applicants, the researcher should document their rationale or justification.

Cutoff scores are different from critical scores in that a cutoff score defines a point on a selection procedure score distribution below which candidates are rejected, while a critical score defines a specified point on a distribution of selection procedure scores above which candidates are considered successful. Critical scores are criterion referenced and may be useful for implementing a certain type of selection procedure (e.g., certification exam), but are not appropriate when no absolute minimum on the selection procedure score distribution can be discerned (e.g., cognitive ability or aptitude test).

When researchers make recommendations concerning the use of a rank ordering method or a cutoff score, the recommendation often takes into account the labor market, the consequences of errors in prediction, the level of a KSAO represented by a chosen cutoff score, the utility of the selection procedure, resources needed to monitor and maintain a list of qualified candidates, and other relevant factors. The goals of the organization may favor a particular alternative. For example, some organizations decide to use a cutoff score rather than rank ordering to increase workforce diversity, recognizing that a reduction also may occur in job performance and utility. Whatever the decision, the researcher should document the rationale for it.

Bands

Bands are ranges of selection procedure scores in which candidates are treated alike. The implementation of a banding procedure makes use of cut scores, and there are a variety of methods for determining bands (Cascio, Outtz, Zedeck, & Goldstein, 1991; Campion et al., 2001).

Bands may be created for a variety of administrative or organizational purposes; they also may be formed to take into account the imprecision of selection procedure scores and their inferences. However, because bands group candidates who have different selection procedure scores, predictions of expected criterion outcomes are less precise. Thus, banding will necessarily yield lower expected criterion outcomes and selection utility (with regard to the criterion outcomes predicted by the selection procedure) than will top-down, rank order selection. On the other hand, the lowered expected criterion outcomes and selection utility may be balanced by benefits such as administrative ease and the possibility of increased workforce diversity, depending on how within-band selection decisions are made. If a banding procedure is implemented, the basis for its development and the decision rules to be followed in its administration should be clearly documented.

Norms

Normative information relevant to the applicant pool and the incumbent population should be presented when appropriate. The normative group should be described in terms of its relevant demographic and occupational characteristics and presented for subgroups with adequate sample sizes. The time frame in which the normative results were established should be stated.

Communicating the Effectiveness of Selection Procedures

Two potentially effective methods for communicating the effectiveness of selection procedures are expectancy analyses and utility estimates.

Expectancies and practical value. Expectancy charts may assist in understanding the relationship between a selection procedure score and work performance. Further, information in the Taylor-Russell Tables (Taylor & Russell, 1939) identifies what proportions of hired candidates will be successful under different combinations of test validity (expressed as correlation coefficients), selection ratios, and percentages of current employees that are satisfactory performers.

Utility. Projected productivity gains or utility estimates for each employee and the organization due to use of the selection procedure may be relevant in assessing its practical value. Utility estimates also may be used to compare the relative value of alternative selection procedures. The literature regarding the impact of selection tests on employee productivity has provided several means to estimate utility (Brogden, 1949; Cascio, 2000**;** Cronbach & Gleser, 1965; Hunter, Schmidt, & Judiesch, 1990; Naylor & Shine,

1965; Raju, Burke, & Normand, 1990**;** Schmidt, Hunter, McKenzie, & Muldrow, 1979). Some of these express utility in terms of reductions in some outcome of interest (e.g., reduction in accidents, reduction in person-hours needed to accomplish a body of work). Others express utility in dollar terms, with the dollar value obtained via a regression equation incorporating a number of parameters, such as the increment in validity over current practices and the dollar value of a standard deviation of performance. Still others express utility in terms of percentage increases in output due to improved selection. The values for terms in these models are often estimated with some uncertainty, and thus the result is a projection of gains to be realized if all of the model assumptions hold true. Often researchers do not conduct follow-up studies to determine whether projected gains are, in fact, realized. Under such circumstances, the results of utility analyses should be identified as estimates based on a set of assumptions, and minimal and maximal point estimates of utility should be presented when appropriate to reflect the uncertainty in estimating various parameters in the utility model.

Appropriate Use of Selection Procedures

Inferences from selection procedure scores are validated for use in a prescribed manner for specific purposes. To the extent that a use deviates from either the prescribed procedures or the intended purpose, the inference of validity for the selection procedure is likely to be affected.

Combining Selection Procedures

Personnel decisions are often made on the basis of information from a combination of selection procedures. The individual components as well as the combination should be based upon evidence of validity. Changes in the components or the mix of components typically require the accumulation of additional evidence to support the validity of inferences for the altered procedure. When a multiple hurdle approach is employed, the original validation data may be relied on for those components that remain intact. However, the effectiveness of the selection procedure as a whole may be reduced as a result of the introduction of a predictor of unknown quality.

When a compensatory approach is used, the addition or deletion of a selection procedure component can fundamentally change the inferences that may be supported. Under these circumstances, the original validation evidence may not be sufficient when there are alterations to the selection procedures that are not supported by a follow-up validation effort.

Using Selection Procedures for Other Purposes

The selection procedure should be used only for the purposes for which there is validity evidence. For example, diagnostic use of a selection procedure that has not been validated in a way to yield such information should be avoided. Likewise, the use of a selection procedure designed for an educa-

tional environment cannot be justified for the purpose of predicting success in employment settings unless the education tasks and the work performed in the validation research or their underlying requirements are closely related, or unless the relevant research literature supports this generalization.

Recommendations

The recommendations based on the results of a validation effort should be consistent with the objectives of the research, the data analyses performed, and the researcher's professional judgment and ethical responsibilities. The recommended use should be consistent with the procedures used in, and the outcomes from, the validation research including the validity evidence for each selection procedure or composite score and the integration of information from multiple sources. In addition, the researcher typically considers the cost, labor market, and performance expectations of the organization, particularly when choosing a strategy to determine who is selected by the procedure. Tight labor markets may necessitate acceptance of a lower score on the selection procedure. Also, the organization's expectations regarding work force diversity may influence the use of test information for that organization.

Technical Validation Report

Reports of validation efforts should include enough detail to enable a researcher competent in personnel selection to know what was done, to draw independent conclusions in evaluating the research, and to replicate the study. The reports must accurately portray the findings, as well as the interpretations of and decisions based on the results. Research findings that may qualify the conclusions or the generalizability of results should be reported. The following information should be included:

Identifying Information

The report should include the author(s), dates of the study, and other information that would permit another researcher to understand who conducted the original research.

Statement of Purpose

The purpose of the validation research should be stated in the report.

Analysis of Work

The report should contain a description of the analysis of work, including procedures employed, the participants in the process, data analyses, and results.

Search for Alternative Selection Procedures

The researcher should document any search for selection procedures (including alternate combinations of these procedures) that are substantially equally valid and reduce subgroup differences.

Selection Procedures

Names, editions, and forms of selection procedures purchased from publishers should be provided as well as descriptions and, if appropriate, sample items. When proprietary tests are developed, the researcher should include a description of the items, the construct(s) that are measured, and sample items, if appropriate. Typically, copies of tests or scoring procedures should not be included in technical reports or administration manuals in order to protect the confidentiality of operational items. The rationale for the use of each procedure and basic descriptive statistics, including appropriate reliability estimates for the sample in the research study when feasible, also should be included. If raters are an integral part of the selection procedure (as in some work samples), the reliability of their ratings should be determined and documented.

Relationship to Work Requirements

The report should provide a description of the methods used by the researcher to determine that the selection procedure is significantly related to a criterion measure or representative of a job content domain. Establishing the relationship of a selection procedure to job content is particularly important when conducting a job content validation study.

Criterion Measures (When Applicable)

A description of the criterion measures, the rationale for their use, the data collection procedures, and a discussion of their relevance, reliability, possible deficiency, freedom from contamination, and freedom from or control of bias should be provided in detail.

Research Sample

The sampling procedure and the characteristics of the research sample relative to the interpretation of the results should be described. The description should include a definition of the population that the sample is designed to represent, sampling biases that may detract from the representativeness of the sample, and the significance of any deviations from representativeness for the interpretation of the results. Data regarding restriction in the range of scores on predictors or criterion measures are especially important. When a transportability study is conducted to support the use of a particular selection procedure, the relationship between the original validation research sample and the population for which the use of the selection procedure is proposed should be included in the technical report.

Results

All summary statistics that relate to the conclusions drawn by the researcher and the recommendations for use should be included. Tables should present complete data, not just significant or positive results. The sample size, means, standard deviations, and intercorrelations of variables measured and other information useful to the interpretation of the results should be presented and clearly labeled. Both uncorrected and corrected values should be presented when corrections are made for statistical artifacts such as restriction of range or unreliability of the criterion.

Scoring and Transformation of Raw Scores

Methods used to score items and tasks should be fully described. When performance tasks, work samples, or other methods requiring some element of judgment are used, a description of the type of rater training conducted and scoring criteria should be provided.

Derived scales used for reporting scores and their rationale should be described in detail in the research report or administration manual. Whether using derived scores or locally produced labels (such as "qualified," "marginal," or "unqualified"), the researcher should clearly describe the logical and psychometric foundations.

Normative Information

Parameters for normative data provide researchers and users with information that guides relevant interpretations. Such parameters often include demographic and occupational characteristics of the normative sample, time frame of the data collection, and status of test takers (e.g., applicants, incumbents, college students).

When normative information is presented, it should include measures of central tendency and variability and should clearly describe the normative data (e.g., percentiles, standard scores). Norm tables usually report the percent passing at specific scores and may be useful in determining the effects of a cutoff score. Expectancy tables indicate the proportion of a specific sample of candidates who reach a specified level of success and are often used to guide implementation decisions.

Recommendations

The recommendations for implementation and the rationale supporting them (e.g., the use of rank ordering, score bands, or cutoff scores, and the means of combining information in making personnel decisions) should be provided. Because implementation rules like those applied to cutoff scores sometimes do change, subsequent modifications should be documented and placed in an addendum to the research report or administration manual.

Caution Regarding Interpretations

Research reports and/or administration manuals should help readers make appropriate interpretations of data and should warn them against common misuses of information.

References

There should be complete references for all published literature and available technical reports cited in the report. Technical reports completed for private organizations are often considered proprietary and confidential, and the researcher cannot violate the limitations imposed by the organization. Consequently, some technical reports that may have been used by the researcher are not generally available.

Administration Guide

The term "test administrator" refers to those individuals responsible for day-to-day activities such as scheduling testing sessions, administering the selection procedure, scoring the procedure, maintaining the databases, and reporting scores or results.

An administration guide should document completely the information needed to administer the selection procedure, score it, and interpret the score. When the selection procedure is computer-based or in a format other than paper-and-pencil, the administration guide should also include detailed instructions on the special conditions of administration. While this document is sometimes a part of a technical report, it is often separate so that confidential information in the validity study is protected and administrators are provided with only the information needed to administer the selection procedure. In other situations, the test user will write parts of the administration guide since the researcher may not know the organization's specific policies or the details of its implementation strategies. In deciding whether two separate documents are needed, the researcher should consider access to each document, the sensitivity of information to be included, the purpose of each document, and the intended audience.

Administration guides developed by a publisher are often supplemented with addenda that cover local decisions made by the user organization. Consequently, not all the information listed below will be found in every administration guide from a publisher. However, the researcher in the user organization should try to provide answers or guidance for the issues raised.

The information developed for users or examinees should be accurate and complete for its purposes and should not be misleading. Communications regarding selection procedures should be stated as clearly and accurately as possible so that readers know how to carry out their responsibilities competently. The writing style of all informational material should be writ-

ten to meet the needs of the likely audience. Normally, the following information should be included in an administration guide.

Introduction and Overview

This section of the report should inform the reader of the purpose of the assessment procedure and provide an overview of the research that supports the use of the procedure. The introduction should explain why the organization uses formal, validated selection procedures, the benefits of professionally developed selection procedures, the importance of security, and the organization's expectations regarding the consistency of their use. Care must be taken in preparing such documents to avoid giving the reader an impression that an assessment program is more useful than is really the case.

Contact Information

The administration guide should provide information about whom to contact if there are questions or unanticipated problems associated with the selection procedure.

Selection Procedures

The selection procedure should be thoroughly described. Names, editions, and forms of published procedures as well as information for ordering materials and ensuring their security should be provided. Although entire tests are not usually included in administration guides for security reasons, sample items are helpful. When proprietary tests are developed, the researcher should include a description of the items, the construct(s) that are measured, and sample items.

Applicability

The description of the selection procedure should indicate to whom the procedure is applicable (e.g., candidates for "x" job) and state any exceptions (e.g., exemptions for job incumbents) to test requirements. It may also be useful to indicate that use of a selection procedure is based on one or more validation efforts that focused on specific jobs/job requirements and that these efforts define the boundaries of any test applications. If the organization has rules about when tests are administered, these rules must be clearly stated in the administration guide used by the organization. For example, some organizations only administer a selection procedure when there is a job vacancy. Other organizations may administer selection procedures periodically in order to build pools of qualified candidates.

Administrators

The administration guide should state the necessary qualifications of administrators and the training required to administer selection procedures in general, as well as the specific selection procedure described in the administration guide. Administrators should receive training in the administration of selection procedures. Administrators must understand that failure to follow standardized protocol may render the research results and operational scores irrelevant to some degree. The researcher must be both insistent and persuasive to gain understanding with regard to both the nature of and the need for standardized administration of tests or other procedures. Periodic retraining may be needed to reinforce the administration rules. Observational checks or other quality control mechanisms should be built into the test administration system to ensure accurate and consistent administration.

Information Provided to Candidates

Many organizations use test brochures or test orientation materials to inform candidates about the employee selection process. Some organizations also provide informational sessions prior to the administration of a selection procedure. When appropriate, the researcher should consider providing candidates with clearly written, uniform information about the selection procedure such as the purpose, administrative procedures, completion strategies, time management, feedback, confidentiality, process for requesting accommodation for disability, and other relevant user policies. Whenever possible, both the content and the process for orienting candidates should be standardized. The administration guide should describe these materials and provide information on how the administrator may obtain them. The rules for distribution should be explicitly stated in order to facilitate consistent treatment of candidates.

Guidelines for Administration of Selection Procedures

The researcher should use the administration guide as an opportunity to convey the organization's requirements for selection procedure administration. In addition to detailed instructions regarding the actual administration of the selection procedure, the administration guide often includes rules and tips for providing an appropriate testing environment as well as ensuring the candidate's identity.

For security reasons, the identity of all candidates should be confirmed prior to administration. Administrators should monitor the administration to control possible disruptions, protect the security of test materials, and prevent collaborative efforts by candidates. The security provisions, like other aspects of the *Principles,* apply equally to computer and Internet-administered sessions.

Administration Environment

There are a number of factors that potentially affect test administration: appropriate workspace, adequate lighting, and a quiet, comfortable setting, free of distractions. The researcher should consider these conditions and their potential effects on test performance. At a minimum, selection procedure administration should be in an environment that is responsive to candidates' concerns about the selection procedures and maintains their dignity.

Scoring Instructions and Interpretation Guidelines

The researcher should provide the selection procedure administrator or user with details on how the selection procedure is to be scored and how results should be interpreted. The administration guide should help readers make appropriate interpretations of data and warn them against common misuses of information.

Processes that will ensure accuracy in scoring, checking, and recording results should be used. This principle applies to the researcher and to any agent to whom this responsibility has been delegated. The responsibility cannot be abrogated by purchasing services from an outside scoring service. Quality control checks should be implemented to ensure accurate scoring and recording.

Instructions for scoring by the user should be presented in the administration guide in detail to reduce clerical errors in scoring and to increase the reliability of any judgments required. Distinctions among constructs should be described to support the accuracy of scoring judgments. Scoring keys should not be included in technical reports or administration manuals and should be made available only to persons who score or scale responses.

If Computer-Based Test Interpretation (CBTI) is used to process responses to a selection procedure, the researcher should provide detailed instructions on how the CBTI is to be used in decision making.

Test Score Databases

Organizations should decide what records of assessment administrations and scores are to be maintained and should provide detailed information (or reference detailed information) regarding record keeping and databases. In addition, policies on the retention of records (e.g., duration, security, accessibility, etc.) and the use of archival data over time should be established and communicated. Raw scores should be kept because data reported in derived scales may limit further research. Databases should be maintained for sufficient time periods to support periodic audits of the selection process.

Reporting and Using Selection Procedure Scores

The researcher must communicate how selection procedure scores are to be reported and used. Results should be reported in language likely to be

interpreted correctly by persons who receive them. The administration guide should also indicate who has access to selection procedure scores.

Administrators should be cautioned about using selection procedure information for uses other than those intended. For example, although selection procedure data may have some validity in determining later retention decisions, more potentially relevant measures such as performance ratings may be available. Furthermore, if the pattern of selection procedure scores is used to make differential job assignments, evidence is required demonstrating that the scores are linked to, or predictive of, different performance levels in the various assignments of job groupings.

Candidate Feedback

In addition to reporting selection procedure scores to others within the organization, the researcher should include information on how to provide feedback to candidates, if such feedback is feasible or appropriate. Feedback should be provided in clear language that is understandable by candidates receiving the feedback, and should not violate the security of the test or its scoring.

Nonstandard Administrations (See Also Candidates With Disabilities)

The administration guide should cover nonstandard selection procedure administrations. Such administrations encompass not only accommodated selection procedure sessions, but also sessions that were disrupted (e.g., power failures, local emergency, and illness of a candidate), involved errors (e.g., questions and answer sheet did not match, timing mistake), or were nonstandard in some other way.

The administration guide should establish a clear process to document and explain any modifications of selection procedures, disruptions in administration, or any other deviation from established procedures in the administration, scoring, or handling of scores. While it is impossible to predict all possible occurrences, the researcher should communicate general principles for how deviations from normal procedures are to be handled.

Reassessing Candidates

Generally, employers should provide opportunities for reassessment and reconsidering candidates whenever technically and administratively feasible. In some situations, as in one-time examinations, reassessment may not be a viable option. In order to facilitate consistency of treatment, the administration guide should clearly explain whether candidates may be reassessed and how reassessment will take place. In some organizations, specific time intervals must elapse. In others, although difficult to evaluate, significant developmental activities must have occurred prior to reassessment.

Corrective Reassessment

Users in conjunction with researchers should consider when corrective reassessment is appropriate. Critical errors on the part of the administrator (e.g., timing mistakes, use of nonmatching selection procedure booklet and answer sheet) and extraordinary disturbances (e.g., fire alarm, acutely ill assessee) are usually justifications for reassessment. The administration guide should cover procedures and guidelines for granting corrective reassessment and documenting all requests.

Security of the Selection Procedure

Selection procedure items that are widely known (through study, coaching, or other means) in an organization are usually less effective in differentiating among candidates on relevant constructs. Maintenance of test security therefore limits the type and amount of feedback provided to candidates. The more detail on candidate responses provided, the greater the security risk. The administration guide should emphasize the importance of safeguarding the content, scoring, and validity of the selection procedure.

Selection procedures usually represent a significant investment on the part of the organization for development and validation. The administration guide should point out the value of the selection procedure itself and the cost of compromised selection procedures in terms of the additional research required and the possibility of a less capable candidate being hired.

Practices for ensuring the security of selection procedure documents (e.g., numbering test booklets and maintaining records of the numbers; keeping used and unused selection procedures in a secure, locked facility; collecting scratch paper after administration sessions) and selection procedure scoring should be communicated.

Selection procedure scores must be kept secure and should be released only to those who have a need to know and who are qualified to interpret them. Special practices may be required to protect confidential materials and selection procedure information that exist in electronic forms. Although security practices may be difficult to apply in the case of employment interviews, the importance of security as a means of preserving their standardization and validity should be considered. Organizations are encouraged to develop policies that specify the length of time that confidential information is retained. When confidential information is destroyed, the user should consider ways of maintaining its security such as having selection personnel supervise the destruction of the documents.

References

When other useful documents are mentioned, they should be referenced fully. When the documents are internal publications, the means of acquiring those documents should be described.

Other Circumstances Regarding the Validation Effort and Use of Selection Procedures

Influence of Changes in Organizational Demands

Because organizations and their work forces are dynamic in nature, changes in organizational functioning may occur and subsequent selection procedure modifications may be necessary. Changing work requirements may lead to adjustments in cutoff scores or the introduction of a new assessment; both would require further study of the existing selection procedure. If advised of such circumstances, the researcher should examine each situation on its own merits and make recommendations regarding the impact of organizational change on the validation and use of any selection procedure.

Review of Validation and Need for Updating the Validation Effort

Researchers should develop strategies to anticipate that the validity of inferences for a selection procedure used in a particular situation may change over time. Such changes may occur because of changes in the work itself, worker requirements, or work setting. Users (either on their own or with researcher assistance) of a selection procedure should periodically review the operational use of the assessment instrument and the data at hand (including timeliness of normative data if appropriate) to determine if additional research is needed to support the continued use of the selection procedure. When needed, the research should be brought up to date and reported. The technical or administration guides should be revised (or an addendum added) if changes in research data or use of procedures make any statement or instruction incorrect or misleading.

Candidates With Disabilities

Assessing candidates with disabilities may require special accommodations that deviate from standardized procedures. Accommodations are made to minimize the impact of a known disability that is not relevant to the construct being assessed. For example, an individual's upper extremity motor impairment may affect a score on a measure of cognitive ability although the motor impairment is not related to the individual's cognitive ability. Accommodations may include, but are not limited to, modifications to the environment (e.g., high desks), medium (e.g., Braille, reader), time limit, or content. Combinations of accommodations may be required to make valid inferences regarding the candidate's ability on the construct(s) of interest.

Professional judgment is required on the part of the user and the developer regarding the type or types of accommodations that have the least negative impact on the validity of the inferences made from the selection procedure scores. Empirical research is usually lacking on the effect of given

accommodations on selection procedure performance for candidates with different disabilities or varying magnitudes of the same disability.

Responsibilities of the Selection Procedure Developers, Researchers, and Users

Researchers and individuals charged with approving the accommodation for an organization should be knowledgeable about the availability of modified forms of the selection procedure, psychometric theory, and the likely effect of the disability on selection procedure performance. Users may choose to modify the original selection procedure, develop a modified procedure for candidates with disabilities, or waive the selection procedure altogether and use other information regarding the candidate's job-related knowledge, skills, abilities or other characteristics. While empirical research to demonstrate comparability between the original procedure and the modified procedure may not be feasible in most instances, the individuals developing the modifications should make attempts when possible to limit the modifications, consistent with legal responsibility, to those that allow, insofar as is possible, the comparability of procedures.

Development and validation. Although most employers have too few cases for extensive research, the principles set forth in this document in the preparation of modified selection procedures for candidates with disabilities should be followed to the extent possible. Modified procedures should be pilot-tested with candidates whose disabilities resemble those of the target population when possible and feasible. Practical limitations such as small sample size often restrict the ability of the researcher to statistically equate modified versions of the selection procedure to the original form. These considerations also limit efforts to establish the reliability of the scores and the validity of the inferences made from these scores. Nevertheless, the reliability of selection procedure scores and the validity of inferences based on these scores should be determined whenever possible. In the rare case when it is possible, the effects of administration of the original form of the selection procedure to candidates with disabilities also should be explored.

Documentation and communications regarding accommodations. Descriptions of the modifications made, the psychometric characteristics of the modified selection procedures, and the performance of candidates with disabilities on the original form of the procedure, if available, should be included in the documentation. In addition, selection procedure users should always document the modifications and conditions under which the procedure was administered and keep that information secure and separate from the assessment data in the organization's records. Legal considerations may prohibit giving decision makers information on whether a candidate's score was earned with a selection procedure accommodation and the nature of the modification. However, users may designate those scores earned with an accommodation in such a way to permit special handling in data analysis.

Selection procedure modification.   The test user should take steps to ensure that a candidate's score on the selection procedure accurately reflects the candidate's ability rather than construct-irrelevant disabilities.   One of these steps is a dialog with the candidate with the disability about the accommodations possible.   In some cases, the construct being assessed cannot be differentiated from the disability (e.g., proofreading test taken by a sight-impaired candidate).   Other times, the disability does not affect selection procedure performance and no accommodation is necessary.   Components of a selection procedure battery should be considered separately in determinations of modifications.   To the extent possible, standardized features of administration should be retained in order to maximize comparability among scores.   Approval of prespecified, routine accommodations not expected to affect the psychometric interpretation of the selection procedure scores (e.g., adjusting table height) may be delegated to administrators.

Maintaining consistency with assessment use in the organization.   The selection procedures used when assessing candidates with disabilities should resemble as closely as possible the selection procedures used for other candidates.   The selection procedures are developed for the purpose of making selection decisions, not for the purpose of assessing the extent of a candidate's disability.   The addition of a procedure designed to assess the existence or degree of a disability is inappropriate as a selection tool.