Other Publications by the Association for Assessment in Counseling


For ordering information contact: Dr. Brian Glaser, AAC Treasurer at bglaser@coe.uga.edu or visit the AAC website at http://aac.ncat.edu to download order forms.

© 2003, Association for Assessment in Counseling. All rights reserved. Production by any means is unlawful without permission by the Association for Assessment in Counseling.

Acknowledgements

AAC would like to thank AERA for permission to quote selected standards from Standards for Educational and Psychological Testing for this publication. The full publication can be ordered from AERA at http://www.aera.net/products/standards.htm. AAC would also like to thank the American Counseling Association (ACA) for permission to reprint portions of their ethics statements. This document is based on a previous publication created by AAC, but published by ACA; we wish to thank ACA for its willingness to allow AAC to revamp and disseminate this new publication.

Cover by Reynold Wong, Monterey, CA
Preface

The Association for Assessment in Counseling (AAC) is an organization of counselors, counselor educators, and other professionals that advances the counseling profession by providing leadership, training, and research in the creation, development, production, and use of assessment and diagnostic techniques.

The increasing diversity in our society offers a special challenge to the assessment community, striving always to assure fair and equitable treatment of individuals regardless of race, ethnicity, culture, language, age, gender, sexual orientation, religion or physical ability. This is especially important given the increased emphasis place on assessment spawned by national and state legislation and educational reform initiatives.

This document, *Standards for Multicultural Assessment*, is an attempt to create and maintain an awareness of the various assessment standards that have been produced by various professional organizations. It is a compilation of standards produced by several professional associations.

This publication is based on a study completed by the Committee on Diversity in Assessment under the direction of the AAC Executive Council. The first version of this document was published in 1992, and was also published as an article in *Measurement and Evaluation in Counseling and Development* (Prediger, 1994). The original publication was prompted by a request from Jo-Ida Hansen, Chair of the 1991-1992 Committee on Testing of the American Association for Counseling and Development (now ACA). The original publication was prepared by Dale Prediger under the direction of the AAC Executive Council.

Because of advances in professional standards in the past decade, it was necessary to update and expand upon the first version. This revised document was created by a committee of members from the AAC, chaired by Dr. Wendy Charkow-Bordeau along with committee members, Drs. Debbie Newsome and Marie Shoffner. This publication was commissioned by the Executive Council of the Association for Assessment in Counseling.

AAC also wishes to thank Drs. Pat Nellor Wickwire and Janet Wall for their care and assistance in finalizing this document and coordinating its production.

AAC hopes that all counselors, teachers, and other assessment professionals find this document to be useful in improving their assessment practices.
# STANDARDS FOR MULTICULTURAL ASSESSMENT (2nd Ed.)

## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>1</td>
</tr>
<tr>
<td>Definition of Multicultural and Diverse Populations</td>
<td>1</td>
</tr>
<tr>
<td>Classification of Standards</td>
<td>2</td>
</tr>
<tr>
<td>Source Documents</td>
<td>2</td>
</tr>
<tr>
<td>The Standards</td>
<td>3</td>
</tr>
<tr>
<td>Selection of Assessment Instruments</td>
<td>4</td>
</tr>
<tr>
<td>Content and Purpose</td>
<td>5</td>
</tr>
<tr>
<td>Norming, Reliability, Validity</td>
<td>4</td>
</tr>
<tr>
<td>Administration and Scoring of Assessment Instruments</td>
<td>5</td>
</tr>
<tr>
<td>Interpretation and Application of Assessment Results</td>
<td>7</td>
</tr>
<tr>
<td>References</td>
<td>9</td>
</tr>
</tbody>
</table>
Standards for Multicultural Assessment

Purpose

The Association for Assessment in Counseling (AAC), a division of the American Counseling Association (ACA), presents this revised compilation of professional standards. Although AAC believes that tests, inventories, and other assessment instruments can be beneficial for members of all populations, AAC recognizes that the increasing diversity in client backgrounds presents special challenges for test users. The standards assembled here address many of these challenges that are specifically related to the assessment of multicultural populations.

Although a number of standards in this compilation have relevance for the use of assessment instruments in psychological screening, personnel selection, and placement, they were selected because they have special relevance for counseling and for multicultural and diverse populations. Standards that apply in the same way for all populations (e.g., general standards for norming, scaling, reliability, and validity) are not included. Readers may consult the source documents and other publications for universal testing standards.

AAC urges all counselors to subscribe to these standards and urges counselor educators to include this compilation in programs preparing the “culturally competent counselor” (Sue, Arredondo, & McDavis, 1992, p. 447). Finally, AAC supports other professional organizations in advocating the need for a multicultural approach to assessment, practice, training, and research.

Definition of Multicultural and Diverse Populations

A precise definition of multicultural and diverse populations is evolving. The multicultural competencies outlined by Sue et al. (1992), and then revised by Arredondo and Toporek (1996), define the following five major cultural groups in the United States and its territories: African/Black, Asian, Caucasian/European, Hispanic/Latino, and Native American. Arredondo and Toporek differentiated between these cultural groups, which are based on race and ethnicity, and diversity, which applies to individual differences based on age, gender, sexual orientation, religion, and ability or disability.

In revising the Standards for Multicultural Assessment, an inclusive definition of multiculturalism and diversity was used. For the purposes of this document, multicultural
and diverse populations include persons who differ by race, ethnicity, culture, language, age, gender, sexual orientation, religion, and ability.

**Source Documents**

Five documents which include professional standards for assessment in counseling were used as sources for this compilation.


4. **Multicultural Counseling Competencies and Standards** (COMPS). (Association for Multicultural Counseling and Development, 1992). These standards can be viewed in the 1996 article by Arredondo and Toporek. Full reference information is listed below in the reference section.

5. **Code of Ethics and Standards of Practice of the American Counseling Association** (ETHICS). (ACA, 1996). Ordering information can be obtained from ACA, 5999 Stevenson Avenue, Alexandria, VA, 22304-3300. The ethical code and standards of practice may also be viewed on-line at http://www.counseling.org/resources/ethics.htm.

**Classification of Standards**

Sixty-eight standards specifically relevant to the assessment of multicultural and diverse populations were identified in a reading of the five source documents. The content and intent of these standards were analyzed and classified. Assessment roles, functions, and tasks cited in these standards were clustered into three major groups.

**Selection of Assessment Instruments**

- Content and Purpose (n=13)
- Norming, Reliability, and Validity (n=18)

**Administration and Scoring of Assessment Instruments** (n=16)

**Interpretation and Application of Assessment Results** (n=21)
The 68 standards are listed below by cluster and source.

**Selection of Assessment Instruments: Content and Purpose**

1. Evaluate procedures and materials used by test developers, as well as the resulting test, to ensure that potentially offensive content or language is avoided. (CODE, Section A-7)
2. Select tests with appropriately modified forms or administration procedures for test takers with disabilities who need special accommodations. (CODE, Section A-8)
3. For individuals with disabilities, alternative measures may need to be found and used.
4. Care should be taken to select tests that are fair to all test takers. (RUST)
5. Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain. (SEPT 7.4)
6. In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct. (SEPT, Standard 7.7)
7. Linguistic modifications recommended by test publishers, as well as the rationale for modifications, should be described in detail in the test manual. (SEPT, Standard 9.4)
8. In employment and credentialing testing, the proficiency language required in the language of the test should not exceed that appropriate to the relevant occupation or profession. (SEPT, Standard 9.8)
9. Inferences about test takers’ general language proficiency should be based on tests that measure a range of language features, and not on a single linguistic skill. (SEPT, Standard 9.10)
10. Tests selected for use in individual testing should be suitable for the characteristics and background of the test taker. (SEPT, Standard 12.3)
11. Culturally competent counselors understand how race, culture, and ethnicity may affect personality formation, vocational choices, manifestation of psychological disorders, help-seeking behavior, and the appropriateness or inappropriateness of counseling approaches. (COMPS, 13)
12. Culturally competent counselors have training and expertise in the use of traditional assessment and testing instruments. They not only understand the technical aspects of the instruments but also are aware of the cultural limitations. This allows them to use test instruments for the welfare of clients from diverse cultural, racial, and ethnic groups. (COMPS, 29)
13. Counselors are cautious when selecting tests for culturally diverse populations to avoid inappropriateness of testing that may be outside of socialized behavioral or cognitive patterns. (ETHICS, Section III.C.5)
Selection of Assessment Instruments: Norming, Reliability, and Validity

1. Evaluate the available evidence on the performance of test takers of diverse subgroups. Determine to the extent feasible which performance differences may have been caused by factors unrelated to skills being assessed. (CODE, Section A-9).

2. Technical information should be reviewed to determine if the test characteristics are appropriate for the test taker (e.g., age, grade level, language, cultural background). (RUST)

3. Where there are generally accepted theoretical or empirical reasons for expecting that reliability coefficients, standard errors of measurement, or test information functions will differ substantially for various subpopulations, publishers should provide reliability data as soon as feasible for each major population for which the test is recommended. (SEPT, Standard 2.11)

4. If a test is proposed for use in several grades or over a range of chronological age groups and if separate norms are provided for each grade or age group, reliability data should be provided for each age or grade population, not solely for all grades or ages combined. (SEPT, Standard 2.12)

5. When significant variations are permitted in test administration procedures, separate reliability analyses should be provided for scores produced under each major variation if adequate sample sizes are available. (SEPT, Standard 2.18)

6. Norms, if used, should refer to clearly described populations. These populations should include individuals or groups to whom test users will ordinarily wish to compare their own examinees. (SEPT, Standard 4.5)

7. When credible research reports that test scores differ in meaning across examinee subgroups for the type of test in question, then to the extent feasible, the same forms of validity evidence collected for the examinee population as a whole should also be collected for each relevant subgroup. Subgroups may be found to differ with respect to appropriateness of test content, internal structure of test responses, the relation of test scores to other variables, or the response processes employed by the individual examinees. Any such findings should receive due consideration in the interpretation and use of scores as well as in subsequent test revisions. (SEPT, Standard 7.1)

8. When credible research reports differences in the effects of construct-irrelevant variance across subgroups of test takers on performance on some part of the test, the test should be used if at all only for the subgroups for which evidence indicates that valid inferences can be drawn from test scores. (SEPT, Standard 7.2)

9. When empirical studies of differential prediction of a criterion for members of different subgroups are conducted, they should include regression equations (or an appropriate equivalent) computed separately for each group or treatment under consideration or an analysis in which group or treatment variables are entered as moderator variable. (SEPT, Standard 7.6)

10. When a construct can be measured in different ways that are approximately equal in their degree of construct representation and freedom from construct-irrelevant variance, evidence of mean score differences across relevant subgroups of examinees should be considered in deciding which test to use. (SEPT, Standard 7.11)
11. When credible research evidence reports that test scores differ in meaning across subgroups of linguistically diverse test takers, then to the extent feasible, test developers should collect for each linguistic group studied the same form of validity evidence collected for the examinee population as a whole. (SEPT, Standard 9.2)

12. When a test is translated from one language to another, the methods used in establishing the adequacy of translation should be described, and empirical and logical evidence should be provided for score reliability and the validity of the translated test’s score inferences for the uses intended in the linguistic groups to be tested. (SEPT, Standard 9.7)

13. When multiple language versions of a test are intended to be comparable, test developers should report evidence of test comparability. (SEPT, Standard 9.9)

14. When feasible, tests that have been modified for use with individuals with disabilities should be pilot tested on individuals who have similar disabilities to investigate the appropriateness and feasibility of the modifications. (SEPT, Standard 10.3)

15. When sample sizes permit, the validity of inferences made from test scores and the reliability of scores on tests administered to individuals with various disabilities should be investigated and reported by the agency or publisher that makes the modification. Such investigations should examine the effects of modifications made for people with various disabilities on resulting scores, as well as the effects of administering standard unmodified tests to them. (SEPT, Standard 10.7)

16. When relying on norms as a basis for score interpretation in assisting individuals with disabilities, the norm group used depends upon the purpose of testing. Regular norms are appropriate when the purpose involves the test taker’s functioning relative to the general population. If available, normative data from the population of individuals with the same level or degree of disability should be used when the test taker’s functioning relative to individuals with similar disabilities is at issue. (SEPT, Standard 10.9)

17. When circumstances require that a test be administered in the same language to all examinees in a linguistically diverse population, the test user should investigate the validity of the score interpretations for test takers believed to have limited proficiency in the language of the test. (SEPT, Standard 11.22)

18. Counselors carefully consider the validity, reliability, psychometric limitations, and appropriateness of instruments when selecting tests for use in a given situation or with a particular client. (ETHICS, Section E.6.a)

Administration and Scoring of Assessment Instruments

1. Provide and document appropriate procedures for test takers with disabilities who need special accommodations or those with diverse linguistic backgrounds. Some accommodation may be required by law or regulation. (CODE, Section B-2)

2. For individuals with disabilities, accommodations in test taking procedures may need to be employed. Appropriate modifications of testing materials and procedures in order to accommodate test takers with special needs are to be arranged. (RUST)

3. Include notes on any problems, irregularities, and accommodations in the test records. (RUST)
4. A systematic and objective procedure is in place for observing and recording environmental, health, emotional factors, or other elements that may invalidate test performance and results; deviations from prescribed test administration procedures, including information on test accommodations for individuals with special needs, are recorded. Carefully observe, record, and attach to the test record any deviation from the prescribed test administration procedures. Include information on test accommodations for individuals with special needs. (RUST)

5. The testing or assessment process should be carried out so that test takers receive comparable and equitable treatment during all phases of the testing or assessment process. (SEPT, Standard 7.12)

6. Testing practice should be designed to reduce threats to the reliability and validity of test score inferences that may arise from language differences. (SEPT, Standard 9.1)

7. When testing an examinee proficient in two or more languages for which the test is available, the examinee’s relative language proficiencies should be determined. The test generally should be administered in the test taker’s most proficient language, unless proficiency in the less proficient language is part of the assessment. (SEPT, Standard 9.3)

8. When an interpreter is used in testing, the interpreter should be fluent in both the language of the test and the examinee’s native language, should have expertise in translating, and should have a basic understanding of the assessment process. (SEPT, Standard 9.11)

9. People who make decisions about accommodations and test modifications for individuals with disabilities should be knowledgeable of existing research on the effects of the disabilities in question on test performance. Those who modify tests should also have access to psychometric expertise for so doing. (SEPT, Standard 10.2)

10. If a test developer recommends specific time limits for people with disabilities, empirical procedures should be used, whenever possible, to establish time limits for modified forms of timed tests rather than simply allowing test takers with disabilities a multiple of the standard time. When possible, fatigue should be investigated as a potentially important factor when time limits are extended. (SEPT, Standard 10.6)

11. Those responsible for decisions about test use with potential test takers who may need or may request specific accommodations should (a) possess the information necessary to make an appropriate selection of measures, (b) have current information regarding the availability of modified forms of the test in question, (c) inform individuals, when appropriate, about the existence of modified forms, and (d) make these forms available to test takers when appropriate and feasible. (SEPT, Standard 10.8)

12. Any test modifications adopted should be considered appropriate for the individual test taker, while maintaining all feasible standardized features. A test professional needs to consider reasonably available information about each test taker’s experiences, characteristics, and capabilities that might impact test performance, and document the grounds for the modification. (SEPT, Standard 10.10)

13. If a test is mandated for persons of a given age or all students in a particular grade, users should identify individuals whose disabilities or linguistic background indicates the need for special accommodations in test administration and ensure that those accommodations are employed. (SEPT, Standard 11.23)
14. Counselors provide for equal access to computer applications in counseling services. (ETHICS, Section A.12.c)

15. When computer applications are used in counseling services, counselors ensure that: (1) the client is intellectually, emotionally, and physically capable of using the computer application; (2) the computer application is appropriate for the needs of the client; (3) the client understands the purpose and operation of the computer applications; and (4) a follow-up of client use of a computer application is provided to correct possible misconceptions, discover inappropriate use, and assess subsequent needs. (ETHICS, Section A.12.a)

16. Prior to assessment, counselors explain the nature and purposes of assessment and the specific use of results in language the client (or other legally authorized person on behalf of the client) can understand, unless an explicit exception to this right has been agreed upon in advance. (ETHICS, Section E.3.a)

**Interpretation and Application of Assessment Results**

1. Interpret the meaning of the test results, taking into account the nature of the content, norms or comparison groups, other technical evidence, and benefits and limitations of test results. (CODE, Section C-1)

2. Review the procedures for setting performance standards or passing scores. Avoid using stigmatizing labels. (CODE, Section C-4)

3. For individuals with disabilities, interpretations need to be made in light of the modifications in the test or testing procedures. (RUST)

4. When test results are influenced by irrelevant test taker characteristics (e.g., gender, age, ethnic background, cheating, availability of test preparation programs) the use of the resulting information is invalid and potentially harmful. (RUST)

5. Factors such as the test taker’s group membership and how that membership may impact the results of the test is a critical factor in the interpretation of test results. Specifically, the test user should evaluate how the test taker’s gender, age, ethnicity, race, socioeconomic status, marital status, and so forth, impact on the individual’s results. (RUST)

6. If local examinees differ materially from the population to which the norms refer, a user who reports derived scores based on the published norms has the responsibility to describe such differences if they bear upon the interpretation of the reported scores. (SEPT, Standard 4.7)

7. In testing applications involving individualized interpretations of test scores other than selection, a test taker’s score should not be accepted as a reflection of standing on a characteristic being assessed without consideration of alternate explanations for the test taker’s performance on that test at that time. (SEPT, Standard 7.5)

8. When scores are disaggregated and publicly reported for groups identified by characteristics such as gender, ethnicity, age, language proficiency, or disability, cautionary statements should be included whenever credible research reports that test scores may not have comparable meaning across different groups. (SEPT, Standard 7.8)

9. When tests or assessments are proposed for use as instruments of social, educational, or public policy, the test developers or users proposing the test should fully and accurately inform policymakers of the characteristics of the tests.
as well as any relevant and credible information that may be available concerning the likely consequences of test use. (SEPT, Standard 7.9)

10. When the use of a test results in outcomes that affect the life chances or educational opportunities of examinees, evidence of mean test score differences between relevant subgroups of examinees should, where feasible, be examined for subgroups for which credible research reports mean differences for similar tests. Where mean differences are found, an investigation should be undertaken to determine that such differences are not attributable to a source of construct underrepresentation or construct-irrelevant variance. While initially the responsibility of the test developer, the test user bears responsibility for users with groups other than those specified by the developer. (SEPT, Standard 7.10).

11. When score reporting includes assigning individuals to categories, the categories should be chosen carefully and described precisely. The least stigmatizing labels, consistent with accurate representation, should always be assigned. (SEPT, Standard 8.8)

12. When there is credible evidence of score comparability across regular and modified administrations, no flag should be attached to a score. When such evidence is lacking, specific information about the nature of the modification should be provided, if permitted by law, to assist test users properly to interpret and act on test scores. (SEPT, Standard 9.5 and 10.11)

13. In testing persons with disabilities, test developers, test administrators, and test users should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disabilities and their associated characteristics extraneous to the intent of the measurement. (SEPT, Standard 10.1)

14. In testing individuals with disabilities for diagnostic and intervention purposes, the test should not be used as the sole indicator of the test taker’s functioning. Instead, multiple sources of information should be used. (SEPT, Standard 10.12)

15. Agencies using tests to conduct program evaluations or policy studies, or to monitor outcomes, should clearly describe the population the program or policy is intended to serve and should document the extent to which the sample of test takers is representative of that population. (SEPT, Standard 15.5)

16. Reports of group differences in average test scores should be accompanied by relevant contextual information, where possible, to enable meaningful interpretation of these differences. Where appropriate contextual information is not available, users should be cautioned against misinterpretation. (SEPT, Standard 15.12)

17. Culturally competent counselors possess knowledge about their social impact on others. They are knowledgeable about communication style differences, how their style may clash or facilitate the counseling process with minority clients, and how to anticipate the impact it may have on others. (COMPS, 7)

18. Culturally competent counselors have knowledge of the potential bias in assessment instruments and use procedures and interpret findings keeping in mind the cultural and linguistic characteristics of the clients. (COMPS, 22)

19. Counselors recognize that culture affects the manner in which clients' problems are defined. Clients' socioeconomic and cultural experience is considered when diagnosing mental disorders. (ETHICS, Section E.5.b)

20. Counselors are cautious in using assessment techniques, making evaluations, and interpreting the performance of populations not represented in the norm group on which an instrument was standardized. They recognize the effects of age, color, culture, disability, ethnic group, gender, race, religion, sexual
orientation, and socioeconomic status on test administration and interpretation and place test results in proper perspective with other relevant factors. (ETHICS, Section E.8)

21. In reporting assessment results, counselors indicate any reservations that exist regarding validity or reliability because of the circumstances of the assessment or the inappropriateness of the norms for the person tested. (ETHICS, Section E.9.a)

References


AAC Executive Council

President – Dr. Janet Wall, Sage Solutions  
Past President – Mrs. Patricia Jo McDivitt, Data Recognition Corporation  
President-Elect – Dr. Brad Erford, Loyola College of Maryland  
Secretary – Dr. Debbie Newsome, Wake Forest University  
Treasurer – Dr. Brian Glaser, University of Georgia  
Member-at-Large Publications – Dr. David Jepsen, University of Iowa  
Member-at-Large Awards – Dr. Claire Miner, Assessment and Counseling Services  
Member-at-Large Membership – Dr. Donna Gibson, The Citadel  
Governing Council Representative – Dr. F. Robert Wilson, University of Cincinnati

Association for Assessment in Counseling

Vision: The Association for Assessment in Counseling (AAC) is an organization of counselors, counselor educators, and other professionals that advances the counseling profession by providing leadership, training, and research in the creation, development, production, and use of assessment and diagnostic techniques.

Mission: The mission of AAC is to promote and recognize scholarship, professionalism, leadership, and excellence in the development and use of assessment and diagnostic techniques in counseling.

Purposes: AAC is positioned to fulfill seven fundamental purposes:

- **Administration and Management**: to provide long range planning, policies, organizational structure, operating procedures, and resources to fulfill AAC’s missions;
- **Professional Development**: to promote professional development which enhances competence in assessment, evaluation, measurement, and research for counselors, counselor educators, and other professionals who develop or use assessment and diagnostic tools and techniques;
- **Professionalization**: to promote the professionalization of counseling through the appropriate use of assessment;
- **Research and Knowledge**: to promote the development and dissemination of knowledge regarding assessment procedures used in counseling;
- **Human Development**: to promote concern for human rights as integral to all assessment activities and to serve as a resource to counselors, counselor educators, and other professionals concerning the assessment aspects of human development;
- **Public Awareness and Support**: to promote and support public policies and legislation that advance the appropriate use of assessment in optimizing human potential;
- **International and Interprofessional Collaboration**: to promote communication and collaboration between AAC and other professional organizations (national and international) in order to address common, assessment-related concerns.

Contact: 3gibsond@citadel.edu for membership information