

Buros Center for Testing

Standards for Accreditation of Testing Programs

Improving the Science and Practice of Testing
www.buros.org

Copyright 2017
The Board of Regents of the University of Nebraska
and the Buros Center for Testing

CONTENTS

Introduction.....	1
Section 1: The Testing Program.....	3
Purpose of the Testing Program	
Validity	
Structure and Resources of the Testing Program	
Section 2: Examination Content.....	4
Content Framework and Test Specifications	
Item Development and Selection	
Section 3: Form Development and Review	5
Pilot Testing	
Creation of Final Exam Forms	
Psychometric Review of Operational Tests	
Comparability across Forms, Formats, and Language	
Section 4: Examination Administration.....	7
Eligibility and Application	
Administration Sites	
Test Administrators and Proctors	
Procedures for Administration	
Record Keeping	
Section 5: Fairness and Diversity.....	10
Accommodations	
Fairness for Diverse Groups	
Section 6: Scoring and Reporting.....	11
Scoring and Scaling	
Determining Cut Scores	
Norm-Referenced Interpretation	
Score Reporting	
Section 7: Exam Security and Privacy.....	13
Exam Material Security	
Security and Privacy of Examinee Data	

Buros Center for Testing

Standards for Accreditation of Testing Programs

INTRODUCTION

The Buros Center for Testing is the world's premier institution for the evaluation and review of tests. The Buros Center for Testing *Standards for Accreditation of Testing Programs* are intended for use in accreditation of testing programs that are proprietary or otherwise not commercially available. The *Standards for Accreditation of Testing Programs* provide transparency regarding the process by which Buros evaluates testing programs for accreditation. More information regarding our accreditation process is available under the Psychometric Consulting section of our website (buros.org/psychometric-consulting).

The *Standards for Accreditation of Testing Programs* are highly consistent with the current leading professional standards for testing: the 2014 *Standards for Educational and Psychological Testing*,¹ developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. The current document synthesizes key requirements of the *Standards for Educational and Psychological Testing* in a format intended to facilitate the Buros accreditation process. Although the *Standards for Accreditation of Testing Programs* are not explicitly intended to guide test development per se, testing programs may find these standards useful for such purposes as well.

Testing programs seeking accreditation by Buros may wish to review the *Standards for Accreditation of Testing Programs* in conjunction with their application or in determining whether to seek accreditation. The outcome of the evaluation will be based on a holistic review of the extent to which the testing program demonstrates meeting these standards. While programs should strive to meet every standard, it is not a requirement to receive accreditation. Testing programs completing the accreditation process will receive a report identifying ways in which the program might be improved and an explanation of why accreditation was or was not granted. Providing formative feedback is an important component of our accreditation process.

A wide variety of testing programs may seek accreditation from Buros. Credentialing tests and higher education assessment are the most common applications, but the standards are intended to be used with virtually any type of testing program. For testing programs with certain features, other specific requirements may apply. In such instances, Buros may

¹ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

incorporate additional relevant standards in the evaluation process and will inform clients when additional evaluative criteria apply.

The Buros Center for Testing is perhaps best known for its test reviews of commercially available tests and assessments in the *Mental Measurements Yearbook* series. Publishers of commercially available tests and assessments are invited to submit their products for review consideration. More information regarding reviews of commercially available tests is available in the Test Reviews and Information section of our website. Although the *Standards for Accreditation of Testing Programs* are not explicitly intended as standards for *Mental Measurements Yearbook* reviews, test publishers, reviewers, or readers may find it to be a useful resource, along with the *Standards for Educational and Psychological Testing*.

SECTION 1: THE TESTING PROGRAM

Purpose of the Testing Program

1.1 The testing program should clearly define the purpose of the testing program, the constructs to be assessed, and the population for which the test is appropriate.

1.2 The testing program should explicitly outline how test scores and other test results should be used and interpreted, as well as describe limitations of these interpretations.

1.3 The testing program should provide potential test users with a description of anticipated unintended or inappropriate test score uses and interpretations.

1.4 If a credential, certificate, or similar recognition is awarded based on exam performance, the scope and limitations of such recognitions should be made clear to relevant parties.

Validity

1.5 The defined purpose of the test scores should be supported by a documented validity framework. The validity framework should provide multiple sources of validity evidence in support of each intended use and interpretation of scores and should be integrated to provide a coherent justification for the intended uses and interpretations of the test scores.

1.6 The testing program should pursue an ongoing program of validity research by which it continually gathers and considers a variety of evidence sources to support each intended use and interpretation of test scores. The testing program should provide evidence that results from the full body of validity research, including potentially unfavorable results, are used to make improvements to the testing program.

Structure and Resources of the Testing Program

1.7 The testing program should demonstrate that the relationship between the testing program and any related association, organization, or agency (e.g., licensing board) ensures the independence of the testing program and its related functions.

1.8 If the testing program provides both education and testing, it should administratively and financially separate the educational and testing functions related to any high stakes decisions made about examinees.

1.9 The testing program should demonstrate that its staff possess the knowledge, skills, and professional credentials necessary to conduct the testing program and/or that it has made use of non-staff consultants and professionals to sufficiently supplement staff knowledge and skills.

SECTION 2: EXAMINATION CONTENT

Content Framework and Test Specifications

2.1 The test developer should provide clear documentation of the method used to define the content of the exam. Included in this documentation should be detailed descriptions of any analyses or judgmental procedures as well as their results, that were used at each major step of the process.

2.2 The knowledge, skills, abilities, and judgments to be assessed by the exam should be determined through an accepted content specification methodology such as a job or practice analysis or other comparable method.

2.3 The panel of subject matter experts providing judgments for the job analysis or other content specification method should be demographically and technically representative and possess significant experience with the intended examinee population and the content and skills measured.

2.4 The test content framework should be developed using a psychometrically sound methodology to translate results from the job analysis or content specification method into a comprehensive list of the knowledge, skills, abilities, and cognitive (or non-cognitive) dimensions required for and important to the construct(s) of interest.

2.5 Detailed test specifications should be derived consistent with the test content framework and should, at a minimum, consist of the relative percentages or number of items assigned to each content area, item format, and cognitive (or non-cognitive) dimension.

2.6 The testing program should systematically and periodically reevaluate practices and contents in the tested domain(s) to assure the test content framework remains current in accordance with the intended use and interpretation of exam scores.

Item Development and Selection

2.7 The testing program should provide clear documentation of the procedures used in the development, review, and selection of items.

2.8 Exam items should be written by qualified persons who have knowledge of the content domain and have participated in adequate training in item writing that is relevant to the exam format and content.

2.9 All items should undergo a rigorous review process conducted by qualified and adequately trained persons to evaluate item content, readability, appropriateness of the cognitive level, and the item classification relative to the test specifications.

2.10 The testing program should have a procedure in place to maintain a sufficiently large and adequately balanced item pool that appropriately represents the test specifications.

SECTION 3: FORM DEVELOPMENT AND REVIEW

Pilot Testing

- 3.1** The testing program should pilot test any potential exam items to determine the psychometric adequacy of items prior to their use on any operational forms.
- 3.2** The testing program should provide clear documentation of the procedures used to pilot test the exam items. The conditions under which the pilot data are gathered should be the same as those by which the operational items will be administered.
- 3.3** The testing program should provide clear documentation of the characteristics of the pilot test sample. The pilot sample should be of adequate size and representative of the intended population of examinees with respect to ability (on the construct being measured) and relevant demographic characteristics (e.g., gender, ethnicity, geographic location).
- 3.4** The pilot test stage should incorporate evaluation of examinees' understanding of and interaction with test items, using methodology such as cognitive labs or statistical analysis if possible.
- 3.5** If the exam is computer/internet based, the pilot test process should include a test of the system capabilities, including capacity of the computer hardware and server.

Creation of Final Exam Forms

- 3.6** The testing program should provide clear documentation of the process used to assemble the final operational exam form(s) or, for programs without fixed forms of the examination, to select items for operational administration.
- 3.7** Psychometric analyses should be conducted with the pilot data to determine which items from the item pool should be included in the operational exam forms and to detect potential item bias.
- 3.8** Selection of items for inclusion on the final form(s) should incorporate consideration of the items' psychometric properties in addition to adherence to the test specifications.
- 3.9** The alignment between the final exam form(s) and the test specifications should be objectively evaluated and documented. Testing programs without fixed exam forms should conduct relevant evaluations of the correspondence between the test specifications and the results of the item selection algorithm as utilized in operational administration.

Psychometric Review of Operational Tests

3.10 The testing program should regularly conduct psychometric reviews of all operational items using accepted empirical methods such as classical test theory, item response theory, or other statistical models. Evaluation and documentation of item performance should, at a minimum, include item discrimination, item difficulty, and differential item functioning statistics.

3.11 At least annually, the testing program should conduct a psychometric review of its examination form(s). The review should include the following information for all examination form(s) administered since the last reporting period:

- a. number of examinations administered;
- b. descriptive statistics for exam score distributions, including mean, median, standard deviation, and range;
- c. reliability of the total exam score as well as for key points on the score scale, and decision consistency coefficients (if applicable);
- d. if applicable, the number and percentage of examinees passing the examination or classified into performance categories, including separate reporting for first-time examinees and for retake examinees;
- e. if applicable, indices of model fit.

3.12 The testing program should provide documentation of the psychometric reviews and any actions that were taken based on the results.

Comparability across Forms, Formats, and Language

3.13 If the testing program utilizes multiple exam forms, the process by which equivalence of forms is evaluated and ensured should be clearly documented.

3.14 If the examination is administered in more than one format, evidence should be collected and evaluated to ensure equivalence across examination formats and to ensure examinee performance will result in reliable and valid interpretation of scores regardless of exam format.

3.15 If a test is adapted from one language to another, the testing program should provide evidence supporting intended interpretations of test scores for the adapted exam version. The testing program should provide clear and detailed documentation of the process by which the adapted version was created and of the qualifications and training of those completing the adaptation.

3.16 If adapting an exam from one language or location to another, the testing program should consider the impact of linguistic and cultural differences related to relevant details of exam design, content, or administration, with the consultation of personnel competent in the languages and culture of both the original and adapted version of the examination.

SECTION 4: EXAMINATION ADMINISTRATION

Eligibility and Application

4.1 Eligibility requirements should be documented and should clearly specify necessary qualifications or characteristics of the population for whom the test is intended.

4.2 The testing program should provide clear documentation of the process by which exam candidates apply to take the exam, including instructions given to applicants; the testing program should provide clear documentation of its procedures for reviewing applications, by which it assures that all candidates who are permitted to sit for the exam meet the eligibility requirements.

4.3 In advance of testing, the testing program should provide to examinees information about the purpose of the test, intended use of scores, the scope of test content, testing procedures and format, scoring criteria, and any available alternative formats, including different language formats.

4.4 The testing program should have in place a reasonable retake policy specifying the conditions, including any mandatory wait periods, under which an examinee may retake the exam.

Administration Sites

4.5 When delivered as paper-based examinations, exams should be scheduled far enough in advance to allow for timely shipment of supplies to administration sites or other necessary preparation of materials.

4.6 The testing program should offer the examination at a sufficient number of sites to ensure reasonable access to the exam administration site for as large a percentage of candidates as is practicable, taking into account security concerns.

4.7 If the testing program administers a computer-based or internet-based exam, the testing program should ensure the administration sites conform to necessary technology requirements and that test administration conditions are comparable, if different device types are used.

4.8 Sites chosen for administering examinations should conform to legal requirements for safety, health, and accessibility for all qualified examinees and should maintain the integrity of scores and security. Sites should ensure the following:

- a. accessibility for qualified examinees with disabilities, in accordance with applicable legislation, whether it be at the main site or at an alternative site meeting all other requirements;
- b. adherence to all fire safety and occupancy codes of the jurisdiction in which they are located;
- c. a quiet, well-lit environment with minimal distraction;

- d. sufficient spacing between each examinee or other appropriate and effective methods to preclude any examinee from viewing another examinee's test responses;
- e. acoustics that allow each examinee to hear instructions clearly, using an electronic audio system if necessary;
- f. ventilation and temperature appropriate for health and comfort of examinees.

Test Administrators and Proctors

4.9 Responsibilities, duties, and qualifications of test administrators and monitors/proctors should be directed toward assuring standardized, secure examination administration and fair and equitable treatment of examinees.

4.10 The testing program should provide documentation of the responsibilities and minimum qualification criteria for test administrators and monitors/proctors.

4.11 The testing program should provide suitable training for test administrators and monitors/proctors to enable them to fulfill the above responsibilities and to adequately respond to any testing incidents, disruptions, or anomalies during administration, including security concerns.

4.12 The number of approved monitors/proctors assigned to a test administrator should be sufficient to allow each examinee to be observed and supervised to assure conformance to standardization and security requirements.

4.13 The testing program should have administrators or monitors/proctors sign a confidentiality statement assuring they will not reveal or reproduce any of the test information.

Procedures for Administration

4.14 The testing program should provide each test administrator with a manual detailing the procedures and requirements for all aspects of the examination administration process.

4.15 The test administration should be conducted in a standardized fashion according to the procedures documented in the test administration manual. Any disruptions to the procedure should be documented by the test administrator/proctor and reported to the testing program.

4.16 The examinee should be presented with detailed instructions specifying how to proceed through the exam and clearly stating any rules or procedures the examinee is expected to follow during the administration, including enforced policies regarding personal technology devices such as phones.

4.17 The testing program should establish and enforce procedures by which examinee identity is verified when taking the test to prevent intentional or unintentional identification errors.

Record Keeping

4.18 The testing program should provide evidence of practices and procedures that are implemented to safeguard and keep records of examinee's examination results confidential.

4.19 The testing program should specify the length of time that records of test administrations and score reports will be maintained. The specific length of time should be sufficient for reasonably anticipated needs, such as score appeals, and in accordance with any applicable regulations or laws.

4.20 Records maintained by testing programs should identify the examination form and/or version and specify the date the examination was taken for each examinee.

SECTION 5: FAIRNESS AND DIVERSITY

Accommodations

5.1 The process by which the testing program determines eligibility of candidates to receive accommodations should be clearly documented and be compliant with the guidelines set forth by the Americans with Disabilities Act, the Individuals with Disabilities Education Act, and any other relevant legislation.

5.2 Sufficiently in advance of test administration, the testing program should give clear instructions to candidates about the eligibility criteria for receiving accommodations, which accommodations are allowable, the process by which they may apply for accommodations, and the documentation that must accompany accommodation requests.

5.3 The testing program should enact consistent policies to ensure eligible applicants receive appropriate accommodations; testing programs should provide a rationale for disallowing specific accommodations thought to interfere with accurate measurement of the intended test construct. Procedures by which candidates who are denied accommodations requests may appeal the decision should be specified and shared with such candidates.

5.4 Where accommodations require additional administration personnel, arrangements should be such that neither the security of the examination contents nor the validity of score interpretations is compromised.

Fairness for Diverse Groups

5.5 Throughout the test development process, samples of either examinees or of subject matter experts should include sufficiently large numbers of individuals from demographic groups that are representative of the overall target examinee population, potentially by oversampling these groups if necessary for adequate subgroup sample sizes.

5.6 The testing program should guard against bias by taking steps throughout the test and item development process to evaluate and reduce the influence of extraneous factors on test scores or item responses, using quantitative and judgmental methods.

5.7 The testing programs should provide evidence to support claims that the test can be appropriately used with test takers from diverse backgrounds.

5.8 Panels of experts providing judgements at various stages of exam development should be representative of the diversity in the examinee population in order to identify and prevent potential test or item bias related to culture, gender, language, disability, etc.

SECTION 6: SCORING AND REPORTING

Scoring and Scaling

6.1 The testing program should have clear documentation describing the procedures used for scoring an exam including any procedures used to ensure the accurate recording of item raw scores.

6.2 For paper-based tests, materials containing examinees responses should be shipped to the testing program or scoring facility in a timely and secure manner.

6.3 If the testing program utilizes raters or subjective scorers for any part of the exam, the testing program should provide adequate training for the scorers and have in place a process by which the accuracy of these scores is checked on an ongoing basis.

6.4 Clear documentation should be provided for all procedures related to scaling of the exam scores, including development of the scale, procedures used to translate raw scores to the reporting scale, and rationale for the scaling method.

6.5 The development of the score scale and the translation of raw scores to scale scores should be based on a psychometrically sound method appropriate for the intended interpretation of test scores.

Determining Cut Scores

6.6 If applicable, performance standards should be determined using an accepted standard-setting method appropriate for the exam. If the standard-setting method relies on judgments from panels of subject-matter-experts, those panels should receive adequate training in the standard-setting method.

6.7 Panelists should be provided with a clear description of the purpose of the standard-setting workshop, the intended use of the cut score(s) being estimated, descriptions of the performance standards to be applied, and empirical data to evaluate the impact of their cut-score recommendations.

6.8 Standard-setting panelists should be subject matter experts who are familiar with the target population of examinees and the construct/content being assessed. Panelists selected to participate in the standard-setting workshop should be representative of the stakeholders who are qualified to make decisions about the required proficiency of examinees.

6.9 The procedure and results of the standard-setting workshop should be clearly documented, including the method used to determine the recommended cut score(s), the resulting cut score recommendations, and an estimate of variability in panelists' recommendations. The final cut score(s) adopted and used in practice should be clearly reported.

Norm-Referenced Interpretation

6.10 If the testing program utilizes norm-referenced reporting or interpretation of examinee scores, a norming study should be conducted to collect appropriate data on which to base such score interpretation.

6.11 The selection of individuals included in the norming study should be based on psychometrically sound sampling procedures. The sample selected should be representative of the target population with respect to demographic and other relevant characteristics.

6.12 The norming study should be conducted under the same conditions the operational exam will be administered.

6.13 The documentation of the norming study should provide a clear description of the process, including a description of the sample, administration format, test dates, and results.

Score Reporting

6.14 The testing program should prepare score reports that include a guide to interpreting any scores including any sub-scores. Score interpretations presented in the guide should be in accordance with the test's defined purpose and intended uses of the scores.

6.15 If applicable, the testing program should describe the professional skills and qualifications required to interpret test results.

6.16 The testing program should provide evidence of the reliability and/or precision of each score presented in the score report. If subscores are reported, the testing program should provide evidence of adequate reliability and precision of these subscores. Standard errors or other measures of uncertainty should be presented in a way that is understandable to readers.

6.17 The testing program should have a procedure in place by which score reports will be delivered to authorized recipients within a specified time frame.

6.18 If a testing program is considered to be high stakes for examinees, the testing program should have in place a formal appeals procedure by which examinees can appeal ineligibility to take the exam or accuracy of their scores. This process should include a procedure by which exam scores will be investigated and scoring issues addressed. Test takers should have the opportunity to question the appropriateness of a keyed answer or scoring rubric.

SECTION 7: EXAM SECURITY AND PRIVACY

Exam Material Security

7.1 The testing program should have effective procedures in place to protect the security of the exam items, forms, and administration records, whether physical or stored as data, at all stages of the process; procedures should include routine monitoring to allow for timely detection of potential security breaches.

7.2 When tests are administered in paper format, security of the examination materials should be maintained in shipments to and from the administration site.

7.3 The examinations should be administered in a manner that maximizes the security of the exam contents. This set of procedures includes ensuring only the examinees and authorized proctors see the contents of the exam, before, during, and after administration.

7.4 For computerized exams, testing programs should establish and enforce security protocols that effectively protect test materials and test items from unauthorized access.

7.5 The testing program should establish procedures to be followed in any instance where the security of an examination or the integrity of scores is suspected to be compromised. These procedures should be specific for identifying, handling, and reporting suspected or alleged cheating incidents, lost or stolen booklets, intentional or unintentional divulging of test items by examinees or administration personnel, or any other incidents perceived to have potentially damaged the security of the examination or to have reduced the credibility of examinees' scores.

Security and Privacy of Examinee Data

7.6 The testing program should have effective procedures in place to protect the security and privacy of examinees' personal information, responses, and scores at all stages of the process.

7.7 The testing program should provide documentation of procedures ensuring the confidential release of examinee scores to the examinee and other authorized persons or organizations (e.g., school, licensing agency), as indicated in the program's policy documents.

7.8 For computerized exams, the testing program should provide for appropriate precautions that protect the security of any candidate responses, scores, or personal information as the information is transmitted or stored as data.