

adoption and implementation of voluntary affirmative action programs by users who have no obligation under Federal law to adopt them; but are not intended to impose any new obligations in that regard. The agencies issuing and endorsing these guidelines endorse for all private employers and reaffirm for all governmental employers the Equal Employment Opportunity Coordinating Council's "Policy Statement on Affirmative Action Programs for State and Local Government Agencies" (41 FR 38814, September 13, 1976). That policy statement is attached hereto as appendix, section 17.

TECHNICAL STANDARDS

§ 1607.14 Technical standards for validity studies.

The following minimum standards, as applicable, should be met in conducting a validity study. Nothing in these guidelines is intended to preclude the development and use of other professionally acceptable techniques with respect to validation of selection procedures. Where it is not technically feasible for a user to conduct a validity study, the user has the obligation otherwise to comply with these guidelines. See sections 6 and 7 above.

A. *Validity studies should be based on review of information about the job.* Any validity study should be based upon a review of information about the job for which the selection procedure is to be used. The review should include a job analysis except as provided in section 14B(3) below with respect to criterion-related validity. Any method of job analysis may be used if it provides the information required for the specific validation strategy used.

B. *Technical standards for criterion-related validity studies*—(1) *Technical feasibility.* Users choosing to validate a selection procedure by a criterion-related validity strategy should determine whether it is technically feasible (as defined in section 16) to conduct such a study in the particular employment context. The determination of the number of persons necessary to permit the conduct of a meaningful criterion-related study should be made by the user on the basis of all relevant information concerning the selection proce-

dures, the potential sample and the employment situation. Where appropriate, jobs with substantially the same major work behaviors may be grouped together for validity studies, in order to obtain an adequate sample. These guidelines do not require a user to hire or promote persons for the purpose of making it possible to conduct a criterion-related study.

(2) *Analysis of the job.* There should be a review of job information to determine measures of work behavior(s) or performance that are relevant to the job or group of jobs in question. These measures or criteria are relevant to the extent that they represent critical or important job duties, work behaviors or work outcomes as developed from the review of job information. The possibility of bias should be considered both in selection of the criterion measures and their application. In view of the possibility of bias in subjective evaluations, supervisory rating techniques and instructions to raters should be carefully developed. All criterion measures and the methods for gathering data need to be examined for freedom from factors which would unfairly alter scores of members of any group. The relevance of criteria and their freedom from bias are of particular concern when there are significant differences in measures of job performance for different groups.

(3) *Criterion measures.* Proper safeguards should be taken to insure that scores on selection procedures do not enter into any judgments of employee adequacy that are to be used as criterion measures. Whatever criteria are used should represent important or critical work behavior(s) or work outcomes. Certain criteria may be used without a full job analysis if the user can show the importance of the criteria to the particular employment context. These criteria include but are not limited to production rate, error rate, tardiness, absenteeism, and length of service. A standardized rating of overall work performance may be used where a study of the job shows that it is an appropriate criterion. Where performance in training is used as a criterion, success in training should be properly measured and the relevance of the training should be shown either

through a comparison of the content of the training program with the critical or important work behavior(s) of the job(s), or through a demonstration of the relationship between measures of performance in training and measures of job performance. Measures of relative success in training include but are not limited to instructor evaluations, performance samples, or tests. Criterion measures consisting of paper and pencil tests will be closely reviewed for job relevance.

(4) *Representativeness of the sample.* Whether the study is predictive or concurrent, the sample subjects should insofar as feasible be representative of the candidates normally available in the relevant labor market for the job or group of jobs in question, and should insofar as feasible include the races, sexes, and ethnic groups normally available in the relevant job market. In determining the representativeness of the sample in a concurrent validity study, the user should take into account the extent to which the specific knowledges or skills which are the primary focus of the test are those which employees learn on the job.

Where samples are combined or compared, attention should be given to see that such samples are comparable in terms of the actual job they perform, the length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or that these factors are included in the design of the study and their effects identified.

(5) *Statistical relationships.* The degree of relationship between selection procedure scores and criterion measures should be examined and computed, using professionally acceptable statistical procedures. Generally, a selection procedure is considered related to the criterion, for the purposes of these guidelines, when the relationship between performance on the procedure and performance on the criterion measure is statistically significant at the 0.05 level of significance, which means that it is sufficiently high as to have a probability of no more than one (1) in twenty (20) to have occurred by chance. Absence of a statistically significant relationship between a selection proce-

dures and job performance should not necessarily discourage other investigations of the validity of that selection procedure.

(6) *Operational use of selection procedures.* Users should evaluate each selection procedure to assure that it is appropriate for operational use, including establishment of cutoff scores or rank ordering. Generally, if other factors remain the same, the greater the magnitude of the relationship (e.g., correlation coefficient) between performance on a selection procedure and one or more criteria of performance on the job, and the greater the importance and number of aspects of job performance covered by the criteria, the more likely it is that the procedure will be appropriate for use. Reliance upon a selection procedure which is significantly related to a criterion measure, but which is based upon a study involving a large number of subjects and has a low correlation coefficient will be subject to close review if it has a large adverse impact. Sole reliance upon a single selection instrument which is related to only one of many job duties or aspects of job performance will also be subject to close review. The appropriateness of a selection procedure is best evaluated in each particular situation and there are no minimum correlation coefficients applicable to all employment situations. In determining whether a selection procedure is appropriate for operational use the following considerations should also be taken into account: The degree of adverse impact of the procedure, the availability of other selection procedures of greater or substantially equal validity.

(7) *Overstatement of validity findings.* Users should avoid reliance upon techniques which tend to overestimate validity findings as a result of capitalization on chance unless an appropriate safeguard is taken. Reliance upon a few selection procedures or criteria of successful job performance when many selection procedures or criteria of performance have been studied, or the use of optimal statistical weights for selection procedures computed in one sample, are techniques which tend to inflate validity estimates as a result of chance. Use of a large sample is one safeguard: cross-validation is another.

(8) *Fairness.* This section generally calls for studies of unfairness where technically feasible. The concept of fairness or unfairness of selection procedures is a developing concept. In addition, fairness studies generally require substantial numbers of employees in the job or group of jobs being studied. For these reasons, the Federal enforcement agencies recognize that the obligation to conduct studies of fairness imposed by the guidelines generally will be upon users or groups of users with a large number of persons in a job class, or test developers; and that small users utilizing their own selection procedures will generally not be obligated to conduct such studies because it will be technically infeasible for them to do so.

(a) *Unfairness defined.* When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores.

(b) *Investigation of fairness.* Where a selection procedure results in an adverse impact on a race, sex, or ethnic group identified in accordance with the classifications set forth in section 4 above and that group is a significant factor in the relevant labor market, the user generally should investigate the possible existence of unfairness for that group if it is technically feasible to do so. The greater the severity of the adverse impact on a group, the greater the need to investigate the possible existence of unfairness. Where the weight of evidence from other studies shows that the selection procedure predicts fairly for the group in question and for the same or similar jobs, such evidence may be relied on in connection with the selection procedure at issue.

(c) *General considerations in fairness investigations.* Users conducting a study of fairness should review the A.P.A. Standards regarding investigation of possible bias in testing. An investigation of fairness of a selection procedure depends on both evidence of validity and the manner in which the selection

procedure is to be used in a particular employment context. Fairness of a selection procedure cannot necessarily be specified in advance without investigating these factors. Investigation of fairness of a selection procedure in samples where the range of scores on selection procedures or criterion measures is severely restricted for any subgroup sample (as compared to other subgroup samples) may produce misleading evidence of unfairness. That factor should accordingly be taken into account in conducting such studies and before reliance is placed on the results.

(d) *When unfairness is shown.* If unfairness is demonstrated through a showing that members of a particular group perform better or poorer on the job than their scores on the selection procedure would indicate through comparison with how members of other groups perform, the user may either revise or replace the selection instrument in accordance with these guidelines, or may continue to use the selection instrument operationally with appropriate revisions in its use to assure compatibility between the probability of successful job performance and the probability of being selected.

(e) *Technical feasibility of fairness studies.* In addition to the general conditions needed for technical feasibility for the conduct of a criterion-related study (see section 16, below) an investigation of fairness requires the following:

(i) An adequate sample of persons in each group available for the study to achieve findings of statistical significance. Guidelines do not require a user to hire or promote persons on the basis of group classifications for the purpose of making it possible to conduct a study of fairness; but the user has the obligation otherwise to comply with these guidelines.

(ii) The samples for each group should be comparable in terms of the actual job they perform, length of time on the job where time on the job is likely to affect performance, and other relevant factors likely to affect validity differences; or such factors should be included in the design of the study and their effects identified.

(f) *Continued use of selection procedures when fairness studies not feasible.*

If a study of fairness should otherwise be performed, but is not technically feasible, a selection procedure may be used which has otherwise met the validity standards of these guidelines, unless the technical infeasibility resulted from discriminatory employment practices which are demonstrated by facts other than past failure to conform with requirements for validation of selection procedures. However, when it becomes technically feasible for the user to perform a study of fairness and such a study is otherwise called for, the user should conduct the study of fairness.

C. Technical standards for content validity studies—(1) Appropriateness of content validity studies. Users choosing to validate a selection procedure by a content validity strategy should determine whether it is appropriate to conduct such a study in the particular employment context. A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job. Selection procedures which purport to measure knowledges, skills, or abilities may in certain circumstances be justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined as provided in section 14C(4) below, and if that knowledge, skill, or ability is a necessary prerequisite to successful job performance.

A selection procedure based upon inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs, such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability. Content validity is also not an appropriate strategy when the selection procedure involves knowledges, skills, or abilities which an employee will be expected to learn on the job.

(2) Job analysis for content validity. There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative im-

portance and, if the behavior results in work product(s), an analysis of the work product(s). Any job analysis should focus on the work behavior(s) and the tasks associated with them. If work behavior(s) are not observable, the job analysis should identify and analyze those aspects of the behavior(s) that can be observed and the observed work products. The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

(3) Development of selection procedures. A selection procedure designed to measure the work behavior may be developed specifically from the job and job analysis in question, or may have been previously developed by the user, or by other users or by a test publisher.

(4) Standards for demonstrating content validity. To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job. In the case of a selection procedure measuring a knowledge, skill, or ability, the knowledge, skill, or ability being measured should be operationally defined. In the case of a selection procedure measuring a knowledge, the knowledge being measured should be operationally defined as that body of learned information which is used in and is a necessary prerequisite for observable aspects of work behavior of the job. In the case of skills or abilities, the skill or ability being measured should be operationally defined in terms of observable aspects of work behavior of the job. For any selection procedure measuring a knowledge, skill, or ability the user should show that (a) the selection procedure measures and is a representative sample of that knowledge, skill, or ability; and (b) that knowledge, skill, or ability is used in and is a necessary prerequisite to performance of critical or important work behavior(s). In addition, to be content valid, a selection procedure measuring a skill or ability should either closely approximate an

observable work behavior, or its product should closely approximate an observable work product. If a test purports to sample a work behavior or to provide a sample of a work product, the manner and setting of the selection procedure and its level and complexity should closely approximate the work situation. The closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity. As the content of the selection procedure less resembles a work behavior, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content valid, and the greater the need for other evidence of validity.

(5) *Reliability.* The reliability of selection procedures justified on the basis of content validity should be a matter of concern to the user. Whenever it is feasible, appropriate statistical estimates should be made of the reliability of the selection procedure.

(6) *Prior training or experience.* A requirement for or evaluation of specific prior training or experience based on content validity, including a specification of level or amount of training or experience, should be justified on the basis of the relationship between the content of the training or experience and the content of the job for which the training or experience is to be required or evaluated. The critical consideration is the resemblance between the specific behaviors, products, knowledges, skills, or abilities in the experience or training and the specific behaviors, products, knowledges, skills, or abilities required on the job, whether or not there is close resemblance between the experience or training as a whole and the job as a whole.

(7) *Content validity of training success.* Where a measure of success in a training program is used as a selection procedure and the content of a training program is justified on the basis of content validity, the use should be justified on the relationship between the content of the training program and the content of the job.

(8) *Operational use.* A selection procedure which is supported on the basis of content validity may be used for a job if it represents a critical work behavior (i.e., a behavior which is necessary for performance of the job) or work behaviors which constitute most of the important parts of the job.

(9) *Ranking based on content validity studies.* If a user can show, by a job analysis or otherwise, that a higher score on a content valid selection procedure is likely to result in better job performance, the results may be used to rank persons who score above minimum levels. Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance.

D. *Technical standards for construct validity studies*—(1) *Appropriateness of construct validity studies.* Construct validity is a more complex strategy than either criterion-related or content validity. Construct validation is a relatively new and developing procedure in the employment field, and there is at present a lack of substantial literature extending the concept to employment practices. The user should be aware that the effort to obtain sufficient empirical support for construct validity is both an extensive and arduous effort involving a series of research studies, which include criterion related validity studies and which may include content validity studies. Users choosing to justify use of a selection procedure by this strategy should therefore take particular care to assure that the validity study meets the standards set forth below.

(2) *Job analysis for construct validity studies.* There should be a job analysis. This job analysis should show the work behavior(s) required for successful performance of the job, or the groups of jobs being studied, the critical or important work behavior(s) in the job or group of jobs being studied, and an identification of the construct(s) believed to underlie successful performance of these critical or important work behaviors in the job or jobs in question. Each construct should be named and defined, so as to distinguish

§ 1607.15

29 CFR Ch. XIV (7-1-10 Edition)

it from other constructs. If a group of jobs is being studied the jobs should have in common one or more critical or important work behaviors at a comparable level of complexity.

(3) *Relationship to the job.* A selection procedure should then be identified or developed which measures the construct identified in accord with subparagraph (2) above. The user should show by empirical evidence that the selection procedure is validly related to the construct and that the construct is validly related to the performance of critical or important work behavior(s). The relationship between the construct as measured by the selection procedure and the related work behavior(s) should be supported by empirical evidence from one or more criterion-related studies involving the job or jobs in question which satisfy the provisions of section 14B above.

(4) *Use of construct validity study without new criterion-related evidence—(a) Standards for use.* Until such time as professional literature provides more guidance on the use of construct validity in employment situations, the Federal agencies will accept a claim of construct validity without a criterion-related study which satisfies section 14B above only when the selection procedure has been used elsewhere in a situation in which a criterion-related study has been conducted and the use of a criterion-related validity study in this context meets the standards for transportability of criterion-related validity studies as set forth above in section 7. However, if a study pertains to a number of jobs having common critical or important work behaviors at a comparable level of complexity, and the evidence satisfies subparagraphs 14B (2) and (3) above for those jobs with criterion-related validity evidence for those jobs, the selection procedure may be used for all the jobs to which the study pertains. If construct validity is to be generalized to other jobs or groups of jobs not in the group studied, the Federal enforcement agencies will expect at a minimum additional empirical research evidence meeting the standards of subparagraphs section 14B (2) and (3) above for the additional jobs or groups of jobs.

(b) *Determination of common work behaviors.* In determining whether two or more jobs have one or more work behavior(s) in common, the user should compare the observed work behavior(s) in each of the jobs and should compare the observed work product(s) in each of the jobs. If neither the observed work behavior(s) in each of the jobs nor the observed work product(s) in each of the jobs are the same, the Federal enforcement agencies will presume that the work behavior(s) in each job are different. If the work behaviors are not observable, then evidence of similarity of work products and any other relevant research evidence will be considered in determining whether the work behavior(s) in the two jobs are the same.

DOCUMENTATION OF IMPACT AND VALIDITY EVIDENCE

§ 1607.15 Documentation of impact and validity evidence.

A. *Required information.* Users of selection procedures other than those users complying with section 15A(1) below should maintain and have available for each job information on adverse impact of the selection process for that job and, where it is determined a selection process has an adverse impact, evidence of validity as set forth below.

(1) *Simplified recordkeeping for users with less than 100 employees.* In order to minimize recordkeeping burdens on employers who employ one hundred (100) or fewer employees, and other users not required to file EEO-1, et seq., reports, such users may satisfy the requirements of this section 15 if they maintain and have available records showing, for each year:

(a) The number of persons hired, promoted, and terminated for each job, by sex, and where appropriate by race and national origin;

(b) The number of applicants for hire and promotion by sex and where appropriate by race and national origin; and

(c) The selection procedures utilized (either standardized or not standardized).

These records should be maintained for each race or national origin group (see section 4 above) constituting more