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 scientific literature and the testing professional'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. 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 testing professional 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 affect them.

Defining the organization’s needs, objectives, and constraints.

Testing professionals use their expertise and experience to assist the organization in refining its goals and objectives. Different units of the organization may have different and sometimes competing and conflicting objectives. For instance, one unit may prefer rigorous selection standards even though they create hardships for another unit responsible for recruiting qualified applicants.

Organizations often consider costs (price, time, administrative effort) when choosing among selection procedures. These costs should be weighed against the benefits of the proposed selection system through a cost-benefit analysis.

The testing professional is encouraged to work with all units (e.g., human resources, internal or outsourced recruiting, labor relations, legal, compliance, information technology) that may have an effect on or be affected by the selection procedure and with other relevant stakeholders (e.g., internal or external individuals and groups such as labor organizations, work councils, advocacy groups, customers). The testing professional provides accurate information regarding the benefits and limitations of various strategies in meeting the organization’s goals based on past experience and the relevant body of scientific research. In all situations, the testing professional and the organization’s representatives should factor in the desires of the various stakeholders and determine the relative levels of consideration to be given to each point of view.

Climate and culture. Testing professionals 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. Testing professionals must consider the attitudes and commitments of organization leaders and employees who are faced with intense competition, mergers, stakeholder demands, and other corporate events that may influence the perceived
relative importance of selection research. Testing professionals also may need to take into
account the legal and labor environment when deciding on validation approaches or selection
instruments. In addition, many HR functions are inter-related with actions in one area affecting
other areas. For example, changes in the selection standards often impact the level and extent
of training required. Global selection systems should also take into consideration locally accepted
practices and the organization’s ability to execute the selection procedure accurately and reliably,
regardless of location.

**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, synthetic validation, content-oriented strategy) as well as affect the
feasibility and method for pilot testing procedures.

Even when the number of workers is sufficient to conduct a local validation study, their
availability and willingness to participate in a 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,
or workers who have experienced organizational restructuring may be skeptical about the purpose
of the research and its effect on their own positions. Careful consideration should also be given
to the timing of the data collection for the validation study. For example, attempting to collect
assessment data or manager ratings during the unit’s busy season or when the organization is
downsizing can affect the quality of the data.

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 base cutoff scores on the performance of incumbents may find that those cutoff
scores are too high, and thus inappropriate, if the organization’s workforce is more capable than
the applicant pool. 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 validation and implementation
efforts may include, but are not limited to, the workers themselves, managers, supervisors,
trainers, customers, archival records, business performance metrics, and research conducted
internal and external to the organization (including meta-analyses and sources such as O*NET).
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
testing professional’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 desire selection procedures that
are predictive of important outcomes, easy and quick to administer, 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 legally 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 (e.g., childhood experiences) and
personality inventory items may be viewed as an invasion of privacy, even if they can be shown
to be related conceptually and empirically to the criterion measures or the requirements of the
job. Cultures also differ in the acceptability of different kinds of selection procedures, so
candidates’ willingness to complete the assessment should be taken into consideration.
Some organizations prefer selection procedures that provide information regarding the strengths and developmental needs of the test taker. In such cases, procedures that measure knowledge or content that can be learned (e.g., software) 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 an effective selection procedure. Nevertheless, the testing professional 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 implementation of the selection procedure. The testing professional must use professional judgment in determining the appropriate information to provide and the communication format and style that will be most effective. Testing professionals encourage organizations to consider the effects of participation in the validation effort on employees, managers, and business/organizational units. For example, organizations typically decide that data from a concurrent validation or selection system development study will be kept confidential and not be used for subsequent employment-related decisions. Organizations may also limit the number of performance ratings a manager is asked to make to minimize the demands on the manager's time and promote high quality ratings.

Understanding Work and Worker Requirements

In many businesses and industries, the nature of work changes rapidly. Factors such as changes in technology, mission, security context, strategy, organizational structure, the applicant pool, or customer demands result in substantive and frequent changes in work behaviors and requirements. A new work analysis should be conducted when test developers or users have reason to believe that the nature of the work performed has changed meaningfully since any prior analysis was conducted.

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 in which the importance or relevance of a criterion indicator or 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 jobs and 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 a variety of factors, including the number of workers, their work locations, their demographic characteristics, their performance-related characteristics (e.g., amount of experience, training, proficiency), shift or other work
cycles, and other variables that might influence the work analysis. Inclusion of a broad sample of incumbents (or other SMEs) is likely to increase the representativeness of the results.

**Documentation of the results.**

The methodology, data collection methods, analyses, results, and implications of the work analysis 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. See Technical Validation Report section for more information about documenting results.

**Selecting Assessment Procedures for the Validation Effort**

The testing professional 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. The result of this step is often the test plan, which describes the predictor constructs that will be measured with each assessment procedure. An example of this might be a KSAO-by-test matrix, often included in a technical report.

**Review of research literature and the organization’s objectives.**

Testing professionals should become familiar with not only the organization’s objectives for the selection system but also research relevant to the constructs to be measured. The research literature can be 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 should be considered for each instrument. Some of the more important psychometric considerations include reliability, evidence supporting the validity of the intended inferences, and differences among subgroups.

**Scoring considerations.**

The testing professional must ensure that administration and scoring tasks can be completed accurately and consistently across candidates and locations. For all testing modalities, regardless of format, medium, or platform, test professionals should ensure that scoring rubrics are standardized, reliable, and appropriate in order to allow test users to make score-based inferences consistent with the content and intent of the assessment.

**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 rater-evaluated individual responses.

Medium refers to the method of delivery of the selection procedure content. For example, a
A measure of cognitive ability could be presented via paper-and-pencil, computer, video, or orally. There are advantages and disadvantages in selecting a particular medium. For example, computer-administered procedures may reduce the demands on administrators and enhance standardization.

Testing professionals may choose to use multiple media in test administration; however, changing the medium may affect the construct being measured and threaten the equivalency of scores across media. For example, converting a paper-and-pencil situational judgment test to a video in which the situations will be acted out will reduce the reading component of the test. Also, administering speeded tests of cognitive ability on computer rather than paper-and-pencil may alter the construct being measured (Mead & Drasgow, 1993).

A number of considerations are important when evaluating different 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, breadth of recruiting, and accessibility of testing opportunities are all important considerations. Organizational decision-makers may find that unproctored internet-based assessments allow for developing a larger applicant pool. An alternative is a proctored assessment approach, which may be more costly and require applicants to travel to a test site, although it more readily allows for standardization of measurement, verification of applicant identity, and verification of applicant performance. Professional judgment must be used in determining the appropriate medium for test administration in light of the organization’s goals.

In addition to understanding that scores from the same test delivered via different media or using different response formats may be non-comparable, developers of selection systems should be cognizant that format and medium can affect mean score differences among subgroups (Hough, Oswald, & Ployhart, 2001). Over time, assessment systems may demonstrate changes in scores or even validity if test material becomes compromised.

Acceptability to the candidate.

In addition to the organization’s needs and objectives, testing professionals should also 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, length of the process, the modality of the online assessment, 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 too closely resemble the work may be perceived as obsolete when the work changes and may assess KSAOs that are not needed at entry because they are learned during on-the-job training. 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 some, despite the demonstrated validity. 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 times when 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 factors in the selection process. In addition, recruiters sometimes resist or attempt to circumvent the use of selection procedures because it increases the need for additional candidates. Testing professionals should consider approaches designed to minimize negative perceptions of a selection procedure.
Alternate forms.

Alternate forms of a selection procedure (including item banks for adaptive tests and/or unproctored tests) may be needed to reduce practice effects and enhance security. Alternate forms may allow the organization to avoid security breaches and continue assessment after a security breach. Testing professionals can provide information to organizations to help them balance these advantages with the increased costs for development and validation of alternate forms. If alternate forms 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. Monitoring score distribution qualities across multiple test forms for parallel structure is important.

Selecting the Validation Strategy

Once testing professionals 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(s), testing professionals must decide what validation strategy or strategies will be pursued to accumulate evidence to support the intended inference(s). Clearly, the strategy selected must be feasible in the organizational context, and it must meet the project goals within the constraints imposed by the situation.

Fit to objectives, constraints, and selection procedures.

In choosing a validation strategy, the testing professional 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 a second scenario, another organization wanting 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. In a third scenario, neither option may be available when a position is unique to the organization, and in this case, the organization may use 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 (Silzer & Jeanneret, 2011). 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 or
competencies. Validation research studies of clinical judgments are clearly an exception rather than the rule (Church & Rotolo, 2011; Kwaske, 2008; Ryan & Sackett, 1998; Silzer & Jeanneret, 2011). However, both validity generalization and content-oriented validation strategies may be appropriate when designing an individual assessment strategy. 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 (e.g., Morris, Daisley, Wheeler, & Boyer, 2015) and would, therefore, 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 the relationships between predictor scores and other variables (criteria), the nature of those criteria is determined by the outcomes from the work analysis, including worker requirements (e.g., KSAs or competency model) and proposed uses of the selection procedures. Professional judgment should be exercised in selecting 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 as opposed to operational performance management measures used for administrative purposes. Other performance information may also be useful (e.g., training program scores, sales, error rates, customer ratings, and productivity indices). Attention to avoiding bias against demographic groups is important when selecting criteria, and consideration should be given to psychometric characteristics of 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 appropriate. Examples include absenteeism, turnover, and other organizational citizenship behaviors. The testing professional should be cautious about deficiencies or contaminating factors in all indices.

**Relevance and psychometric considerations.**

Criteria are typically expected to represent some organizationally relevant construct (e.g., work performance, citizenship behavior, counterproductive behavior), 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.
Although criteria should demonstrate adequate levels of reliability, calculation of an appropriate reliability estimate may be influenced by the data available for the study and organizational constraints. For example, one can typically calculate some form of criterion reliability estimate in any local validation study. However, depending upon the inferences the testing professional desires to make regarding the criterion scores, the reliability estimate that the testing professional is able to calculate (based on the data on hand) may not appropriately reflect the types of measurement error of interest. When reporting criterion reliability, the testing professional should describe the type of reliability estimate and sources of error that are reflected in (and ignored by) the reliability index.

Data Collection

The 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, those who provide criterion data, those who will use the test, and other appropriate stakeholders. 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, who will have access to the results, and how security of test and criterion data will be maintained over time. Supervisors who provide criterion ratings and others who provide archival criterion data will want to know the logistics of data collection, ultimate use, provisions for confidentiality, and data security protections. 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 work analysis, test, and criterion data should be shared with those who generated it. Periodic updates to stakeholders on project status, go-live dates, responsibilities, and process as well as a final briefing on the project results are recommended.

Pilot testing.

The testing professional should determine the extent to which pilot testing is necessary or useful to ensure that data collection will go smoothly. 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 during the validation research 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 testing professional. It is also a major concern to all those involved in the research. Those who provide 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 testing professional should carefully decide what level of anonymity or confidentiality can be established, communicate it to participants, and maintain it thereafter. The testing professional 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 stored in different databases, collected at different times, or collected by different methods. Online data collection presents additional confidentiality challenges, such as insuring the security of the data collected.

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.

Issues of quality control and test security become particularly salient in unproctored internet testing (UIT) or remotely proctored internet testing (RPIT) environments. In these contexts, mechanisms and procedures should be adopted that diminish the chances of the assessment content being compromised, reduce the opportunity for cheating on the assessment, and help ensure positive identification of the individual completing the assessment. Test users considering the use of UIT or RPIT should be familiar with the advantages and disadvantages of these assessment options, as well as emerging best practices in these areas (e.g., Industrial and Organizational Psychology, Volume 2 Issue 1, 2009; International Test Commission. (2006). International guidelines on computer-based testing and Internet-delivered testing. International Journal of Testing. 6, 143-172. ).

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 and management.

As part of the data collection process, measures and procedures should be included to facilitate later analyses of the quality of data provided by validation study participants. For example, the testing professional should consider including content or mechanisms to help identify careless or insufficient effort responding (Huang, Curran, Keeney, Poposki, & DeShon, 2012; Meade & Craig, 2012). Raters of job performance should be asked about factors that could influence their ability to provide quality performance ratings, such as their level of familiarity with the ratee’s performance, opportunities to observe the ratee’s performance, and length of supervision of the ratee. All decision rules used when preparing the data for analyses should be clearly documented and appropriately justified.

Although becoming less common given advances in technology, a double-entry process should be considered to help ensure accurate entry when data are manually entered. Regardless of whether data are manually entered or captured through technology, values for all variables in
the resulting data set should be checked for out-of-range values or, in the case of a technology-
enabled data collection, missing data that may be indicative of a technology glitch. Data should
also be checked for logical inconsistencies that often arise when extracting data for a validation
study from multiple sources, for example, checking that demographic information (e.g., sex, race,
age, tenure) obtained from archival and self-report data match. Clear decision rules for handling
any inconsistencies should be documented.

If archival data are included in the validation study, extra precautions should be taken prior to
analyzing such data. Issues involving data privacy, data integrity, and consistency of variable
naming and definitions over time are all critical factors to consider.

**Missing data and outliers.**

Often, one or more data points are missing and/or outliers exist in the data set. The testing
professional must examine each situation on its own merits and follow a strategy for handling
missing data and/or outliers based on professional judgment informed by best practices cited
in the literature on handling missing data and outliers.

When analyzing data collected for validation studies, two commonly recommended strategies
are full information maximum likelihood (FIML) and multiple imputation (MI) approaches (Enders,
2010). Default options for handling missing data in common statistical packages (i.e., listwise and
pairwise deletion, mean imputation) are often poor choices (Wilkinson & APA’s Taskforce on
Statistical Inference, 1999; see Enders, 2010; Graham, 2009; Little & Rubin, 2002; and Newman,
2014 for methods for dealing with missing data.) When there are missing data, the testing
professional should provide (a) a summary of missing data patterns and the nature of the
missingness (e.g., missing at random, missing completely at random, missing not at random) and
(b) justification for the missing data technique adopted for analyses.

Testing professionals should also check their data for both univariate and multivariate
outliers (Aguinis, Gottfredson, & Joo, 2013). Documentation should include how outliers were
defined and identified. If clear outliers are found, sensitivity analyses should be performed to
evaluate the effects of including and excluding outliers on the validation study results, or robust
estimation/analysis techniques should be used that account for the presence of outliers. Orr,
Sackett, and DuBois (1991) report that most testing professionals oppose dropping outliers
unless there is evidence that the data point is erroneous. Dropping outliers to obtain more
favorable results is not appropriate.

**Descriptive statistics.**

Most data analyses will begin with descriptive statistics for predictor and criterion variables
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 estimates.

**Appropriate analyses.**

Data analyses should be appropriate for the method or strategy undertaken. Data are
frequently collected as part of the analysis of work during the pilot or field testing of
predictor/criterion measures and during the validation effort itself. Data analytic methods used
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. For example, the presence of
non-independence (clustering of individuals) in the predictor-criterion data being analyzed can
affect the accuracy/quality of inferences and should be considered. (For a review of non-
independence issues and their potential effects on evaluating predictor-criterion relations, see
Differential prediction.

Organizations vary in their goals, and competing interests within an 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, testing professionals 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 for all groups.

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 substantial amount of overprediction may also lead to a consideration of dropping the predictor for all groups and/or investigating alternate selection procedures for all groups.

Combining selection procedures into a selection system.

As noted earlier, the testing professional must exercise professional judgment regarding the outcomes of the overall selection system to determine those predictors that should be included in the final selection system and the method of combination and sequencing that will meet the goals of the organization (cf. Combining predictors and combining criteria). The methods used for combining and sequencing predictors should be clearly documented and justified. When combining predictors to form an overall score or make an overall decision, organizations may have different goals and values. For example, some organizations may put more emphasis on maximizing validity relative to minimizing subgroup differences. In contrast, other organizations may put more emphasis on minimizing subgroup differences relative to maximizing validity or may desire striking a balance between maximizing validity and minimizing subgroup differences.

Multiple hurdles versus compensatory models.

Taking into account the purpose of the assessment and the outcomes of the validity study, the testing professional 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, hybrid approaches are possible (e.g., 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. Testing professionals should be aware that the method of combining test scores might influence 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 the use of (a) a cutoff score and (b) rank order/top-down selection. A cutoff score defines the point on a selection procedure score
distribution below which candidates are rejected. There is no single method for establishing cutoff scores; several potentially viable options exist (Mueller, Norris, & Oppler, 2007). For example, cutoff scores may be criterion referenced when the predictor score can be linked to a meaningful performance threshold. 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. When there is an indication of nonmonotonicity in predictor-criterion relationships, this finding should be taken into consideration in determining how to use those scores for making personnel decisions (e.g., Converse & Oswald, 2014).

When data are not locally available to evaluate linearity and monotonicity in predictor-criterion relations, testing professionals should consider findings from past research and their implications for proposed use of scores. For example, though research evidence suggests relations between measures of cognitive ability and job performance are linear (e.g., Arneson, Sackett, & Beatty, 2011; Coward & Sackett, 1990), findings with regard to linearity of relations between other types of predictors (e.g., personality measures) and job performance have been mixed, and is an open area of research (e.g., Carter, Dalal, Boyce, O’Connell, Kung, & Delgado, 2014).

Beyond the factors noted above, professional judgment is necessary when setting any cutoff score, and when deciding between use of cutoff scores, top-down selection, or score bands (addressed in the next section). These decisions are typically driven by the goals of the organization and may be based on factors such as the estimated cost-benefit ratio, the number of vacancies and the selection ratio, the labor market, expectancy of success versus failure, the consequences of failure on the job, other consequences of selection decision errors, the relative emphasis on the performance and diversity goals of the organization, judgments as to the level a KSAs/competency or performance required by the work, and the utility of the selection procedure. Whatever the decision, the testing professional should document the rationale for it.

When evaluating or recommending cutoff scores for selection procedures, it may be useful to consider conditional standard errors for the selection measure/composite on which the cutoff scores are being set in the vicinity of those cutoff scores. Documentation should indicate the model used to compute the conditional standard errors (Brennan, 1998; Qualls-Payne, 1992; Raju, Price, Oshima, & Nering, 2007). One might also provide estimates of the percentage of applicants who would be classified in the same way (i.e., pass/fail) on two or more replications of the selection procedure at the given cutoff score (Haertel, 2006).

**Bands.**

Bands are ranges of selection procedure scores in which candidates are treated alike. The implementation of a banding procedure makes use of cutoff scores (i.e., to delineate predictor score ranges that define the bands), and there are a variety of methods for determining bands (Campion et al., 2001; Cascio, Outtz, Zedeck, & Goldstein, 1991). 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 generally 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. 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 utility estimates 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, testing professionals 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 compensatory approach is used, the addition or deletion of a selection procedure component can fundamentally change the inferences that might be supported. Under these circumstances, the original validation evidence might not be sufficient to support the altered selection procedure.

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 educational 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 testing professional’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 testing professional typically considers the cost, labor market, effects on protected groups as well as workforce diversity, and performance expectations of the organization, particularly when choosing a strategy to determine who is selected by the procedure.

Technical Validation Report

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

Identifying information.

The report should include the authors, their credentials, their affiliations, dates of the study, and other information that would permit another testing professional 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, the characteristics of the participants in the process, any judgments made by SMEs, instructions that were provided to participants in data collection efforts for their specific tasks, and data analyses and results including reliability/precision. If any of the documents used in the analysis of work were translated, then a description of the translation and adaptation procedures should be included.

Search for alternative selection procedures.

The report should document any search for selection procedures (including alternate combinations of the procedures) that show equal or greater validity for the given selection situation with an accompanying reduction in subgroup differences.

Development of selection procedures.

Names, editions, and forms of selection procedures purchased from publishers should be provided as well as technical descriptions and, if appropriate, sample item content. When proprietary selection tools are developed, the testing professional should include a description of the content, including the construct(s) measured by the content, and the process by which the content was developed, if appropriate. Typically, content and scoring algorithms 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 statistical procedure and results of relevant analyses performed should be included. If raters are an integral part of the selection procedure, as in some work samples, then the reliability and agreement of their ratings should be determined and documented.

Establishing validity.

The report should provide a description of the validation studies conducted such that another testing professional could reproduce the analyses and results. The report should also include methods used by the testing professional to determine that the selection procedure is statistically and practically significantly related to a criterion measure and/or representative of a job content domain. Establishing the relationship of a selection procedure to job content and KSAOs is particularly important when conducting a job content validation study, both to justify the use of the selection procedure and to provide substantive support for its validity.

Criterion validation studies, when conducted, should report the following in detail: a description of the criterion measures; the rationale for their use; the data collection procedures; and a discussion of the measures’ relevance, reliability, possible deficiencies, possible sources of contamination, and freedom from or control of biasing sources of variance. If the testing professional developed the criterion measure, then the report should include the rationale and steps taken to develop it, so it can be well understood and, if needed, replicated in future validation studies.
Research sample.

The sampling procedure and the characteristics of the research sample relative to the appropriate 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, the significance of any deviations from representativeness for the interpretation of the results, and any statistical power analysis results. Data informing the potential 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. Test developers should make clear whether psychometrics in the technical report refer to candidates or incumbents, and results for concurrent validation studies should not be represented as the results for predictive validation studies.

Results.

All summary statistics that relate to the conclusions drawn by the testing professional and the recommendations for use should be included. Complete statistical results related to the development and validation, not just statistically significant or supportive 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 and algorithms used to score content should be fully described. For example, when weighted scores, derived scales, or composite or categorical scores are used, rationale should be provided in detail. 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.

Normative information.

Parameters for normative data provide testing professionals 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., candidates, incumbents, students). When normative information is presented, it should include measures of central tendency and variability (and skewness when appropriate), and it should clearly describe the normative data (e.g., percentiles, standard scores). Normative 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 of selection procedures and the rationale supporting the recommendations (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. Some implementation rules may change over time (e.g., those applied to cutoff scores), and, subsequent modifications should be documented and placed in an addendum to the research
report or administration manual.

Caution regarding interpretations.

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

Technology enabled selection procedures.

If the selection procedure is technology enabled, the researcher should document the technology requirements and any technology-based accommodations that can be provided by the administrator for test takers with disabilities.

References.

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

Administration Information

Individuals with test administration responsibilities include those 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. Those who have responsibilities related to the technology supporting administration, such as programming scoring algorithms, maintaining web interfaces for testing and score reporting, and ensuring that updated or expanded test content is incorporated into the existing testing system, should be considered as part of a test administration team. The accuracy of their work is the responsibility of the test administration lead.

Those with day-to-day administration responsibilities should be aware of any personal limitations (physical, perceptual, cognitive) that might affect their ability to administer and/or score a test fairly and accurately, and they should not administer assessments when they cannot meet the demands of their roles or if there are barriers to their effective delivery of responsibilities.

Complete documentation should be available with regard to administering the selection procedure, scoring it, and interpreting the score, regardless of the mode of assessment delivery (e.g., paper-and-pencil, computerized, Internet/web based). Although this documentation is sometimes a part of a technical report, it is often separate so that confidential information in the validation study is protected and administrators are provided with only the information necessary and appropriate to administer the selection procedure. In other situations, the test user in the organization will develop some of the administration information and procedures because the testing professional may not know the organization’s specific policies or the details of its implementation strategies. In deciding whether separate documents are needed, the testing professional should consider who has access to each document, the sensitivity of the information to be included, the purpose of each document, and the intended audiences.

Administration information developed by a publisher is often supplemented with addenda that cover local decisions made by the user organization. Consequently, not all the information listed below will be found in administration documentation from a publisher or vendor. However, the testing professional in the user organization should try to provide answers or guidance for the issues raised.
The information developed for users or examinees should be clear, accurate, and complete for its purposes. Communications regarding selection procedures should be stated as clearly and accurately as possible so that readers know how to carry out administrative responsibilities competently. The writing style of all informational material should be appropriate to address the understanding and needs of the likely audience. When a test is to be administered in multiple countries and in multiple languages, documentation and supporting materials required for administration may need to undergo appropriate translation procedures. Normally, the following information should be included as administration documentation:

**Introduction and overview.**

This section of the documentation should inform the reader of the purpose of the assessment procedure and provide an overview of the empirical 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 assessment security, and the degree of consistency required in administration. Care must be taken in preparing such documents to avoid giving the reader an impression that an assessment program is more useful or applicable than is really the case.

**Contact information.**

The administration documentation should provide information about whom to contact in case questions or unanticipated problems associated with the selection procedure arise.

**Selection procedures.**

The selection procedures 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 documentation for security reasons, providing sample items that represent all relevant aspects of the test can be very helpful. When proprietary tests are developed, the testing professional 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., job candidates for a specified job) and state any exceptions to test requirements (e.g., exemptions for job incumbents). Information on applicability to testing individuals with disabilities and individuals from different cultural and linguistic groups should be included. If the organization has rules about when tests are administered, these rules must be clearly stated in the administration documentation used by the organization. For example, some organizations only administer a selection procedure when there is a job vacancy, while other organizations may administer selection procedures periodically in order to build pools of qualified candidates.
Administration responsibilities.

The administration documentation should state the necessary qualifications of those with different administrative responsibilities (e.g., for maintaining a scoring algorithm, for handling retesting requests) and the training required to administer selection procedures in general, as well as training for the specific selection procedure of interest. Training should emphasize that failures in following the standardized protocols may render any research results and the meaning and interpretation of operational scores irrelevant to some degree. Testing professionals should document the nature of and the need for standardized administration of tests or other procedures. Periodic training may be needed to maintain understanding and compliance to the administration rules, especially when the people who are involved in administration change. Observational checks or other quality control mechanisms should be built into the test administration system to ensure accurate and consistent administration. Pass rates or mean scores of the assessment should be reviewed periodically to look for spikes, which may indicate the scoring key has been compromised.

Information provided to candidates.

Many organizations provide information to candidates about the employee selection process (via brochures, web pages, emails, videos), and such information should be clear, pertinent, and timely. Depending on the test, the population of test takers, and the circumstances, the administrator should consider what information about the selection procedure to provide candidates. For example, information about the intended test use, administrative procedures, test format and interface, test completion strategies (e.g., opportunity to go back and change item responses), time parameters, feedback and access to scores (e.g., who will have access and how long data will be retained), confidentiality protections and conditions under which records may be released, processes for requesting accommodation for disability, warnings about improper candidate behavior and responsibility to respect copyright laws, retesting policies, and other relevant user policies as appropriate might be provided. Administrators might also convey whether and how test takers may review and correct their personal information, as well as how to appeal when test scores are cancelled or withheld (as in credential and licensure test settings), or when allegations of misconduct occur. Regardless of what information is provided to candidates, it should be clear and consistent. Both the content and the process for orienting candidates should be standardized whenever possible. The administration documentation should describe these materials and indicate how they are provided to candidates (e.g., via open website or email). The rules for distribution should be explicitly stated in order to facilitate consistent treatment of candidates.

Guidelines for administration of selection procedures.

The testing professional should use the administration documentation 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 documentation may include rules and tips for providing an appropriate testing environment as well as ensuring the candidate’s identity. Some technology-enabled tests may require that test takers receive instruction and practice prior to administration. Test administrators are responsible for ensuring that any such instruction and practice are provided. When the test taker is responsible for his/her own testing environment (e.g., unproctored internet testing), the administrator still has the responsibility of informing the test taker of environmental factors likely to affect performance and of the characteristics of an appropriate testing environment. Further, those with test administration responsibilities also are responsible for informing test takers of any instructions
regarding security (e.g., identification verification, setting up web cams, verification codes) and
the consequences for the test taker of not following test security procedures. Appeals processes
when testing irregularities have been detected should be conveyed.

Reasonable effort should be made to ensure the integrity of test scores (e.g., verifying
identities of test takers). When appropriate, test administrators should be trained on how to take
precautions against cheating, how to detect and prevent opportunities to cheat, and how to
monitor and detect cheating as it occurs. Those who use technologies designed to detect
irregularities (e.g., particular answer patterns or erasure patterns, plagiarism) are responsible for
their appropriate use. Administrators should monitor the administration to control possible
disruptions, protect the security of test materials, and prevent collaborative efforts by candidates.
Although older versions of tests are sometimes made available by the test user for practice
purposes, in general, tests should not be made available to the public or resold to unqualified test
users. 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. Examples include (but
are not limited to) an appropriate workspace; adequate lighting; a quiet, comfortable setting, free
of distractions; and the extent to which the test is technology enabled and the corresponding
effects of requirements such as browser, monitor size, and touch screen. The testing professional
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’
corresponds about the selection procedures and maintains their dignity. When effects of the
environment on test performance are known, test takers should be informed which specific test-
taking conditions may have consequences (e.g., potential lowered performance). Administrators
should inform test takers on the general environmental conditions conducive for test taking when
individuals are responsible for their own testing environments (e.g., unproctored internet testing).

**Scoring instructions and interpretation guidelines.**

Testing professionals should provide the selection procedure administrators or users with
details on how the selection procedure is to be scored and how results should be interpreted.
Note that the documentation provided in commercially available test manuals may not provide
sufficient or complete documentation with regard to the proper application of the selection
procedure. Administration documentation should provide objective information regarding any role
of the test administrator in the intended interpretation of test scores, the positive and negative
consequences of test use, and protecting the security of test content and the privacy of test takers.
The administration documentation should therefore help readers make appropriate
interpretations of scores and related information and warn them against common misuses.

Processes should be followed to ensure accuracy in scoring, checking, and recording
results. This principle applies to the testing professional and to any agent to whom this
responsibility has been delegated. The responsibility cannot be ignored or substituted by
purchasing services from an outside scoring service. Quality control checks and routine
monitoring should be implemented to ensure accurate scoring and recording. Procedures for
rescoring of tests when mistakes are suspected should be clear to administrators.

Instructions for scoring by the user should be presented in the administration documentation
in detail to reduce clerical errors in scoring and to increase the reliability of any required judgments.
Distinctions among measured 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 scoring and interpretation procedures (e.g., automated feedback reports) are used to process responses to a selection procedure and generate reports, the testing professional should provide detailed instructions on how they are to be used in decision making. When relevant to the interpretation of test scores, the conditions under which the test was administered (e.g., unproctored setting, accommodated test conditions) should be shared with the test user.

**Test score databases.**

Organizations should decide what records of assessment administrations and scores are to be maintained and should provide detailed information regarding record keeping and databases (or should reference that detailed information). In addition, policies on the retention of records (e.g., duration, security, accessibility) and the use of archival data over time should be established and communicated as appropriate. Raw item data and scores should be retained, 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.**

Documentation provided by the testing professional 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 documentation 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 assignments to jobs or job groupings, evidence is required to support those assignments, such as by demonstrating that the scores are linked to, or predictive of, different performance levels across those jobs or job groupings.

**Candidate feedback.**

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

**Nonstandard administrations (see also Candidates with Disabilities).**

The administration documentation 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. Note that in some cases reporting of a nonstandard administration may be left to the test taker (e.g., a dropped internet connection), and whenever such reporting is invoked, the reporting and procedures for doing so must be clearly explained to the test taker.

The administration documentation should establish a clear process to document and explain any changes to selection procedures, disruptions in administration, or any other deviation
from established procedures in the administration, scoring, or handling of scores. Although it is impossible to predict all possible occurrences, the testing professional 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. To facilitate consistency of candidate treatment, the administration documentation should clearly explain whether candidates may be reassessed and how reassessment will take place. In some organizations, specific time intervals must elapse before reassessment occurs. In other organizations, significant developmental activities must have occurred prior to reassessment.

Corrective reassessment.

Users in conjunction with testing professionals 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 test taker) usually justify reassessment. The administration documentation should cover procedures and guidelines for granting corrective reassessment and documenting all requests. When test takers are remote from any administrative personnel (e.g., unproctored internet testing), it is important that the test taker be informed of the conditions under which he/she might ask for a corrective reassessment, the required timing of such requests, the documentation required, and the procedure for doing so.

Security of the selection procedure.

Selection procedure content that is widely known to job candidates in an organization (through study, coaching, internet resources, or other means) is usually less effective in differentiating among job candidates on relevant constructs. Maintenance of test security therefore is required, which necessarily limits the type and amount of test feedback provided to candidates. The more detail on candidate responses that is provided, the greater the security risk. The administration documentation should emphasize the importance of safeguarding the content, scoring, and validity of the selection procedure as well as monitoring for overexposure of the content.

Selection procedures usually represent a significant investment on the part of the organization for development and validation. The administration documentation 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 and risk of less capable candidates being hired.

It is important to communicate, exercise, and enforce practices that protect the security of selection procedure documents (e.g., verification codes for test access, rotation of content) and the security of selection procedure scoring. Procedures for the security of testing administrator training materials and previous test editions should be documented.

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. International laws regarding data privacy change often and should be consulted in making these determinations. 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 content, standardization,
and validity should be considered. Organizations are encouraged to develop policies that specify the length of time confidential information is to be 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.

When other 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 workforces are dynamic in nature, changes in organizational functioning may occur, and subsequent selection procedure modifications may be necessary. Changing work requirements may lead to the introduction of a new assessment or adjustments in cutoff scores for existing ones; both would require further study of the existing selection procedure. If advised of such circumstances, the testing professional should examine each situation on its own merits and make recommendations to the organization 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.

Testing professionals 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, or the emergence of new jobs. Users of a selection procedure (either on their own or with testing professional assistance) should periodically review the operational use of the assessment instrument using the available data (including timeliness of normative data if appropriate) to determine whether 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. There is also a possible need for evidence that score interpretations continue to be appropriate when there is a change in test format, mode of administration, instructions, or language used in administering a test; the greater the changes, the more likely the need. The technical or administration documentation should be revised accordingly (or an addendum added) if changes in research data or use of procedures make any statement or instruction incorrect or misleading.

Testing Accommodations

Assessing candidates with disabilities may require special accommodations that deviate from standardized procedures in order to remove construct-irrelevant barriers that otherwise interfere with test takers’ ability to demonstrate their standing on job-relevant constructs. 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 lower a score on a measure of cognitive ability because of the candidate’s difficulty taking the test, even though 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), the testing medium (e.g., Braille, text reader), the testing time limit, or test content. Combinations of accommodations may be required to make valid inferences regarding the candidate’s standing on the construct(s) of interest. Procedures for implementing accommodations should be as standardized as possible, including rules for determining who is eligible for an accommodation and how the accommodation is to be administered. Test users should document these procedures and are responsible for monitoring their appropriate implementation. Test takers should be informed of the process and requirements for obtaining any needed accommodation and the confidentiality provisions regarding their disability status.

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 with varying magnitudes of the same disability. Note that a test may be modified so that it no longer assesses the same construct but still provides useful information; if a test is modified, such information should be documented. For example, an individual with dyscalculia may need a calculator for certain items on a broader mathematics problem-solving assessment, rendering the test modified but still useful for indicating something about the individual's skills. If a test no longer assesses the same construct as the original, these test scores can no longer be directly compared with scores from the unmodified test. When score reports are made, it is appropriate to indicate deviation from standard administration procedures and discuss how such deviation may affect results and interpretation to the extent permitted by law.

Responsibilities of the selection procedure developers, testing professionals, and users related to accommodation.

Testing professionals and individuals charged with approving accommodations should be knowledgeable about the availability of accommodated forms of the selection procedure, psychometric theory, and the likely effect of the disability on selection procedure performance. In many employee selection contexts, empirical research to demonstrate comparability between the original procedure and the altered procedure will not be feasible. When changes mean the test no longer assesses the same construct, this is considered a modification of the test itself. Users may choose to alter the original selection procedure, develop an altered procedure for candidates with disabilities, or waive the selection procedure altogether and use other information regarding the candidate’s job-related KSAs or competencies. Implications of these latter changes should be considered seriously, because they create potential challenges in terms of standardization, fairness, job relevance, and other issues.

Selection procedure accommodation and 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 possible accommodations. In some cases, the construct cannot be assessed without reasonably accommodating the disability. Other times, the disability does not affect performance on the selection procedure, and, therefore, no accommodation is necessary. Components of a selection procedure battery should be considered separately when determining appropriate accommodations. To the extent possible, standardized features of administration should be retained in order to maximize comparability among scores. Approval of pre-specified, commonly used accommodations that are irrelevant to selection procedure scores and their psychometric interpretation (e.g., adjusting table height) may be delegated to administrators.

Development and validation. Although most employers have too few cases of accommodated tests for extensive research, the principles set forth in this document in the preparation of altered selection procedures for candidates with disabilities should be followed to the extent possible. Altered procedures should be pilot-tested with candidates whose disabilities resemble those of the target population when possible and feasible; at the very least, this provides practical experience in ensuring the altered procedure can be made operational and run smoothly. Practical limitations, such as small sample size, often restrict the ability of the testing professional to statistically equate data from accommodated versions of the selection procedure to data from the original form, thereby challenging the strict comparability of scores. These considerations also limit efforts to establish the reliability of the accommodated scores and the validity of the inferences made from these scores. Nevertheless, the reliability of accommodated 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 and appropriate, the effects of administration of the original form of the selection procedure to candidates with disabilities also should be examined.
**Documentation and communications regarding accommodations and modifications.** Descriptions of the changes made, the psychometric characteristics of the accommodated or modified selection procedures, and, when sufficient volume of test takers makes it feasible, statistics summarizing the performance of candidates with disabilities on the accommodated or modified forms of the procedure, and the original forms if available should be included in the documentation. Legal considerations may prohibit giving decision-makers information on whether a candidate's score was earned under a selection procedure accommodation and the nature of the accommodation. However, test users may designate those scores earned with an accommodation in such a way to permit special handling in data analysis.

**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. To be clear, selection procedures are developed for the purpose of making selection decisions, not for the purpose of assessing the existence or 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.

**Candidate Linguistic and Cultural Background**

The language in which an assessment is administered may depend upon the job requirements. When appropriate, test administrators should inform test takers of linguistic options.