CALIFORNIA COMMUNITY COLLEGES

Standards for Assessment Test Instrument Review

2021

Table of Contents

Note: These standards and this document are under review and being updated.

SECTION ONE: ASSESSMENT IN THE STUDENT EQUITY AND ACHIEVEMENT PROGRAM

Assessment is a holistic process through which each college collects information about students in an effort to facilitate their success by ensuring their appropriate placement into the curriculum. Assessment is one of the core services provided to students through the Student Success and Support Program (SSSP) established under the Seymour-Campbell Student Success Act of 2012. Per the Education Code, section 78211.5(a), the purpose of the Act was "to increase California community college student access and success by providing effective core matriculation services, including orientation, assessment and placement, counseling, and other education planning services, and academic interventions." The Student Equity and Achievement Program replaced the SSSP program on June 29, 2020, but maintains these same (and additional) aims.

Assessment is governed by California Education Code, sections 78210-78219 and California Code of Regulations, title 5, sections 55502-55532. In addition to these statutes and regulations, colleges must adhere to the standards provided in this document when implementing and managing any assessment instrument used for course placement. The Education Code and title 5 sections referenced are included in Appendix A. The core requirements of placement assessments and their review are summarized below.

Per the California Code of Regulations (CCR), title 5, section 55522, high school performance data should be used as the primary source for placement in English and mathematics (or quantitative reasoning) for all U.S. high school graduates (or the equivalent). While districts are allowed to use multiple measures in placing students, any form of assessment must be submitted to the Chancellor's Office for review and approval. Per Education Code 78213, a community college district or college shall not use any assessment instrument related to Education Code 78213 without the authorization of the board of governors. The board of governors may adopt a list of authorized assessment instruments and shall establish an advisory committee to review and make recommendations concerning all assessment instruments used by districts and colleges related to Education Code 78213. California Code of Regulations (CCR), title 5, section 55522, further stipulates that assessment tests and instruments for use in placing students in English, mathematics (or quantitative reasoning), or English as a Second Language (ESL) courses must be approved by the Chancellor's Office, along with guidelines for their use by community college districts.

Education Code 78213 further defines assessment as "the process of gathering information about a student regarding the student's study skills, English language proficiency, computational skills, aptitudes, goals, learning skills, career aspirations, academic performance, and need for special services. Assessment methods may include, but not necessarily be limited to, interviews, standardized tests, attitude surveys, vocational or career aptitude and interest inventories, high school or postsecondary transcripts, specialized certificates or licenses, educational histories, and other measures of performance."

Additionally, per Education Code 78213, assessment instruments must meet the following requirements:

(1) Assessment instruments shall be sensitive to cultural and language differences between students, and shall be adapted as necessary to accommodate students with disabilities.
(2) Assessment instruments shall be used as an advisory tool to assist students in the selection of appropriate courses.
(3) Assessment instruments shall not be used to exclude students from admission to community colleges.

California Code of Regulations (CCR), title 5, section 55522, further details what is required of California Community Colleges:

- Use only assessment instruments approved by the California Community Colleges Chancellor's Office, with the exception of limited field testing for new or alternative tests.
- Evaluate assessment instruments to meet content validity, cut score validity, minimization of bias, reliability, and disproportionate impact standards defined by the Chancellor's Office.
- Implement a plan to address any disproportionate impact identified, in consultation with the Chancellor's Office.
- Adopt and clearly communicate procedures regarding the college's sample test preparation, placement decisions, and retest policies.
- Use assessment instruments solely for the purpose for which they were developed and for which purposes they have been validated.
- Prohibit use of assessment instruments to exclude a student from admission to the college, from any particular course, or educational program, except in the case of nursing programs or special part-time or full-time students under Education Code section 76002, based on an assessment that involves multiple measures and complies with title 5.
- Any placement decision must be supported by data and validated to ensure that no tool or measure is being used to preclude students from enrolling in a course they have a legal right to access.
- Placement practices should be designed to fulfill the requirements of title 5 § 55522 and § 55522.5.

Title 5 further requires students with disabilities to be provided necessary accommodations (section 55526(a)). In addition, title 5 addresses student responsibilities (section 55530) and institutional responsibilities (section 55531), including assessment, as well as exemption policies for assessment and other services (section 55532) that colleges may choose to implement.

These state regulations provide the context for establishing the standards for assessment review. A test that provides information gathered to make course placement decisions regarding individual students must be submitted for approval from the Chancellor's Office to ensure its validity, reliability, and fairness (freedom of bias).

The Assessment Instrument Approval Process: Broad Overview

The ultimate responsibility for the validity and successful use of assessment instruments and procedures and resulting course placement rests with local colleges. Approval of an assessment instrument by the Chancellor's Office allows a college to use the instrument; however, approval does not automatically endorse **how** a local college uses the test. Each college must sufficiently document that the test is used appropriately, regardless of whether the test has been created by a second-party publisher or was locally developed or managed.

Any test used to assist with the appropriate placement of students into different levels of instruction, classes, or programs must receive approval from the Chancellor's Office. Although assessment is broadly defined in Education Code (see above), these standards focus largely on tests specifically, although guidance is also provided for other measures used in placement decisions. The Standards for Educational and Psychological Testing (2014; hereafter referred to as Joint Standards) define a test as "an evaluative device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process."

When requesting approval of a test instrument, second-party publishers and colleges locally developing or managing a test must take the following steps:

1. Compile and submit information on assessment instruments to the Chancellor's Office.
2. Review the preliminary approval recommendation from the Chancellor's Office and respond to any questions or requests for additional information.
3. Review the final report regarding the determination of approval and prepare an appeal to determination, if needed.
4. Continue to collect and document the validity and the impact of using the instrument.

Tests that have not received approval from the Chancellor's Office may **not** be used to place students. They may, however, be used on an experimental or pilot basis, such as to conduct research needed to obtain approval, or they may be used to assess student progress.

The remainder of this document provides standards for the review of test instruments. While these standards define the criteria for judging the acceptability of an instrument, they should not be considered complete instructions for test approval and validation, nor should they be used as the only guidance needed to perform the necessary documentation.

Applicants for approval are strongly urged to identify experienced researchers, psychometricians, and other relevant experts to assist them in conducting the research and providing the necessary information required in the approval process. It is also highly recommended to become familiar with the Standards for Educational and Psychological Testing (2014), the Code of Fair Testing Practices in Education (2004) and the Uniform Guidelines for Employee Selection Procedures (1978).

SECTION TWO: STANDARDS FOR THE EVALUATION OF TEST INSTRUMENTS

National Standards Governing the Evaluation of Test Instruments

The Standards for Educational and Psychological Testing (2014; hereafter referred to as Joint Standards) represent the joint efforts of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) to establish guidelines and criteria for the development, use, and evaluation of tests. The Joint Standards are intended to apply to a wide range of test instruments and procedures that sample, evaluate, and score an individual's behavior through a standardized process. As noted above, the Joint Standards (p. 183) define a test as "an evaluative device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process."

The Joint Standards is the primary guidance document for assessment and measurement professionals in the United States and many other countries worldwide. It has been referenced in federal law and cited in Supreme Court and other judicial decisions, where it has been recognized as setting the gold standard for the testing profession. Given the document's recognized authority on tests and testing practices, the Joint Standards also serves as the primary reference for this document (the CCC Assessment Standards).

Two additional reference documents have guided the development of the CCC Assessment Standards: the Code of Fair Testing Practices in Education (2004) and the Uniform Guidelines for Employee Selection Procedures (1978), hereafter referred to as the EEOC Guidelines.

The Code of Fair Testing Practices was developed by the Joint Committee on Testing Practices as a guide for professionals to help them meet "their obligation to provide and use tests that are fair to all test takers" (p.3). The document is consistent with the Joint Standards but focuses more narrowly on testing in education and presents guidelines separately for test developers and test users.

The EEOC Guidelines developed by the Equal Employment Opportunity Commission (EEOC) were considered in the development of the CCC Assessment Standards for their specific focus on the proper use of tests for fair and equitable selection purposes. In particular, the EEOC Guidelines' criteria regarding the identification of adverse impact and requirements regarding the documentation of evidence greatly informed the assessment review criteria for disproportionate impact and the testing of special groups in this document.

The CCC Assessment Standards are intended to align with these three reference sources.

Criteria for the Evaluation of Assessment Instruments

The CCC Assessment Standards are organized around five key areas of review:

1. Fairness
2. Validity
3. Reliability and errors of measurement
4. Scaling, norming, score comparability, and cut scores
5. Test administration, scoring, reporting, and interpretation

Each key area specifies several criteria that must be considered during assessment review, with references that have been paraphrased from the Joint Standards.

Fairness

The diversity of the test-taking population requires an evaluation of the appropriateness and fairness of test use with special groups, such as individuals of different linguistic backgrounds or with disabilities. The special characteristics of these groups may require test accommodations to minimize barriers irrelevant to measuring student knowledge. Such barriers may be found in test content, setting, instructions, response format, access, or opportunity to learn. It is important that test developers and test users are cognizant of the potential presence of these barriers and take appropriate measures to mitigate them. Criteria that enhance measurement and evaluation of diverse populations are presented below:

Criterion 1. Testing Special Groups: General. Decisions regarding appropriate test selection, provision of test accommodations or other modifications of testing procedures, and test score interpretation with respect to individuals with special characteristics and needs must always be made by someone who has expertise in testing these special groups or in consultation with someone who possesses this expertise.

Criterion 2. Test Design for Non-Native English Speakers. Tests and test procedures must minimize threats to validity and reliability that may arise from language differences. Any modifications made to accommodate individuals with limited language proficiency must be described in detail. When a test is used with linguistically diverse test takers, information must be provided for appropriate test use and test score interpretation. Translated tests must be evaluated for reliability, validity, and comparability with the English version.

Criterion 3. English Language Proficiency Tests for Non-Native English Speakers**.** Assessments of English language proficiency must be comprehensive in scope (e.g., reading, grammar, writing, listening, and speaking) and must reflect the language requirements of the academic environment. It is insufficient to assess only a single language skill such as reading or writing in order to draw inferences about an individual's overall language proficiency.

Criterion 4. Test Design for People with Disabilities. Expertise in psychometrics and populations with disabilities is required for modifications on tests provided to individuals with disabilities. Knowledge of the effects of various disabilities on test performance is essential. Until validity data are obtained for test scores resulting from non-standardized testing conditions, test documentation must issue cautions for the interpretation of test scores. Pilot testing of any modified assessment instruments is strongly advised with persons having the same or similar disability. When feasible, time limits should be modified for students with disabilities on the basis of expert judgment and/or previously conducted reliability and validity studies.

Criterion 5. User Responsibility. When modified forms of a test are available for special populations, they are to be used with these students. Proper selection of appropriate norms to facilitate valid score interpretation for these groups is essential. Using personnel who have been specifically trained for the group or person to be tested is also strongly encouraged for test administration.

Validity

Validity is the most fundamental concern when evaluating an assessment process and placement decision. Validity is demonstrated through a variety of evidence sources that support the specific interpretation of test scores and their use. It is important to point out that a test or assessment method itself is not validated but rather the interpretation and use of the score. If the score interpretation and use differs across applications, then each specific application requires validation. The following review criteria address various types of validity evidence that are especially germane to the use of assessments for course placement purposes in the California community colleges. Unless otherwise noted, test developers ordinarily have responsibility to provide the information called for by each criterion for their tests.

Criterion 1. Validity: General. Evidence must be provided by the test developer or test user supporting the particular interpretation and use(s) of the test scores. Each evidence source must be described in detail and a rationale be given that explains how the reported evidence supports test score interpretation and use. The rationale given for each evidence source will guide test reviewers in determining the soundness and sufficiency of the evidence.

In some assessment contexts—for example, during the initial implementation of a course placement assessment—validity evidence may not be available for review. In those contexts, test materials must explicitly state that the validity of the assessment has not been established. Test developers or test users must accompany those statements with a plan for test validation that specifies the type of validity evidence to be provided, a rationale for the type of validity evidence that will be provided, and a data collection plan.

If a test developer or test user substantially alters a test or its use, new validity evidence obtained under the changed testing conditions must be presented for the altered test.

Alternatively, the test developer or test user may offer a rationale that addresses why the changes to the test have not substantially altered it. Changed testing conditions can result from changes in item format, test administration procedures, language, instructions, or test content. The extent of the revalidation will likely vary with the nature of the changes.

Criterion 2. Content-Related Evidence. If a test is proposed to represent a defined domain of content and/or skills, then a clear definition of the content domain and rationale for its relevance to the proposed test use(s) must be provided. Additionally, the relationship between item content and the content domain must be described fully and accurately. For example, test developers must provide information describing how individual items map onto the various categories that make up the content domain. Sufficient detail is expected so that test users and reviewers can evaluate the range of content in the assessment instrument and consider its appropriateness. If content experts are used to judge the appropriateness of the selected test content, then their qualifications and independence from the test development must be described.

Criterion 3. Construct-Related Evidence. If a test is proposed to measure a construct, the construct must be specified and well defined. In addition, evidence to support the inference from the test to the construct must be presented. Evidence must show the relationship of the instrument to what it measures, and, conversely, that it does not relate to what it should not measure. The instrument should also show relationships with other variables when such relations can be expected based on psychological or educational theories, including non-cognitive variables that relate to motivation, such as hope or leadership experience.

Criterion 4. Criterion-Related Evidence. If a test is proposed to measure content that is similar to an already existing test and/or dissimilar to another already existing test, then the expected strength of the relationship between performance on the new test and on the criterion test must be demonstrated using correlational or other statistical evidence. All criterion-related studies must be completely described, including a specification of the sample, data collection process, and statistical analyses. In addition, substantive and psychometric information must be provided for each criterion measure along with a rationale for their selection.

Criterion 5. Differential Prediction. Differential prediction should be investigated whenever feasible, and when prior research has established a substantial likelihood for differential prediction to occur with a particular type of test. That is, when groups differ, for example, in terms of demographics, past experiences, or instructional treatment, and when such factors are expected to produce differential performance on the test, then it must be determined if decisions are systematically different for the members of a given group compared to all groups combined. Such information may be useful if disproportionate impact is indicated for certain groups that differ in terms of their performance on a given test.

Criterion 6. Cut Scores. If a test is used to make classification decisions about test takers based on cut scores (e.g., classifying test takers into groups or categories such as course

placement), then the method used to determine the cut scores must be fully documented. Documentation must include information that allows reviewers and other test users to evaluate the appropriateness and rationale for each cut score. When cut scores are based on professional judgment, the qualifications of the judges must be documented. Colleges are also advised that the setting of cut scores should be done in ways that are consistent with other standards on prerequisites such that cut scores cannot be used to preclude students from enrolling in a course in which they have a reasonable likelihood of successful completion.

Reliability and Errors of Measurement

No assessment instrument is free of error. Recognition of this fact requires that the reliability of the assessment instrument and the degree of error associated with test scores be documented. Because error results from many different sources depending on what is measured and the assessment context, the type of reliability evidence that is provided should take into account the error sources that are most relevant and of greatest concern for the assessment instrument. The following criteria for reliability of tests are most applicable to the California community colleges:

Criterion 1. Reliability: General. To determine whether each reported score is sufficiently accurate for the intended use, estimates of reliability and standard errors of measurement must be provided. The sample characteristics, statistics, and the methodology employed to document reliability must be described completely. If theoretical or empirical information suggests that estimates differ by population group, then estimates must be provided for each major population group. Direct performance assessments that rely on human judgment must document adequate levels of scorer consistency using an appropriate statistical method.

As a note of caution, proper documentation of reliability evidence is in itself not sufficient to meet standards for reliability unless the reported estimates of reliability and standard error indicate sufficient accuracy for the test's intended use.

Criterion 2. Type of Reliability Estimates Provided. The type of reliability estimate should be appropriate for the proposed test score use and properly interpreted. For example, coefficients of internal consistency yield information about the agreement among test items but not about the test score’s stability over time.

When tests are speeded by design, appropriate methods yielding non-spurious reliability estimates must be used. When corrected coefficients are reported, uncorrected indices must be presented as well. If a test is scored using human raters and a judgmental process, the degree of error between the human raters must be documented using appropriate statistical methods.

Criterion 3. Specific Reliability Applications. For tests relying on cut scores, reliabilities need to be reported at each cut score or for the score intervals separated by cut scores. Also,

decision consistency information must be reported for selected score points. For computer-adaptive tests, reliabilities must be reported for repeated administrations using different item selections.

Scaling, Norming, Score Comparability, and Equating

The metric or score scale with which test scores are reported is generally chosen to support the interpretability of test scores with respect to their intended meaning and use. Frequently, raw test scores are transformed to facilitate proper test score interpretation. For example, test scores may be transformed to facilitate norm-referenced interpretations; that is, to produce information about a test taker's relative standing within a population or comparison group.

Another important reason to transform test scores is to achieve comparability of test scores across different forms of an assessment. Regardless of the utility achieved by establishing derived scales, the resultant transformed scores can introduce error in the measurement process due to the procedure itself or the sampling methodology used. Important criteria for evaluating the appropriateness of test score transformations used in the California community colleges are presented below. As these criteria suggest, benefits realized by transforming scores must be carefully weighed with any drawbacks.

Criterion 1. Choice of Scales. The method used to compute the transformed (derived) scale or raw score must be clearly delineated. In addition, the rationale must address the relationship between the scaling methodology and the test's purpose.

Criterion 2. Norms. If a test proposes a norm-referenced interpretation of test scores, the reference group to which test scores are compared must be clearly described. Furthermore, the choice of reference group must be appropriate for the proposed interpretation and use of test scores. The methodology for constructing norms, including sampling plan, participation rates, and descriptive statistics, must be exhaustively specified. In addition, the year(s) in which the norming data were collected must be reported. Outdated norms generally do not support the use of test scores with current populations unless an argument can be made that the population has not substantially shifted or changed.

Criterion 3. Comparability of Scores. When test scores based on different test forms and/or different response formats are intended to be interchangeable, data that support the equivalence of the test forms must be provided. When the test content changes across years, the changes must be fully described along with a rationale for each change. Test users must be informed of all test changes and how the test changes impact the comparability of test scores. When the changes to a test are substantial, it must be revalidated.

Standards for Administration, Scoring, and Interpretation of Standardized Tests

To ensure that test takers experience the same test conditions, it is essential that procedures for test administration, scoring, and reporting are carefully documented and made available to test users. Instructions must be clear, accurate, and complete and must enable test users to accurately implement all procedures as well as to make informed decisions when selecting an assessment instrument appropriate for their need. The following criteria speak to the adequacy of documentation and the use of information as it pertains to the assessment needs in the California community colleges.

Criterion 1. Administration and Scoring. The standardized procedures for the administration and scoring of an assessment instrument must be fully described in the accompanying manual. A test manual must identify the qualifications necessary to administer the test appropriately. The standardized procedures for test administration and scoring specified in the manual must be appropriate for the purpose of the assessment instrument. Any modification of standardized test administration procedures or scoring must be fully described in the manual with appropriate cautions noted. A local community college or its representative that develops a test has the same obligation to supply manuals and technical reports as does a commercial test publisher. Standardized scoring instructions and rubrics are essential when the measure is a direct performance assessment.

Criterion 2. Interpretation: General. A test manual should identify qualifications necessary to interpret test results. Screening measures should be used only to identify individuals for further evaluation. Test users must not base decisions on interpretations of test scores unless they have documentation that indicates the validity of the interpretations for the intended use.

Criterion 3. Interpretation: Test for Certification. If a test is used to certify completion of a given education level, both the test content domain and the content domain of instruction at the target education level must be described in sufficient detail so that the agreement between the content domain of instruction and the test content can be assessed. An assessment instrument should not cover material that a student has not had an opportunity to learn. Students must have multiple opportunities to take a test used for certification.

Criterion 4. Test Materials. Test documentation must be readable and understandable. Any claims regarding test properties and characteristics must be limited to those for which data exist to support the claim. Such data must be documented and made available to test users.

Criterion 5. Decision Making. A decision or characterization that will have a major impact on a test taker must not be made solely on the basis of a single test score or single measure. California Code of Regulations, title 5, section 55522 requires that when colleges use an assessment for course placement, “it must be used with one or more other measures to

comprise multiple measures." Title 5, section 55502(i) further defines the "multiple measures" as "a required component of a district's assessment system and refers to the use of more than one assessment measure in order to assess the student." Decisions are to be made in conjunction with other assessment information, previous classroom performance, and/or opinions of advisors familiar with the impacted test taker. Section 7 provides more information on multiple measures.

Criteria Associated with Direct Performance Assessment

The preceding discussion of fairness, validity, reliability, scaling and equating, and administration offers guidelines for the evaluation of tests in general. Additional criteria are useful for the evaluation of direct performance assessments, such as essay writing tests, which require test takers to perform a real-life or simulated task. Direct performance assessments generally require an evaluation of the quality of performance based on explicit performance criteria. Accuracy can play an important role as an evaluation criterion but it may not be the sole criterion. Typically, direct performance assessments are scored by human judges using a scoring rubric or other well-defined scoring procedure. The following sections present specific criteria pertinent to the evaluation of direct performance assessments in the California community college context.

Fairness

Special care needs to be taken when considering the appropriateness of direct performance assessments for special student groups (e.g., students with special needs). Due to the inherently greater complexity of performance-based response formats, there is a greater likelihood of construct-irrelevant factors influencing both test performance as well as the judgment of human scorers. Alternative means of responding to test prompts or tasks may need to be incorporated into the assessment process for some groups (e.g., the visually impaired).

Validity

There are three central validity concerns regarding direct performance assessments:

- the development and selection of appropriate prompts or tasks;
- the development of fair and reliable scoring procedures; and
- the use and re-use of prompts or tasks over time.

The first and second concerns may both be addressed by providing detailed documentation of the development of prompts and tasks and of scoring procedures and their logical connection to the measurement construct. The third concern should be addressed by presenting evidence of the comparability of prompts and tasks across test takers as well as across

assessment occasions. Lastly, evidence should be provided that the assessment instrument can be administered and scored in a standardized (consistent and well-defined) manner.

Reliability and Errors of Measurement

Most direct performance assessments rely on human scorers to accurately and consistently apply scoring rules. Consequently, reliability evidence should focus on the human scorers as a potential source of error and include estimates of inter-rater reliability. When performance assessments vary the types of prompts or tasks across test takers, reliability evidence must be presented separately for each prompt or task. The scorer orientation or training should be articulated and the procedures for scorer calibration (norming) should be specified. Additionally, when appropriate, procedures or evidence must be presented addressing intra-rater reliability issues and concerns to ensure that rater drift does not occur over a scoring period.

Scaling, Norming, Score Comparability, and Equating

When test scores from direct performance assessments are transformed into scale scores, the scaling method must be clearly delineated. Furthermore, a rationale must be given that addresses how the scaling methodology supports the test's purpose.

When a total score is computed on the basis of individual component scores—for example, when individual performance criteria are judged and scored separately as is the case in analytical scoring models—then the methodology used to derive the total score (computations, weighting, etc.) along with a rationale for its use must be provided.

Additionally, when direct performance assessments vary the types of prompts or tasks across test takers, documentation of the equivalence of scoring procedures across prompts or tasks must be presented. If the test content or the scoring procedures change across prompts or tasks but the resulting scale scores are intended to be comparable, the method for maintaining comparability must be described and be shown to be appropriate for its intended use.

Standards for Administration, Scoring, and Interpretation of Standardized Tests

Because of the prevalence of human scoring in direct performance assessments, emphasis must be given on the clear documentation of scoring procedures, including the description of procedures implemented to train scorers. A second concern is the threat to test security that would occur if test prompts or tasks were disseminated. Test developers and test users should describe all measures implemented to minimize such test security threats.

SECTION THREE: THE PROCESS FOR REVIEWING ASSESSMENTS

California Code of Regulations, title 5, section 55522 vests the California Community Colleges Chancellor's Office with the responsibility to approve any assessment test used in placing students in English, mathematics, or English as a Second Language (ESL) courses at the colleges. Section 55522 also requires the Chancellor's Office to establish at least annually a list of approved tests.

The Chancellor's Office is responsible for reviewing tests used for course placement in the community colleges. It is advised and aided by its Assessment Advisory Committee and psychometric experts in this review process and in making decisions regarding approval of placement tests for the colleges. This section outlines the review process.

Step 1. Request for Assessment Approval

A formal written request for assessment approval must be submitted by either a local college developing/managing a test or a second-party test publisher. Appendix C provides an example of the request form that a college must submit. Second-party test publishers must include a cover letter with their application instead of the request form as their requirements for submission differ. Depending on the applicant (college or test publisher) and the type of application (initial submission, follow-up submission, or renewal submission), application materials must sufficiently address relevant criteria described in Section 4. A flowchart and tables outlining criteria for different submission types are provided in Appendix C.

Applications should be sent to the Chancellor's Office before the published submission deadlines to be considered for approval (at assessmentadvisory@cccco.edu). The annual review schedule and required form are published on the Chancellor's Office website. Applications received after the published deadlines may be accommodated if scheduling permits; however, this is not guaranteed. Applicants should anticipate that late submissions will be reviewed during the next semiannual cycle. The corresponding list of approved tests is updated after completion of each review cycle.

Step 2. Preliminary Psychometric Expert Review

The information submitted to the Chancellor's Office will be reviewed by at least two psychometric experts with doctoral degrees in a measurement-related area or who have had five or more years of experience in tests and measurement. They must also have a broad understanding of both theoretical and applied issues for testing. Their evaluation of the instruments is based primarily on the evaluation criteria specific to the California Community Colleges as described in the following sections. Reviewers may also use other guidelines that are commonly accepted by the psychometric profession, such as the Joint Standards, the Code of Fair Testing, and the EEOC Guidelines.

Although much of the documentation on a test instrument is likely to be supplied by the instrument's developer and made available in the assessment approval request, the psychometric experts contracted by the Chancellor's Office may compile additional information relevant to the instrument including, but not limited to, technical reports, test reviews conducted by an independent third party (such as those found in the Mental Measurement Yearbook (MMY) series for second-party tests), review articles that are published in professional journals and books, and technical reports prepared by California community college users or the Chancellor's Office. The quality of the recommendation made by the Assessment Advisory Committee to the Chancellor's Office depends upon the quality of the information considered in the review for approval. Consequently, as much information as possible should be compiled and deliberated.

Upon completing an initial review of the application for approval, the Chancellor's Office's psychometric experts will submit a preliminary assessment evaluation to the applicant (i.e., local college or test publisher) as well as the Chancellor's Office. The applicant has up to 14 days to respond to this preliminary report if it wants to amend its application. However, the Chancellor's Office may reduce the response time if it is necessary to streamline the review process for the specific application cycle. Responses are limited to clarifications of the data previously submitted and/or to additional information that already exists but was not provided by the applicant in the initial submission. Responses should not involve assembling, analyzing, and reporting "new" data gathered in response to the preliminary evaluation.

Upon receiving additional information from the applicant, the psychometric experts may revise the preliminary assessment evaluation. The updated preliminary assessment evaluation will be submitted to the Assessment Advisory Committee to be considered in Step 4.

Step 3. Content Expert Review (For Initial Review of Second-Party Tests)

For the initial review of a new second-party test, the instrument will be reviewed by at least two subject matter experts in each pertinent content area. These subject matter reviewers are California community college faculty members who are knowledgeable about the specific content area courses the instrument assesses.

The content experts must evaluate all of the following:

1) The rationale underlying the assessment instrument stated by the applicant.
2) The items on the instrument.
3) The suggested interpretation of scores relative to the instrument's intended use.
4) The appropriateness of content for the diverse populations served in the California community colleges.

These content reviewers will submit a written report to the Assessment Advisory Committee to be considered in the next step.

Step 4. Assessment Advisory Committee Review

The Assessment Advisory Committee consists of a cross-section of community college faculty, staff, and administrators with expertise in assessment, research, teaching, or instruction. Appendix D presents the Charter of the Assessment Advisory Committee. The Assessment Advisory Committee will review the approval request, test materials and relevant documents, the preliminary assessment evaluation prepared by the psychometric experts, and the review report prepared by the content experts if the submission is a new second-party test. The committee may solicit additional information from test developers or test users, if necessary.

The Assessment Advisory Committee review culminates in a final recommendation report to be written and submitted to the Chancellor's Office. The report summarizes the key points of the application as well as the resulting analysis of the committee. The recommendation will be in one of the following four categories:

- Full Approval
- Provisional Approval
- Probationary Approval
- Not Approved

Only test instruments receiving Full, Provisional, or Probationary Approval may be used by the colleges. The length of time colleges may use a test varies by level of approval. However, a test may only maintain Provisional Approval or Probationary Approval for no more than three (3) years in combination. That is, tests will not maintain approval unless Full Approval is attained within three years. In applying for Full Approval, new evidence to support this designation must be submitted.

The categories for approval are further described below.

Full Approval
Test instruments in this category fully meet **all** relevant standards and criteria. The available evidence indicates a high probability of yielding test scores useful in assisting decision making for a particular community college student.

Provisional Approval
Test instruments in this category meet most but not all relevant standards and criteria, and the tests lack sufficient or recent information to assign the unequivocal Full Approval rating. The expectation of the instrument with Provisional Approval is that the necessary clarifying information to attain Full Approval can and will be provided within one academic year. Failure to submit the required data and/or clarification within one year will result in reclassification

into either a two-year Probationary Approval or the Not Approved category (if the test had Probationary Approval prior to the current Provisional Approval).

Probationary Approval

Test instruments in this category are missing critical information, or noticeable deficiencies are found in the documentation provided. The intended use of these instruments is clearly stated, and some positive information supporting its use is available, but the necessary evidence available for a final judgment is incomplete. To attain this minimal approval standard, the test must satisfy at least one form of validity as well as the fairness/test bias standard. Instruments can only maintain Probationary Approval for a maximum of two years. Failure to submit the required satisfactory evidence within two years will result in reclassification into the Not Approved category.

Not Approved

Test instruments in this category have failed to meet one or more of the essential standards (validity and fairness/test bias as well as a plan to address disproportionate impact) or have failed to meet a condition of title 5.

Step 5. Chancellor's Office Decision

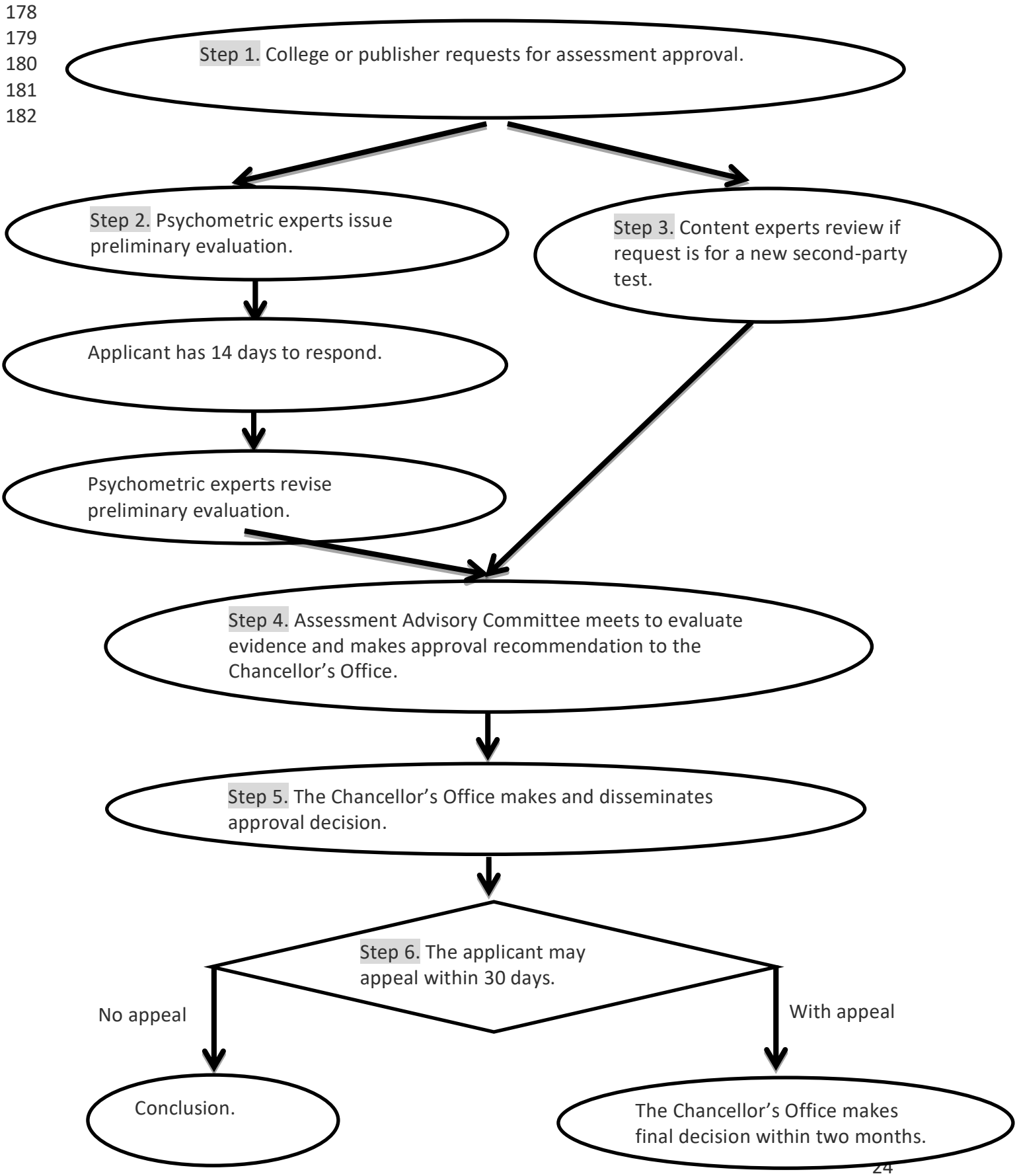
The Chancellor's Office will make the final decision regarding approval. Per title 5, section 55522, instruments that are not on the Chancellor's list of approved tests must not be used for course placement in the California community colleges. The list of approved test instruments is posted on the Chancellor's Office website. The applicant will receive a copy of the final evaluation report and be notified by the Chancellor's Office regarding the approval decision.

Step 6. Appeals Process

A decision by the Chancellor's Office may be appealed by any applicant for approval. Requests for an appeal must be submitted in writing to the Chancellor's Office within 30 days of notification of the Chancellor's Office decision. The request must clearly explain why the decision is being challenged. The Chancellor's Office may determine the next course of action, including the option of convening an appeals workgroup, consisting of Assessment Advisory Committee members, to reconsider the status. The Chancellor's Office must make a determination regarding the appeal within two months of the request.

Validation and documentation of instrument quality is considered an ongoing process; therefore, publishers and colleges are expected to continuously monitor and evaluate their test instruments. Further, once any approval status is attained, that instrument is "approved" for a period not to exceed six years. Well before the end of this six-year period of approval, new supporting materials or documentation must be submitted or the instrument will lose approval. Therefore, second-party publishers with test instruments nearing completion of six years of approval status must resubmit information and documentation during the fifth year

of approval so that continued use can be maintained by colleges. Similarly, colleges with locally developed or managed test instruments are encouraged to resubmit information and documentation during the fifth year of approval, but they may elect to wait to resubmit no later than midway through the sixth year of approval. Section Six of this document provides additional details for the “renewal” of a test’s approval status.


Step 1. College or publisher requests for assessment approval.
Step 2. Psychometric experts issue preliminary evaluation.
Step 3. Content experts review if request is for a new second-party test.
Applicant has 14 days to respond.
Psychometric experts revise preliminary evaluation.
Step 4. Assessment Advisory Committee meets to evaluate evidence and makes approval recommendation to the Chancellor's Office.
Step 5. The Chancellor's Office makes and disseminates approval decision.
Step 6. The applicant may appeal within 30 days.
No appeal
With appeal
Conclusion.
The Chancellor's Office makes final decision within two months.

SECTION FOUR: SPECIFIC CRITERIA FOR TEST USAGE

Section TWO of this document summarizes and abstracts portions of the Joint Standards that are most relevant to the California community college context and provides the general framework for test evaluation. Most tests used by California community colleges are intended to help students enroll in appropriate courses. The tests serve a placement purpose, offering students guidance as to whether they should enroll in a course at the beginning of the sequence of courses in mathematics, for example, or somewhat later in the sequence. Because of the specific nature of these measures and their common use within the California community colleges, more explicit criteria can be written.

In this section, specific criteria for reviewing placement tests are provided to tailor the standards for this specific use of tests. These criteria apply to specific parties responsible for test development and management in the California community colleges. Specifically, two types of tests are differentiated:

- Tests developed or managed by a California community college or district (referred to as “locally developed/managed tests”, and
- Tests developed and maintained by a second-party external to the California community colleges (referred to as “second-party tests”).

The locally developed/managed tests include two types of tests: (1) tests developed by a California community college or district, or (2) tests developed by an independent vendor and not approved by the Chancellor but whose use is deemed appropriate by a college or district, in which case that institution assumes responsibility for bringing the test into compliance with the standards as a locally managed test.

In the case of a locally developed/managed test, a local college (or district) takes on the role of both test developer and test user. In the case of a second-party test, the test vendor is the test developer while local colleges are test users. Therefore, the responsibilities of test developers and those of test users with regard to a second-party test will be presented in separate subsections.

Specific criteria have also been developed for computer-based assessments, including both computer-administered tests and computer-adaptive tests. Advancements in technology and psychometric models in the past decades permit further flexibility of testing; thus, computer-based assessments require additional considerations due to their distinctive testing features.

In summary, the remainder of this section is divided into four subsections:

I. Specific Criteria for Locally Developed/Managed Tests
II. Specific Criteria for Second-Party Tests: Primary Responsibility of the Test Developer
III. Specific Criteria for Second-Party Tests: Primary Responsibility of the Local College or District as Test User
IV. Additional Criteria for Computer-Based Testing

It should be noted that the specific criteria detailed in the following subsections establish minimum requirements necessary to gain an approval status. In general, the expectation is that **all** specific criteria should be met. Under no circumstances can a test instrument gain approval for use without meeting the minimum requirements regarding validity and test fairness. At a minimum, **there must be evidence of the test's validity for the intended purpose; and there must be evidence that the test minimizes cultural/linguistic bias, insensitivity, and offensiveness**. In addition, the college should have a plan to mitigate any disproportionate impact that is identified for any student groups. In the absence of such information, the test will not be approved for use in the CCC.

In general, the appropriate use of a test needs to be considered in light of all applicable standards described in the Joint Standards, the Code for Fair Testing, and the EEOC Guidelines. Consequently, meeting the specific criteria described in the following subsections may not be sufficient to receive a favorable recommendation for using a test instrument. It should also be noted that evaluating the appropriateness and usefulness of a particular assessment instrument is an **ongoing** activity. As student populations, the nature of a test, course prerequisites, and/or placement sequences change over time, it will be necessary for test developers and test users to reevaluate the instrument and its use.

I. Specific Criteria for Locally Developed/Managed Tests

The responsibilities for colleges that locally develop or manage placement tests are delineated in this section. At a minimum, these colleges must provide acceptable documentation addressing **all** of the following:

- Fairness based on bias review findings (for a locally developed test) or adequate fairness evidence provided by the test publisher or another college (for a locally managed test).
- Content validity based on an item-by-item evaluation.
- A brief description of and rationale for the initial setting of cut scores (however, cut score validity evidence is not required to obtain a minimum/probationary level of approval).
- A plan to mitigate any disproportionate impact that is identified for any student groups.

However, meeting the minimum requirements is not sufficient to attain Full Approval status. Colleges that locally develop or manage tests should carefully consider all requirements in this section for Full Approval.

1. Fairness

a. Logical review of bias, insensitivity, and offensiveness

Evidence focusing on the lack of cultural and/or linguistic bias, insensitivity, and offensiveness must be provided. This evidence must involve evaluations of test items by diverse panels of people who demographically represent the college's student population. Federal and state anti-discrimination laws and guidelines should be consulted for identifying relevant protected classes. As a general rule, when a protected class (as designated by race, gender, age, and disability) constitutes at least two percent of the student population, that group should be represented in the investigation. For tests of English language proficiency used for the English learner population (e.g., English as a Second Language or ESL tests), the linguistic and cultural background of test takers must also be considered.

The bias review guidelines, procedures, and training must be described in the application for approval. A description of the panel members' qualifications and demographic representations must also be included. At least two reviewers representing each protected class must be involved in the logical review, and a reviewer can represent more than one class (e.g., gender and ethnicity). Faculty, staff, students, and community participants can serve as reviewers. However, faculty who are involved in the test's development and/or item writing should not participate in the logical review. The determination of potential bias from these investigations must be used to eliminate or minimize sources of test/item bias, insensitivity, and offensiveness.

When a college is managing a second-party instrument for which there is adequate and representative fairness evidence from the test publisher or from another college's study, additional data from the college is not required. In this case, the fairness evidence and the sources should be cited in the approval request. Colleges are encouraged, but not required, to supplement the existing evidence with logical reviews conducted locally.

When assessment procedures are in place to minimize potential bias—for example, by providing students a choice of questions, prompts, or tasks—then the procedure must be clearly stated and described.

b. Disproportionate impact

Disproportionate impact must be monitored on an ongoing basis for various demographic groups (including race, gender, age, and disability). Disproportionate impact must be summarized and reviewed by pertinent college staff at least every three years. Federal and state anti-discrimination laws and guidelines should be consulted for identifying relevant protected classes. For tests of English language proficiency used for the English learner population, the linguistic and cultural background of test takers must also be considered. Colleges must update their disproportionate impact investigations when there has been a significant change in student demographics at the college.

When disproportionate impact is observed, the college shall, in consultation with the Chancellor's Office, develop and implement a plan for addressing the disproportionate impact, including studies of differential prediction. Colleges may consult the California community colleges document titled Ensuring Equitable Access and Success: A Guide to Assessing & Mitigating Disproportionate Impact in Student Success and Support Programs (Aug, 2013) for further details on the definition, identification, and treatment of disproportionate impact. The requirement that disproportionate impact is to be continuously monitored must not be overlooked and colleges should be aware that they will be held to the standard of having plans to mitigate disproportionate impact in assessment.

The initial submission for a new test only needs to include a plan for data collection and analysis to monitor disproportionate impact. Actual data and findings are not expected for the first submission, although disproportionate impact information is welcome if available.

When a test is nearing the end of its initial six-year approval cycle is submitted for approval under the renewal process as described in Section SIX, the college must submit disproportionate impact findings along with mitigation plans or actions (if any disproportionate impact was found). Full approval of a renewal application can only be granted when:

- disproportionate impact findings are reported and judged acceptable, or

- appropriate actions and/or plans to address significant disproportionate impact findings are sufficiently described by the college.

c. Standardization

If the test is revised for testing individuals who cannot take it under standard conditions, there must be documentation of all changes along with the basis for each accommodation. The justification for changing or altering assessment instruments or procedures must be kept on file at the local college.

2. Validity

a. Content-related validity evidence

The college must provide a comprehensive description of the appropriateness of a placement test for a course sequence based on the overlap of the **knowledge and skills measured by the test and the prerequisite knowledge/skills for courses in the sequence**. Content-related validity evidence must be grounded in statements of specific pre-course expectations (i.e., prerequisite skills) that are then linked to the actual tested skills. Information addressing the extent of the alignment between prerequisite skills and the specific content measured by the test provides strong rational evidence of the content representativeness and relevance of the test for the course sequence. In addition, the college must comply with title 5, section 55003 (d), which addresses the allowable purposes for establishing a prerequisite.

Procedurally, local college faculty are to evaluate the content representativeness by participating in an **item-by-item evaluation** of the test content considering the prerequisite skills for each course in the placement sequence. When the test is a performance assessment (e.g., writing samples), the evaluation is with regard to prompts, tasks, and rubrics. At least one instructor (with current or recent teaching assignment) for each course in the sequence should be included in this content review while the involvement of multiple instructors is encouraged. To the extent possible, faculty who were involved in item writing or any other part of the test development process should not serve on the content review panel.

Summary information that describes the process and results of the content review must be provided. Results must include tabulations addressing the degree of match between the prerequisite skills and the assessed content/skills. The necessary prerequisite skills not measured by the instrument must be noted, as well as those skills/content tested that are not relevant to the prerequisite skills. In other words, information presented on content-related validity evidence must address the following questions:

- Does the instrument measure all prerequisite skills that are necessary for appropriate course placement?
- Are there sufficient items and coverage to adequately assess each prerequisite skill?

- Does the instrument assess skills and knowledge that are not prerequisite skills?

The greater the degree of the overlap between course prerequisite skills and those skills measured by the test, the stronger the evidence in support of content-related validity. The extent to which a test measures irrelevant (non-prerequisite) skills must be documented and considered when judging the appropriateness of the test. When a test can be used to "test out" of a course or course sequence (e.g., English as a Second Language), a content-related validity evaluation must provide evidence of the content match between the instrument and the objectives of the alternative course (e.g., native English course).

For direct performance assessments, descriptions of how prompts/tasks and scoring rubrics were developed must be included in the initial submission in addition to the alignment results. A copy of the rubrics must be provided. Including the actual prompts/tasks is at the applicant's discretion, but the test review process may be facilitated if the prompts/tasks are provided. A description of how raters are trained to administer the instrument/task and apply the scoring rules consistently must be documented. Additionally, the methods used for resolving inconsistencies between scorers should be described and justified.

b. Criterion-related or consequential-related validity evidence

Evidence addressing criterion-related or consequential-related validity need only be collected if such a design is implemented in order to provide the empirical validation of local cut scores. (See subsequent Subsection I.2.c.)

c. Evidence addressing adequacy of cut scores

It is the local community college's responsibility to validate its cut scores. Data are to be collected by the individual college to justify the selection of any cut scores or score ranges used for placement recommendations. The adequacy of any cut score may be demonstrated by either a judgmental or an empirical approach. Arbitrary decisions about cut scores (e.g., passing with 70 percent of items correct) are not based on empirical evidence or informed judgments and are therefore not acceptable. However, setting initial cut scores on the basis of an item-by-item analysis of relevance of each test item to "pre-course" skills is an acceptable judgmental methodology provided empirical evidence is later collected to support or modify the cut scores.

A judgmental approach typically focuses on setting the initial cut scores. However, if such judgmental data are to be used as the only evidence to support the adequacy of the cut scores, the college must use a systematic procedure that can be found in the cut-score setting literature (for example, see Cizek & Bunch, 2007, and Zieky, Perie, & Livingston, 2008, for appropriate procedures). In addition to the process and the results, the college must also document the persons involved and their credentials. Individuals involved in this process should be familiar with student learning in the courses in the placement sequence.

When an empirical approach is used to determine cut scores, the results must, at a minimum, demonstrate that individuals who score above the cut score or within a score range have a statistically significant expectancy of success in a specific course for which placement recommendations are made than those who score below the cut score or score range. The "success" may be defined as the appearance of preparation for the course based on instructor ratings, a mid-term grade, or a final course grade of C or higher. Other operational definitions for success may be articulated by the college. Either criterion-related or consequential-related evidence is appropriate for meeting this standard with the determination as to which evidence is most appropriate being based on the following general principle:

- Criterion-related studies are appropriate for a new test that has not yet been approved and whose test scores have **not** been used for course placement.
- Consequential-related studies are appropriate for a previously approved test whose test scores have been used for course placement.

Determining the methods and procedures for carrying out cut score validation studies is a local college decision. However, colleges are expected to follow proper and reasonable investigative approaches, including obtaining sufficient sample sizes and maintaining objective judgments (or impartiality) by using double blind experimentation.

For criterion-related validity studies, a variety of research designs are acceptable including mean comparisons or correlational designs. Appropriate criterion variables may include, but are not limited to, any of the following:

- Student ratings of ability to meet course requirements.
- Instructor ratings of students' abilities to meet course requirements.
- Midterm grades or test scores.
- Final course grades or test scores.

When used as the primary index for criterion-related validity, the coefficient of the correlation between the test score and the criterion must be greater than or equal to .35 (or a comparable effect size if an alternative statistical analysis was performed). Coefficients corrected for range restrictions (either on the test score, on the criterion, or both) are acceptable if a rational foundation is presented for their use.

For consequential-related validity studies, the following research questions are to be considered:

(i) After the first few weeks of a course (e.g., between the fourth and sixth weeks), how do students whose test scores recommended placement into that class evaluate the appropriateness of their course placement (e.g., placed in the correct/proper course, should have been placed in a higher course, should have been placed in a lower course)? The standard is at least 75 percent affirmative endorsement by students.

(ii) After the first few weeks of a course (e.g., between the fourth and sixth weeks), how do instructors evaluate the readiness to undertake the material of the course for individual students whose test scores recommended placement into that course? The standard is at least 75 percent of students are considered properly placed by instructors.

(iii) For students who choose not to follow a test's course placement recommendation, how do these students fare (in terms of material learned, suitability of the placement, and their likelihood of successful matriculation) in the classes into which they choose to enroll, and can such performance be justified/expected?

(iv) What do students and instructors identify as undesirable results of an "incorrect" course placement and what are the consequences (for students, instructors, academic units, and the institution) of such decisions?

Under any approach chosen by a college for investigating consequential-related validity, at a minimum, items (i) and (ii) above must be formally addressed and satisfied for the instrument to be fully approved. Items (iii) and (iv) are optional and supplementary; although they are not required, they can be extremely useful sources of information to colleges that may choose to pursue such lines of inquiry. Other supplemental research questions are possible, and comparable investigative orientations that involve students and instructors are encouraged.

Consequential-related validity evidence results should be reported separately for each course in the placement sequence as well as a cumulative result for the entire sequence.

A minimum sample size of 30 students is required for each course in a consequential-related validity study. When sample sizes are small due to low enrollment via assessment, colleges must attempt to collect information from students and instructors over multiple terms and report results with both cumulative data and single-term data. In addition, the data must be collected within a recent three-year period. If sample sizes are still too small even after collecting data over a recent three-year period, colleges may submit analyses using recent three-year data regardless of the small sample size. However, in this case, colleges must document in writing their efforts to collect, analyze, and monitor all available student data and use due diligence to ensure fair and proper test usage.

d. Validity evidence for subscores

When subscores are used to assist placement recommendations, validity evidence (Subsections I.2.a. through I.2.c.) must be demonstrated for each subscore.

3. Reliability and Errors of Measurement

The local community college should attempt to evaluate all relevant sources of measurement errors. The sources of measurement errors may include score variability over testing occasions, item sampling, alternate forms, or scorers. At least one form of reliability must be provided for Full Approval status. Multiple forms of reliability evidence may be required

depending on the nature of the test (e.g., direct performance assessment) and its score usage. The minimum sample size for a reliability analysis is 50 students.

a. Variability over testing occasions (test-retest reliability)

Test score stability may be assessed by administering the test on two occasions to the same student sample. The design may involve administering the same test form twice (test-retest approach) or using alternate test forms for the two administrations (equivalent-form approach). In order to assess stability, the two testing occasions must be at least two weeks apart. The resulting correlation coefficients between the test scores from two administrations must be .75 or higher.

When a college locally manages a second-party test, the stability coefficients from the test publisher or another college may be cited to meet this requirement.

b. Variability over items (internal consistency reliability)

Internal consistency reliability represents the agreement among test items measuring the same construct. Evidence documenting the internal consistency reliability must be based on appropriate analyses (e.g., alpha coefficient, split-half coefficient, Kuder-Richardson index, or indices based on test information curves). The minimum acceptable value for these internal consistency indices is .80.

When a college locally manages a second-party test, the internal consistency reliability coefficients from the test publisher or another college may be cited to meet this requirement.

c. Variability over parallel forms (equivalent-form or inter-prompt reliability)

When there are multiple test forms, question sets, prompts, or tasks used concurrently to produce interchangeable test scores for placement recommendations, the score variability over those forms/question sets/prompts/tasks must be evaluated and reported. A college may use correlation coefficients to evaluate the equivalency between scores on different forms/question sets/prompts/tasks, and the correlation coefficients must be .75 or higher. As an alternative approach, colleges may conduct random assignments of forms/question sets/prompts/tasks and evaluate the comparability of the resulted score distributions.

When a college locally manages a second-party test, the parallel-form reliability coefficients from the test publisher or another college may be cited to meet this requirement.

d. Variability over raters (inter-scorer reliability)

For direct performance assessments (e.g., writing samples), where score assignments involve subjective judgments, inter-scorer consistency must be evaluated to support the usage of these assessments:

- If inter-scorer correlation coefficients are provided, these coefficients must be .70 or higher.
- If percent agreement indices are provided, they must indicate a high level of consistency. For example, if a 6-point scale is used, it is expected to have at least 90 percent score agreement within a 1-scale-point difference.
- If Cohen's Kappa coefficients (a chance-corrected agreement index) are provided, they must be .40 or higher.
- When appropriate, other recognized chance-corrected agreement indices can also be used to support inter-score reliability.

A description of how raters are trained to administer the instrument/task and apply the scoring rules consistently must be documented. Additionally, the methods used for resolving inconsistencies between scorers should be described and justified.

Colleges **cannot** cite inter-score reliability coefficients from the test publisher or another college to meet this requirement.

e. Standard errors of measurement (SEM)

Standard errors of measurement must be provided for intervals across the score scale or at likely cut points.

f. Reliability of subscores

When subscores are used to assist placement recommendations, reliability evidence (Subsections I.3.a. through I.3.e.) must be demonstrated for each subscore.

II. Specific Criteria for Second-Party Tests: Primary Responsibility of the Test Developer

Second-party publishers' responsibilities are delineated in this section. At a minimum, second-party publishers must provide acceptable documentation addressing **all** of the following:

- Fairness based on **both** empirical and logical analysis findings.
- Specificity of test objectives and test content.
- Acceptable criterion-related or consequential-related validity evidence.
- Accessibility of the testing instruments to students with disabilities or alternative accessible arrangements that are offered for such individuals.

However, meeting the minimum requirements is **not** sufficient to attain Full Approval status. Second-party publishers should carefully consider all requirements in this section for Full Approval.

1. Fairness

a. Logical and empirical review of bias, insensitivity, and offensiveness

Evidence focusing on the lack of cultural and/or linguistic bias, insensitivity, and offensiveness must be provided. Federal and state anti-discrimination laws and guidelines should be consulted for identifying relevant protected classes. As a general rule, when a protected class (as designated by race, gender, age, and disability) constitutes at least two percent of the California community college student population, that group should be represented in the investigation. For tests of English language proficiency used for the English learner population (i.e., English as a second language or ESL tests), the linguistic and cultural background of test takers must also be considered. Test publishers must provide fairness evidence from two types of studies: "logical review" and "empirical review."

In a logical review, test items, prompts, and scoring rubrics must be reviewed by diverse panels of people who demographically represent the California community college's student population. A description of the panel members' qualifications and demographic representations must be included. At least two reviewers representing each protected class must be included in the logical review. Test publishers, employees of test publishers, persons who are involved in test development or item writing, or persons who may have a conflict of interest may not participate in logical reviews. The review guidelines, procedures, and training should be described.

An empirical review of test data and item performance (e.g., differential item functioning) must be conducted on data from students similar to those ordinarily served by the California community colleges. Analysis must be based on sufficient sample sizes for protected classes. Follow-up investigation must be described if any item is flagged by the empirical review.

The determination of potential bias from these investigations must be used to eliminate or minimize sources of test/item bias, insensitivity, and offensiveness. Item removal must be documented and justifications must be provided for item retention or revision.

When assessment procedures are in place to minimize potential bias—for example, by providing students a choice of questions, prompts, or tasks—then the procedures must be clearly stated and described.

b. Testing special groups

Test publishers must agree to provide the test and response forms in alternate media upon request of a college. Testing instruments must be accessible to students with disabilities or publishers must offer alternative accessible arrangements for such individuals (e.g., braille, large print, audiotape, and electronic tests). Such a commitment is required for a test to receive any level of approval from the Chancellor's Office.

2. Validity

a. Content-related validity evidence

Explicit statements of test objectives and table of specifications must be available to inform test users. Test publishers must describe the test content (items and item formats) with sufficient and clear specificity for colleges to evaluate the test's appropriateness for making placement recommendations at their colleges.

Upon request, test publishers must provide an operational test booklet or, in the case of computer-adaptive tests, a sufficient representative sample of operational test items to enable local colleges to conduct an item-by-item review. Retired items or forms are not acceptable. If a sample of items is provided, the number of items must represent a psychometrically sound single form of a traditional fixed form of the test.

b. Criterion-related or consequential-related validity evidence

Data must be presented to indicate that the test is useful for making placement recommendations into courses offered in the California community colleges. Empirical evidence must support the following conclusion: Test takers who achieve scores within some specified range should enroll in a different course or set of courses in comparison with test takers who score outside that range. When submitting evidence to meet this standard, the course content must bear a close logical relationship with courses offered by the California community colleges, and the student samples must be demographically representative of the students ordinarily served by the California community colleges.

Supportive data from at least **six** community colleges in six different districts representing the diversity of courses and students in the California community colleges are required to attain

Full Approval status; supportive data from at least **four** community colleges in four different districts are required to attain Provisional Approval status; and supportive data from at least **three** community colleges in three different districts are required to attain Probationary Approval status. **Additionally, a majority of the colleges included must be California community colleges** to attain a specific approval level status (i.e., 4 of 6, 3 of 4, or 2 of 3 respectively for Full, Provisional, and Probationary Approval status). An instrument can be piloted to gather needed information prior to approval but it cannot be used for course placement purposes until it receives at least Probationary Approval from the Chancellor. Until some level of approval is received, colleges are prohibited from using an unapproved test for making placement recommendations.

Either criterion-related or consequential-related evidence is appropriate for meeting this standard with the determination as to which evidence is most appropriate being based on the following general principle:

- Criterion-related studies are appropriate for a new test that has **not** yet been approved and whose test scores have **not** been used for course placement.
- Consequential-related studies are appropriate for a previously approved test whose test scores have been used for course placement.

For criterion-related validity studies, a variety of research designs are acceptable including mean comparisons or correlational designs. Appropriate criterion variables may include, but are not limited to, any of the following:

- Student ratings of ability to meet course requirements
- Instructor ratings of students' abilities to meet course requirements
- Midterm grades or test scores
- Final course grades or test scores

When used as the primary index for criterion-related validity, the coefficient of the correlation between the test score and the criterion must be greater than or equal to .35 (or a comparable effect size if an alternative statistical analysis was performed). Coefficients corrected for range restrictions (either on the test score, on the criterion, or both) are acceptable if a rational foundation is presented for their use.

For consequential-related validity studies, the following research questions are to be considered:

(i) After the first few weeks of a course (e.g., between the fourth and sixth weeks), how do students whose test scores recommended placement into that class evaluate the appropriateness of their course placemen t(e.g., placed in the correct/proper course, should have been placed in a higher course, should have been placed in a lower course)? The standard is at least 75 percent affirmative endorsement by students.

(ii) After the first few weeks of a course (e.g., between the fourth and sixth weeks), how do instructors evaluate the readiness to undertake the material of the course for individual students whose test scores recommended placement into that course?

The standard is at least 75 percent of students are considered properly placed by instructors.

(iii) For students who choose not to follow a test's course placement recommendation, how do these students fare (in terms of material learned, suitability of the placement, and their likelihood of successful matriculation) in the classes into which they choose to enroll, and can such performance be justified/expected?

(iv) What do students and instructors identify as undesirable results of an "incorrect" course placement and what are the consequences (for students, instructors, academic units, and the institution) of such decisions?

Under any approach chosen by a college for investigating consequential-related validity, at a minimum, items (i) and (ii) above must be formally addressed and satisfied for the instrument to be fully approved. Items (iii) and (iv) are optional and supplementary; although they are not required, they can be extremely useful sources of information to colleges that may choose to pursue such lines of inquiry. Other supplemental research questions are possible, and comparable investigative orientations that involve students and instructors are encouraged.

Determining the methods and procedures for carrying out validation studies is the publisher's decision. However, publishers are expected to follow proper and reasonable investigative approaches, including obtaining sufficient sample sizes and maintaining objective judgments (or impartiality) by using double blind experimentation. Publishers may report aggregated results across colleges; however, publishers must also separately report findings for each participating college. Further, a minimum sample size of 30 students is required for each course in a consequential-related validity study.

c. Validity evidence for subscores

When subscores are used to assist placement recommendations, validity evidence (Subsections II.2.a. through II.2.b.) must be demonstrated for each subscore.

3. Reliability and Errors of Measurement

Test publishers are responsible for evaluating all relevant sources of measurement errors. The sources of measurement errors may include score variability over testing occasions, item sampling, alternate forms, or scorers. Multiple forms of reliability evidence may be required depending on the nature of the test (e.g., direct performance assessment) and its score usage. However, the stability of test performance over testing occasions is particularly relevant for all placement tests in the California community colleges; therefore, test-retest reliability must be evaluated by test publishers. The minimum sample size for a reliability analysis is 50 students.

a. Variability over testing occasions (test-retest reliability)

Test score stability may be assessed by administering the test on two occasions to the same student sample. The design may involve administering the same test form twice (test-retest approach) or using alternate test forms for the two administrations (equivalent-form approach). In order to assess stability, the two testing occasions must be at least two weeks apart. The resulting correlation coefficients between the test scores from two administrations must be .75 or higher.

b. Variability over items (internal consistency reliability)

Internal consistency reliability represents the agreement among test items measuring the same construct. Evidence documenting the internal consistency reliability must be based on appropriate analyses (e.g., alpha coefficient, split-half coefficient, Kuder-Richardson index, or indices based on test information curves). The minimum acceptable value for these internal consistency indices is .80.

c. Variability over parallel forms (equivalent-form or inter-prompt reliability)

When there are multiple test forms, question sets, prompts, or tasks used concurrently to produce interchangeable test scores for placement recommendations, the score variability over those forms/question sets/prompts/tasks must be evaluated and reported. When correlation coefficients are used to evaluate the equivalency between scores on different forms/question sets/prompts/tasks, the correlation coefficients must be .75 or higher.

d. Variability over raters (inter-scorer reliability)

For direct performance assessments (e.g., writing samples), where score assignments involve subjective judgments, inter-scorer consistency must be evaluated to support the usage of these assessments:

- If inter-scorer correlation coefficients are provided, these coefficients must be .70 or higher.
- If percent agreement indices are provided, they must indicate a high level of consistency. For example, if a 6-point scale is used, it is expected to have at least 90 percent score agreement within a 1-scale-point difference.
- If Cohen's Kappa coefficients (a chance-corrected agreement index) are provided, they must be .40 or higher.
- When appropriate, other recognized chance-corrected agreement indices can also be used to support inter-score reliability.

Additionally, the methods used for resolving inconsistencies between scorers should be described and justified.

e. Standard errors of measurement

Standard errors of measurement must be provided for intervals across the score scale or at likely cut points.

f. Reliability of subscores

When subscores are used to assist placement recommendations, reliability evidence (Subsections II.3.a. through II.3.e.) must be demonstrated for each subscore.

III. Specific Criteria for Second-Party Tests: Primary Responsibility of the Local College or District as Test User

When using a second-party test, each college retains the responsibilities for documenting fairness, validity, and reliability of test scores and how they are used locally. For each course sequence in which the test is used for placement, the college must maintain an assessment portfolio containing the following evidence to support the test as appropriate and valid:

- Fairness evidence (entails a review of the test publisher's evidence addressing test bias)
- Validity evidence
- Reliability evidence
  - For direct performance assessments, reliability studies are required.
  - For all other tests, a review of the test publisher's reliability evidence is required.

The assessment portfolio must be updated at least every six years, with the exception of disproportionate impact analyses that must be updated at least every three years. Assessment portfolios do not need to be submitted to the Chancellor's Office.

1. Fairness

a. Review of bias, insensitivity, and offensiveness

Local community colleges must review the evidence addressing test bias supplied by the test publisher to ensure that the results are generalizable to the student demographics at their colleges.

b. Disproportionate impact

Local community colleges must conduct a disproportionate impact study according to the approach described in Section FOUR, Subsection I.1.b. (see p. 28). The disproportionate impact study must be updated every three years.

c. Standardization

If the instrument is revised for testing individuals who cannot take the test under standard conditions, there must be documentation of all changes along with the basis for each accommodation. The justification for changing or altering assessment instruments or procedures must be kept on file at the local college.

2. Validity

a. Content-related validity evidence

Local community colleges must conduct a content-related validity study according to Section FOUR, Subsection I.2.a. (see p. 29).

b. Evidence addressing adequacy of the cut score(s)

Local community colleges must conduct a cut-score validity study according to Section FOUR, Subsection I.2.c. (see p. 30).

3. Reliability and Errors of Measurement

Local community colleges must review the evidence addressing reliability and standard errors of measurement supplied by the test publisher to ensure that the results are generalizable to their colleges.

For direct performance assessments, colleges must conduct local inter-scorer reliability studies to ensure score consistency on their campus.

IV. Additional Criteria for Computer-Based Testing

Assessments can differ in terms of test delivery mode (e.g., paper-and-pencil, oral, or computerized administration). Advancements in technology and psychometric models in the past decades permits further flexibility of testing and allows different forms to be administered to different students. Computer-based tests can be traditional fixed-form tests administered via computer, they can be assessments using complex algorithms to produce random test forms (i.e., linear-on-the-fly testing, or they can be assessments that produce unique test forms based on examinees' ability levels (i.e., computer-adaptive testing or CAT).

In most cases, the criteria presented in previous subsections (I, II, and III) apply to computer-based testing. However, because of the technology and often complex item types involved in computer-based tests, other criteria have been enhanced. Further, as the technology used in computer-based testing continues to evolve, test developers and users are advised to contact the Chancellor's Office to determine if there may have been changes to these specific criteria. Publishers are encouraged to present and discuss their methodology and psychometric results using non-technical language for prospective California community college users.

If a computer-based test was previously approved in a paper-and-pencil format, the prior approval status for the paper-and-pencil test does not automatically transfer to the computer version of the assessment. A primary means to accomplish this is through an evaluation of equivalent-form reliability. The equivalent-form reliability coefficient for the computer-based and paper-and-pencil forms must be at least .80. If this equivalent-form reliability is established, the computer-based test can receive some level of approval for use.

However, publishers are not limited to the equivalent-form design. Alternative data gathering and analysis models will be considered if a rationale is provided—and that rationale is considered valid by the Chancellor's Office. The analysis used must result in a high level of consistency between scores from the different formats. Alternatively, the computer-based version of a previously approved paper-and-pencil test can be submitted as a new application if the claim of test equivalency is not intended.

Computer-adaptive tests rely on item banks (also known as item pools) from which items are chosen as testing is initiated or underway (i.e., a variable collection of questions is administered to examinees). Publishers of these tests must demonstrate the quality of the item banks and the quality of the collections of items selected for administration to students. The publishers must also provide specifications of the item bank, documentation of the algorithms used to select items, and the method to reach a final student's result. This information is needed to ensure that items are representative of and appropriate for the content domain being evaluated.

Administrative manuals must address and detail the technical properties of computer-based tests. Considerations of score interpretation and scoring algorithms, how initiating items are

chosen, stopping rules, item exposure control, security, and any non-traditional administrations must be formally and completely described. Suitable accommodations for examinees with special needs must also be provided.

The manuals made available to test administrators must clearly and explicitly detail instructions specifically applicable to computer-based test administrations (e.g., how to save data, how to re-administer an assessment when there is a power failure or machine lock-up, and how to attend to examinees who evidence significant anxiety related to the testing format).

1. Fairness

a. Logical and empirical review of bias, insensitivity, and offensiveness

All operational items in the item bank must undergo review for bias, insensitivity, and offensiveness. The standards described in Section FOUR, Subsections I.1.a. (see p. 27) and II.1.a. (see p. 35) are applicable to locally developed/managed computer-based tests and second-party computer-based tests, respectively.

b. Technology availability and familiarity

Students' lack of familiarity with technology and complex/innovative item types must not unfairly impact their ability to demonstrate content-related skills with computer-based testing. Therefore, test developers must provide **all** of the following:

- Documentation of required and recommended technology infrastructures, devices, and software.
- Sample interface of testing modules.
- Description of test instructions to examiners and examinees (e.g., time limit, how to respond to different item types, whether items can be skipped and revisited later in the testing).
- Training and assistance for examinees to assure that the testing format does not interfere with the achievement estimate for the examinee.

c. Testing special groups

Testing instruments must be accessible to all students including students with disabilities. Test developers/publishers must provide alternative accessible arrangements (e.g., paper-and-pencil format, braille, large print, audiotape, and electronic tests) when requested. Such a commitment is required for a test to receive any level of approval from the Chancellor's Office.

2. Validity

a. Content-related validity evidence

The standards described in Section FOUR, Subsections I.2.a. (see page 29) and II.2.a. (see page 36) are applicable to locally developed/managed computer-based tests and second-party computer-based tests, respectively.

For all computer-based tests, including adaptive tests, the evidence for supporting content-related validity must also include the following:

- A test blueprint (or table of specifications) so local colleges know the construct being evaluated and how/which items (content and cognitive areas) are being used (distributed) on forms of the test to arrive at a score(s).
- Information for prospective college users on how to access all or a sufficiently large and representative sample of operational test items, performance tasks, and/or test forms.
- For a machine-scored direct performance assessment, a description of the scoring engine, including its development process, scoring criteria, rules, and algorithms. This description must include enough specifics so local colleges can evaluate the content validity of the machine-generated scores. In addition, example papers for each score point must be available to local colleges upon request.

In addition, **for computer-adaptive tests**, the supporting evidence must also include **all** of the following:

- Documentation of the item bank, including pool size and the composition of the pool in terms of construct/content/cognitive domain (i.e., all item features considered in the table of specifications) and psychometric properties.
- Goals of the item-selection mechanism and documentation of algorithms used for item selection during the entire test administration cycle and termination criterion.
- Discussion of rules or restrictions about content specifications, item bundles, enemy items, and exposure control, when applicable.
- When multistage testing or testlet-based designs are implemented, documentation of the routing of first-stage tests, the second-stage tests, and so on must be presented and justified for each stage. In addition, the decision rules moving the examinee on to stages/levels/testlets are to be presented in terms of cognitive and content specifications, and these rules must be justified.
- Sufficient representative samples of active (not retired) test items. Item samples may be in the format of a fixed parallel form of the test or examples of testing forms assembled by computer for students at different proficiency levels.
- When automated item generation is used, this design and methodology must be discussed at length and validation evidence presented that clearly supports the methodology.

- Documentation of the item bank maintenance plan, providing a timeframe and procedures for reviewing, refreshing, retiring, and replacing items in the item pool(s).
- If 20 percent or more of the item bank has changed (additions, removals, or revisions) since the test was last approved, the publisher is required to submit a new application within one year of when the changes reached the 20 percent threshold, even if it is prior to the six-year renewal deadline. A statement must be included in all application materials that the publisher acknowledges and will abide by this requirement.

b. Criterion-related or consequential-related validity evidence

The standards described in Section FOUR, Subsections I.2.b. (see p. 30) and II.2.b. (see p. 36) are applicable to locally developed/managed computer-based tests and second-party computer-based tests, respectively.

c. Validity evidence for subscores

When subscores are used to assist placement recommendations, validity evidence (Subsections IV.2.a. through IV.2.b.) must be demonstrated for each subscore.

d. Other validity evidence

For computer-adaptive tests, the supporting evidence must also include all of the following:

- A description of the classical test theory (CTT) or item response theory (IRT) model(s) used for constructing the test, including the explicit scoring algorithm and scaling employed. An examination of the model fit or dimensionality assumptions must be conducted to support the model(s) used for test construction.
- A description of item calibration studies (including sample size, sample characteristics, and year of data collection) for obtaining item parameter estimates.
- A description of equating/linking/concordance studies when applicable.

3. Reliability and Errors of Measurement

a. Variability over testing occasions (test-retest reliability)

The standards described in Section FOUR, Subsections I.3.a. (see p. 33) and II.3.a. (see p. 38) are applicable to locally developed/managed computer-based tests and second-party computer-based tests, respectively.

b. Variability over items (internal consistency reliability)

For fixed-form computer-based tests (i.e., non-adaptive tests), the standards described in Section FOUR, Subsections I.3.b. (see p. 33) and II.3.b. (see p. 39) are applicable.

For computer-adaptive tests or tests developed with IRT model(s), internal consistency may be demonstrated with IRT-based reliability estimate, test information function, or conditional standard error of measurement.

c. Variability over parallel forms (equivalent-form reliability)

The standards described in Section FOUR, Subsections I.3.c. (see p. 33) and II.3.c. (see p. 39) are applicable to fixed-form computer-based tests. However, equivalent-form reliability is not applicable to computer-adaptive tests and thus not required.

d. Variability over raters (inter-scorer reliability)

The standards described in Section FOUR, Subsections I.3.d. (see p. 33) and II.3.d. (see p. 39) are applicable to locally developed/managed computer-based tests and second-party computer-based tests, respectively.

If student work for a direct performance assessment is machine scored using artificial intelligence scoring algorithms, the consistency between computer scoring and expert human scorers must be demonstrated. The indices and benchmarks specified in Section FOUR, Subsections I.3.d and II.3.d are applicable.

e. Standard errors of measurement

The standards described in Section FOUR, Subsections I.3.e. (see p. 34) and II.3.e. (see p. 39) are applicable to locally developed/managed computer-based tests and second-party computer-based tests, respectively.

f. Reliability of subscores

When subscores are used to assist placement recommendations, the reliability evidence (Subsections IV.3.a. through IV.3.e.) must be demonstrated for each subscore in addition to the overall test score.

SECTION FIVE: CRITICAL MASS APPROVAL OF AN INSTRUMENT

The concept of "critical mass" pertains to situations where evidence on a specific assessment instrument has been accumulated across several California community colleges such that its approval status may be generalizable to other colleges. The principle of critical mass and criteria for its application in attaining approval are outlined in the following paragraphs.

Specifically, any test instrument that has been approved for use in a minimum of six colleges from six different California community college districts as a locally managed instrument may be considered as reaching critical mass of evidence for the instrument to be used in other colleges. The evidence from the six colleges must be of the same version of the instrument. After approval as a critical-mass instrument, the instrument can be used by any California community colleges in the same fashion as a second-party test.

Critical mass can be achieved in one of two ways:

1. A minimum of six colleges from at least six different California community college districts have independently attained some level of approval for the same version of a test.

    - Colleges following this process independently submit as colleges locally developing or managing a test, and the criteria for locally developed/managed tests (i.e., Section FOUR, Subsection I) would be followed.

    - The colleges may have gained approval concurrently or over time. The critical factor is that at least six colleges have approval at the same time.

    - Disproportionate impact studies are not required for a critical mass approval, although disproportionate impact must be addressed in a local college application.

    - There must be at least one validity study using empirical data (e.g., consequential validity or criterion validity) for critical mass approval.

2. A consortium of a minimum of six colleges from at least six different California community college districts collaborate on their submission efforts, and, as a group, provide the required data as one submission. A consortium application must address test fairness, validity, reliability, and accommodations for special groups.

- A consortium of colleges that wishes to follow the critical mass process needs to apply in writing to the Chancellor's Office providing notification of its intent and justification for that approach.

- The concept of critical mass allows the consortium to generalize fairness and reliability of an assessment instrument without having to replicate all of the data collection normally required for a locally developed or managed test. Group studies or aggregated data could be submitted for fairness/test bias and reliability.

- In submitting evidence as a consortium, validity studies must be conducted by each college. Relevant and satisfactory validity evidence--including content validity, criterion validity, or consequential validity--will be accepted. However, there must be a validation study using empirical data other than content validity.

- If the instrument is revised for testing individuals who cannot take the test under standard conditions, there must be documentation of all changes along with the basis for each accommodation.

In either case, for approval as a critical-mass test, the Assessment Advisory Committee and the Chancellor's Office must determine that the six (or more) colleges are indeed from six (or more) different California community college districts **and** that they collectively form a representative sample of the California community college student population including gender, age, race/ethnicity, region, etc.

If the test garnered approval as a critical-mass test based on independent submissions (i.e., not through a consortium of colleges), the critical mass status is retained only for a period of six years from when the first college gained approval, unless that college is approved again in a timely manner under the renewal approval process or a consortium of colleges (as defined above) has been approved in a timely manner through the renewal approval process.

If approval is granted, the test will be placed on the Chancellor's Office's approved list as a critical-mass test. Any California community college can use an approved critical-mass test on its campus. When using a critical-mass test, the local college assumes the same responsibilities as those responsibilities it has when using a second-party test. In other words, the criteria described in Section FOUR, Subsection III are applicable to the college using a critical-mass test.

SECTION SIX: RENEWAL OF AN INSTRUMENT'S APPROVAL STATUS

The length of time a test can be available for use by the colleges varies by specific approval category. (See pp. 21-22 for the tenure interval for each approval category.) Regardless, a test can only be considered approved for a maximum of six years starting when status in any of the three approval categories is attained. After this six-year maximum tenure interval, unless new supporting documentation has been submitted and favorably reviewed for its continued use in the California community colleges, the instrument will be downgraded automatically to the Not Approved status.

Sufficient evidence addressing relevant standards must be submitted in advance of the six-year expiration date to allow for a timely renewal of the instrument to be retained on the Chancellor's Office List of Approved Assessment Instruments. Second-party publishers must resubmit information and documentation during the fifth year of approval so that continued use can be maintained by colleges. Similarly, colleges with locally developed/managed tests should resubmit information and documentation during the fifth year of approval.

The approval renewal process is viewed as a time when tests, evidence, and procedures are to be reexamined relative to their appropriateness and continued use for placement in the California community colleges. This renewal requirement is derived from the premise that collecting and evaluating fairness, validity, and reliability evidence should be an ongoing and continuous process. It should be noted that when changes occurred to the test instrument (e.g., changes in items, scoring method, and/or norms) or to the proposed instrument usage (e.g., different curriculum or course sequence), the test shall be reviewed as a new instrument rather than renewal. Other changes triggering a new test review include:

- For a computer-adaptive test, 20 percent or more items in the item bank have been changed (additions, removals, or revisions).
- For a locally developed/managed test, the student population at the college has changed significantly since the test was last approved.

The extent to which the standards (fairness, validity, and reliability) are to be addressed was detailed in Section FOUR of this document. Table 3 and Table 4 in Appendix D summarize the specific requirements for each standard that colleges locally developing or managing tests and second-party publishers need to meet for a renewal approval. Studies and data used to support instrument renewal must be conducted/compiled within the three-year period prior to the renewal application.

An assessment instrument under renewal review may be placed in any one of the three approval categories or may be placed in the Not Approved category. If a renewed test initially receives Provisional Approval or Probationary Approval, the timelines for attaining Full Approval status are the same as for first-time approval requests described on pages 21-22.

SECTION SEVEN: MULTIPLE MEASURES

California Code of Regulations, title 5, section 55522 requires that when colleges use an assessment for course placement, "it must be used with one or more other measures to comprise multiple measures." Title 5, section 55502(i) further defines the "multiple measures" as "a required component of a district's assessment system and refers to the use of more than one assessment measure in order to assess the student. Other measures that may comprise multiple measures include, but are not limited to, interviews, holistic scoring processes, attitude surveys, vocational or career aptitude and interest inventories, high school or college transcripts, specialized certificates or licenses, education and employment histories, and military training and experience." Other than education, demographic factors or personal student data (such as racial or ethnic origin, political or religious affiliation, trade-union membership, or health status) are not valid multiple measures.

The measures chosen must be evidence-based. Multiple measures should be of different formats to allow students multiple opportunities to illustrate their knowledge and readiness. Therefore, using two or more highly correlated tests does not satisfy the requirement for the use of multiple measures. The exception is the use of EAP, SAT, and ACT scores, which the Chancellor's Office has approved for use as multiple measures, or, in the case of the EAP, as a waiver to place students directly in transfer-level coursework (see the July 2015 memo from the Chancellor's Office). The goal of multiple measures is to achieve a more comprehensive student assessment than relying on a single measure or a single test.

Through continued validation, monitoring and refinement of the multiple measures system, the expectation is that the college will identify an appropriate combination of methods to most accurately assess students' capacity to succeed in courses into which they are placed. An evaluation may include gathering data on consequential-related evidence, course success rates, and the likelihood of a student completing the course sequence from basic skills through transfer-level. See the RP Group's publication, "Validating Placement Systems Comprising Test and Multiple Measure Information."

Effects on students' placement accuracy, student success, and its potential disproportionate impact on course placement recommendations must be investigated and reviewed at least every three years. Collecting and evaluating validity evidence for multiple measures must be an ongoing and continuous process. **Colleges must maintain a portfolio for the multiple measures system that includes an up-to-date description of the system, the evidence of fairness and effectiveness of the system, and the history logs of the evolutions of the system.**

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.) (2014) Standards for educational and psychological testing. Washington, DC: AERA.

Carlson, J. F., Geisinger, K. F., & Jonson, J. L. (Eds.). (2014). The nineteenth mental measurements yearbook. Lincoln, NE: Buros Center for Testing.

Cizek, G. J., & Bunch, M. B. (2007). Standard setting: A guide to establishing and evaluating performance standards on tests. Thousand Oaks, CA: Sage.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. (August 25, 1978). Uniform guidelines on employee selection procedures.

Harris, B. W. (2013). Ensuring equitable access and success: A guide to assessing and mitigating disproportionate impact in student success and support programs. CA: California Community Colleges Chancellor's Office

Hayward, C. (2017). Validating placement systems comprising test and multiple measure information. CA: RP Group.

Joint Committee on Testing Practices (2004). Code of Fair Testing Practices in Education. Washington, DC: Joint Committee on Testing Practices.

Zieky, M. J., Perie, M., & Livingston, S. (2008). Cutscores: A manual for setting standards of performance on educational and occupational tests.

APPENDICES

Appendix A: Relevant Education Code Sections

California Education Code
TITLE 3. POSTSECONDARY EDUCATION

Section 78213.
(a) No district or college may use any assessment instrument for the purposes of this article without the authorization of the board of governors. The board of governors may adopt a list of authorized assessment instruments pursuant to the policies and procedures developed pursuant to this section and the intent of this article. The board of governors may waive this requirement as to any assessment instrument pending evaluation.
(b) The board of governors shall review all assessment instruments to ensure that they meet all of the following requirements:
(1) Assessment instruments shall be sensitive to cultural and language differences between students, and shall be adapted as necessary to accommodate students with disabilities.
(2) Assessment instruments shall be used as an advisory tool to assist students in the selection of appropriate courses.
(3) Assessment instruments shall not be used to exclude students from admission to community colleges.
(c) The board of governors shall establish an advisory committee to review and make recommendations concerning all assessment instruments used by districts and colleges pursuant to this article.
(d) For purposes of this section, “assessment” means the process of gathering information about a student regarding the student’s study skills, English language proficiency, computational skills, aptitudes, goals, learning skills, career aspirations, academic performance, and need for special services. Assessment methods may include, but not necessarily be limited to, interviews, standardized tests, attitude surveys, vocational or career aptitude and interest inventories, high school or postsecondary transcripts, specialized certificates or licenses, educational histories, and other measures of performance.

78214.
(a) All participating districts shall, with the assistance of the chancellor, establish and maintain institutional research to evaluate the effectiveness of the Student Success and Support Program described by this article and of any other programs or services designed to facilitate students' completion of their educational goals and courses of study.
(b) The metrics for this research shall include, but not be limited to:
(1) Prior educational experience, including transcripts when appropriate, as determined by the chancellor.
(2) Educational goals and courses of study.
(3) Criteria for exemption from orientation, assessment, or required counseling or advisement, if applicable.
(4) Need for financial assistance.
(5) Disaggregated data by ethnicity, gender, disability, age, and socioeconomic status, to the extent this information is available.

(6) Academic performance, such as the completion of specified unit thresholds, success in basic skills courses, grade point average, course completion outcomes, transfer readiness, and degree and certificate completion.
(7) Any additional information that the chancellor finds appropriate.
(c) The evaluation provided for by this section shall include an assessment of the effectiveness of the programs and services in attaining at least the following objectives:
(1) Helping students to define their academic and career goals and declare a course of study.
(2) Assisting institutions in the assessment of students' educational needs and valid course placement.
3) Helping support students' successful course completion and goal attainment.
(4) Matching institutional resources with students' educational needs.

California Code of Regulations
Title 5 [sections relevant to assessment]

Section 55502. Definitions
For purposes of this subchapter, the following definitions shall apply:
(a) "Assessment for placement" hereinafter referred to as "assessment" is the process of gathering information about individual students in order to identify their skill level and appropriately direct them to courses for which they are prepared. Information used in the assessment process may include, but is not limited to, information regarding the student's study skills, English language proficiency, computational skills, aptitudes, goals, learning skills, career aspirations, academic performance, and need for special services. Assessment involves the collection of such information for purposes of course placement.
(b) "Assessment test" is a validated, standardized, or locally-developed test used in addition to other measures in the course placement process.
(e) "Disproportionate impact" in broad terms is a condition where access to key resources and supports or academic success may be hampered by inequitable practices, policies, and approaches to student support or instructional practices affecting a specific group. For the purpose of assessment, disproportionate impact is when the percentage of persons from a particular racial, ethnic, gender, age, or disability group, who are directed to a particular service or course placement based on an assessment test or other measure is significantly different from the representation of that group in the population of persons being assessed, and that discrepancy is not justified by empirical evidence demonstrating that the assessment test or other measure is a valid and reliable predictor of performance in the relevant educational setting.
(i) "Multiple measures" are a required component of a district's assessment system and refer to the use of more than one assessment measure in order to assess the student. Other measures that may comprise multiple measures include, but are not limited to, interviews, holistic scoring processes, attitude surveys, vocational or career aptitude and interest inventories, high school or college transcripts, specialized certificates or licenses, education and employment histories, and military training and experience.

Section 55520. Required Services
At a minimum, each community college district shall provide students, except as exempted pursuant to section 55532, with all of the following Student Success and Support Program services:
(a) orientation on a timely basis, pursuant to section 55521.
(b) assessment for all nonexempt students pursuant to section 55522;

Section 55522. English and Mathematics Placement and Assessment
(a) The Chancellor shall establish and update, at least annually, a list of approved assessment tests for use in placing students in English, mathematics, or English as a Second Language (ESL) courses and guidelines for their use by community college districts. When using an English, mathematics, or ESL assessment test for placement, it must be used with one or more other measures to comprise multiple measures.

(1) Districts and colleges are required to use the Chancellor's guidelines for the validation of all assessment tests used for placement to ensure that they minimize or eliminate cultural or linguistic bias and are being used in a valid manner. Based on this evaluation, the district or college shall determine whether any assessment test, method, or procedure has a disproportionate impact on particular groups of students, as defined by the Chancellor. When there is a disproportionate impact on any such group of students, the district or college shall, in consultation with the Chancellor, develop and implement a plan setting forth the steps the district will take to correct the disproportionate impact.
(2) The Chancellor may identify other measures of a student's college readiness that community college districts may use for student placement into the college's curriculum.
(b) Each community college district shall adopt procedures that are clearly communicated to students, regarding the college's sample test preparation, how the student test results will be used to inform placement decisions, and the district's limits on the student's ability to re-test.
(c) Community college districts shall not, except as provided in subdivision (d), do any of the following:
(1) use an assessment test for placement which has not been approved by the Chancellor pursuant to section 55522, except that the Chancellor may permit limited field-testing, under specified conditions, of new or alternative assessment tests;
(2) use any assessment test in a manner or for a purpose other than that for which it was developed or has been otherwise validated;
(3) use any assessment test process to exclude any person from admission to a college, except that a college may determine the admission of special part-time or full-time students under Education Code section 76002 based on an assessment which involves multiple measures and complies with other requirements of this subchapter; or
(4) use any assessment test, method, or procedure to exclude students from any particular course or educational program, except that districts may establish appropriate prerequisites pursuant to sections 55002 and 55003.
(5) use any Student Success and Support Program practice which has the purpose or effect of subjecting any person to unlawful discrimination prohibited by subchapter 5 (commencing with section 59300) of chapter 10.
(d) Notwithstanding the provisions of subdivision (c)(1) and (2), assessment tests approved by the Secretary of the United States Department of Education may be used to determine “ability to benefit” in the process of establishing a student's eligibility for federal financial aid pursuant to title 20 United States Code section 1091(d).
(e) Notwithstanding paragraphs (1), (2), (3) or (5) of subdivision (c) or the provisions of sections 55003 or 55522, a community college district may use an assessment test to select students for its nursing program, provided that:
(1) the district complies with all other provisions of this subchapter;
(2) the assessment test or other measures are used in conjunction with other assessment test, methods, or procedures to select students for enrollment in the nursing program; and
(3) the Chancellor has determined that the assessment test predicts likelihood of success in nursing programs, has approved use of the assessment test for that purpose and has established statewide proficiency cut-off scores for that test pursuant to Education Code section 78261.

Note: Authority cited: Section 11138, Government Code; and Sections 66700 and 70901, Education Code. Reference: Section 11135, Government Code; and Sections 72011, 76002, 78211, 78212, 78213 and 78261, Education Code.

Section 55522.5. English as a Second Language Placement and Assessment
(1) The Chancellor shall establish and update, at least annually, a list of the approved assessment tests and instruments for use in placing students in credit ESL courses and guidelines for their use by community college districts. When using an ESL assessment test for placement into credit ESL coursework, it must be used with one or more other measures to comprise multiple measures.
(2) Districts and colleges are required to use the Chancellor's guidelines for the validation of all assessment tests used for placement to ensure that they minimize or eliminate cultural or linguistic bias and are being used in a valid manner. Based on this evaluation, the district or college shall determine whether any assessment test, method, or procedure has a disproportionate impact on particular groups of students, as defined by the Chancellor. When there is a disproportionate impact on any such group of students, the district or college shall, in consultation with the Chancellor, develop and implement a plan setting forth the steps the district will take to correct the disproportionate impact.
(3) The Chancellor may identify other measures of a student's college readiness that community college districts may use for student placement into the college's curriculum.
(d) Each community college district utilizing approved assessment tests or instruments shall adopt procedures that are clearly communicated to students regarding the college's sample test preparation, how the student test results will be used to inform placement decisions, and the district's limits on the student's ability to re-test.
(e) Community college districts shall not, except as provided in subdivision (g), do any of the following:
(1) Use an assessment test for placement which has not been approved by the Chancellor pursuant to this section, except that the Chancellor may permit limited field-testing, under specified conditions, of new or alternative assessment tests;
(2) Use any assessment test in a manner or for a purpose other than that for which it was developed or has been otherwise validated;
(3) Use any assessment test process to exclude any person from admission to a college, except that a college may determine the admission of special part-time or full-time students under Education Code section 76002 based on an assessment which involves multiple measures and complies with other requirements of this subchapter;
(4) Use any assessment test, method, or procedure to exclude students from any particular course or educational program, except that districts may establish appropriate prerequisites pursuant to sections 55002 and 55003; or
(5) Use any Student Success and Support Program practice which has the purpose or effect of subjecting any person to unlawful discrimination prohibited by subchapter 5 (commencing with section 59300) of chapter 10.

Note: Authority cited: Sections 66700 and 70901, Education Code. Reference: Sections 72011, 76002, 78211, 78212, 78213 and 78261, Education Code; and Section 11135, Government Code.

Section 55526. Accommodations

(a) Student Success and Support Program services for students with disabilities shall be appropriate to their needs, and colleges shall, where necessary, make modifications to the services provided or use alternative tests, methods, or procedures to accommodate the needs of such students. Colleges may require students requesting such accommodations to provide proof of need. Disabled Students Programs and Services (DSPS) is authorized, consistent with the provisions of subchapter 1 (commencing with section 56000), to provide specialized services and modified or alternative services as identified in 55520. Notwithstanding this authorization, participation in the DSPS program is voluntary and no student may be denied necessary accommodations in the assessment process because he or she chooses not to use specialized services provided by these programs.

(b) Student Success and Support Program services for students served by the Extended Opportunity Programs and Services (EOPS) who are disadvantaged by economic, social, and educational status shall be appropriate to their needs, and colleges shall, where necessary, make modification to the services provided or use alternative supports to meet the needs of such students. EOPS is authorized, consistent with the provisions of subchapter 2.5 (commencing with section 56200) of chapter 7 to provide services that are over, above, and in addition to services otherwise provided to all credit-enrolled students. Notwithstanding this authorization, participation in the EOPS program is voluntary and no student may be denied necessary supports because he or she chooses to not use specialized services provided by this program.

(c) Colleges shall ensure that Student Success and Support Program services are accessible for English language learners and are appropriate to their needs. Colleges shall, where necessary, make modifications to the services provided to accommodate the needs of such students. Modified or alternative services for limited or non-English-speaking students may be provided in English as a Second Language programs.

Note: Authority Note: Authority Section 11138, Government Code; Sections 66700, 70901 and 78213, Education Code. Reference: Section 11135, Government Code; and Sections 72011, 78211 and 78213, Education Code.

Section 55530. Student Rights and Responsibilities

(a) All students shall be required to:

(1) identify an education and career goal;

(2) diligently engage in course activities and complete assigned coursework; and

(3) complete courses and maintain progress toward an education goal and completing a course of study.

(b) Nonexempt first time students shall, within a reasonable period of time, be required to:

(1) identify a course of study.

(2) be assessed to determine appropriate course placement.

(3) complete an orientation activity provided by the college.

(4) participate in counseling, advising, or another education planning service pursuant to section 55523 to develop, at a minimum, an abbreviated student education plan.
(c) For the purposes of this section, a first time student is a student who enrolls at the college for the first time, excluding students who transferred from another institution of higher education. For purposes of this section, first time enrollment does not include concurrent enrollment during high school. To the extent that a college has the capacity to require and provide the services identified in (b)(1) through (4) to other students, nothing in this section would preclude a college from doing so.
(d) Nonexempt students who have completed the services identified in (b)(1) through (4) shall be required to complete a comprehensive education plan after completing 15 semester units or 22 quarter units of degree applicable credit course work or prior to the end of the 3rd semester or 4th quarter of enrollment, or a shorter period if required by district or program policy.
(e) Failure to fulfill the required services listed in (b) may result in a hold on a student's registration or loss of registration priority pursuant to section 58108 until the services have been completed.
(f) Information obtained from the matriculation process shall be considered student records and shall be subject to the requirements of subchapter 6 (commencing with section 54600) of chapter 5.

Note: Authority cited: Sections 66700 and 70901, Education Code. Reference: Sections 76000, 76001 and 78212, Education Code.

Section 55531. Institutional Responsibilities
(a) The governing board of each community college district shall adopt policies reflecting the provisions of section 55530, Student Rights and Responsibilities. Colleges shall take steps to ensure that information regarding its matriculation policies are accessible and available to all students during or prior to enrollment (e.g., during orientation) and are included in class schedules, catalogs, or other appropriate communications describing student rights and responsibilities under this subchapter.
(b) Once the student has identified a course of study and completed 15 semester units or 22 quarter units of degree applicable course work, the college must provide the student with an opportunity to develop a comprehensive student education plan pursuant to section 55524 within a reasonable time period. Student responsibilities shall also be identified in the student's education plan developed pursuant to section 55524.
(c) Colleges are required to provide nonexempt students with the services specified in sections 55520, 55521, 55522, 55523, and 55524. Initial implementation of these services is required for first time students identified in section 55530(b) by the fall 2015 term. Beginning with the spring 2015 term, districts shall notify students of the requirements established by this subchapter.
(d) Districts may establish a policy providing that a nonexempt student will have a hold placed on registration or lose registration priority pursuant to section 58108 if a student fails to fulfill the responsibilities set forth in section 55530(b) and (c).
(e) Districts and colleges shall make reasonable efforts to avoid duplication of the orientation, assessment, counseling, advising, or other education planning services, and development of student education plans funded through this subchapter or funded through other programs.
(f) It is the intent of this subchapter that instructional and student services departments at each college shall use multiple sources of data from student education planning efforts and identified courses of study to coordinate course scheduling.

Note: Authority cited: Sections 66700 and 70901, Education Code. Reference: Sections 76000, 76001 and 78212, Education Code.

§ 55532. Exemptions

(a) Community college districts may elect to exempt certain students from participation in orientation, assessment, counseling, advising, or student education plan development, as required by subdivisions (a), (b), (c), or (d) of section 55520. Each district shall establish policies specifying the grounds for exemption. Such policies shall be identified in the Student Success and Support Program plan required under section 55510 and the number of students so exempted shall be reported, by category, to the Chancellor pursuant to section 55511.

(b) Districts may adopt policies that exempt a student from orientation, assessment, counseling, advising, or student education plan development if the student:

(1) has completed an associate degree or higher;

(2) has enrolled at the college for a reason other than career development or advancement, transfer, attainment of a degree or certificate of achievement, or completion of a basic skills or English as a Second Language course sequence;

(3) has completed these services at another community college within a time period identified by the district;

(4) has enrolled at the college solely to take a course that is legally mandated for employment as defined in section 55000 or necessary in response to a significant change in industry or licensure standards.

(5) has enrolled at the college as a special admit student pursuant to Education Code section 76001.

(c) Any student exempt from orientation, assessment, counseling, advising, or student education plan development shall be notified and may be given the opportunity to participate in those services.

(d) District policies shall not exempt a student solely because a student has not selected an education and career goal or course of study.

Note: Authority cited: Sections 66700, 70901, and 78215, Education Code.Reference: Section 78215, Education Code.

Appendix B: Request for Approval or Renewal of a Locally Developed/Managed Test

California Community Colleges Chancellor's Office
Preparing for Approval of a Locally Constructed/Managed Test

As indicated on the Request for Approval Form below, a brief narrative must be submitted summarizing the evidence supporting the use of the test. Based on reviews of previous material submitted, the following information may help you create these summary statements:

1. This narrative should not be a lengthy report. However, enough details should be provided for the Chancellor's Office and the Assessment Advisory Committee to evaluate the quality of the supporting studies provided. Summaries for all five standards (validity, reliability, fairness/test bias, cut or placement scores, and disproportionate impact) should not exceed 20 pages for any one test.

2. According to the Standards for Evaluation of Assessment Instruments, a test will not be approved unless evidence provided supports at least one aspect of validity and fairness/test bias. Therefore, to obtain at least the minimum level of approval, a college must provide sufficient evidence in these two areas to support the use of the test.

3. It is the college's responsibility to provide an integrated argument for claims concerning a standard rather than just presenting facts and letting the reviewer draw his own conclusions.

4. For content validity evidence, the college should describe how a test was selected to match the prerequisite skills of the course(s) for which it places students. A well-documented content validity study linking test items or performance tasks and scoring rubrics to course prerequisite skills based on instructor ratings provides powerful validity data supporting use of the test.

5. For the other types of validity evidence required, colleges have the option to submit either criterion-related or consequential-related validity evidence. Criterion-related validity does not need to be restricted to the correlation coefficients (e.g., studies seeking a .35 correlation) between the proposed placement test and end-of-course grade. Other criteria and types of analyses may be used as the primary evidence when arguing for the validity of the instrument.

6. Logical fairness review procedures should be conducted at the individual item level rather than at an overall test level. A diverse panel reflecting the college's student population (with the emphasis on including panelists from the impacted groups) should be used. The summary of this component should include a description of the number and type of each impacted group included in the review.

7. A description of the process to address any unfair items identified should be included.

8. Data collected for item fairness and disproportionate impact studies should be provided for each impacted student group. For ESL placement tests, the groups would be based on linguistic differences.

9. Data submitted as cut-score validity or disproportionate impact evidence may be different depending on whether the application is an initial request or a renewal. For example, consequential validity study or success ratio data are not needed to validate cut-scores for initial requests, although colleges must submit documentation indicating appropriate procedures were used to determine cut scores. Similarly, only a description of the plan to monitor disproportionate impact is required for initial requests. For renewal applications, evidence that the cut-scores have been re-examined or monitored is required as is direct evidence (i.e., data) on disproportionate impact.

10. Evidence must be presented for generalization of results across forms, courses, and colleges.

    - If two or more forms or prompts of a test are in use, evidence that the forms or prompts are parallel and equated must be provided or evidence in each of the areas indicated in the standards must be presented.

    - If multiple courses are involved, evidence must be presented for each course.

    - If the request is for approval at two or more colleges, evidence must support that the colleges are parallel (the same) in course content, delivery of instruction and student populations served. Otherwise, each college must submit evidence specific to their campus.

Appendix B

California Community Colleges Chancellor's Office
Request for Approval or Renewal of a Locally Developed/Managed Test

The minimum requirements for approval are to provide at least one type of validity evidence and address the fairness standard. Please note that it is not necessary to submit extensive documentation to support your request. Please summarize any data from technical reports or other sources that indicate whether a specific standard has been met at a minimal level for your instrument.

When requested, indicate which areas have been investigated or addressed and those not yet addressed. Studies addressing all of areas indicated in the standards need not be completed in order to request approval of an instrument.

Note: The college must receive authorization from the publisher for use of any locally managed, second-party test.

College:

Address:

Contact Person:

Title:

Telephone Number:

Email:

1. Identify the test with its complete title and its MIS code (if renewal):

2. Which course(s) is this test used to assist with student placement?

3. Have there been investigations of the validity of the use of scores obtained from this test? If you cannot answer yes to one of the options below, do not submit this request until some validity evidence is available.

- ☐ YES, all required studies have been completed. A brief narrative is attached that summarizes the information from all such investigations.
- ☐ YES, but not all required studies have been completed. A brief narrative is attached that summarizes the procedures and findings from all such investigations.

  Projected completion date for required studies not completed: _____

4. Have there been investigations of the reliability of scores obtained from this test?

- ☐ YES. A brief narrative is attached summarizing the procedures and findings from all such investigations.
- ☐ NO. Projected completion date: _____

5. Have there been investigations of test bias? (If your response is no to this question, do not submit this request until some test bias evidence is available. Note also that the required evidence may be different depending on whether this is an initial or renewal request for an instrument.

- ☐ YES. A brief narrative is attached summarizing the procedures and findings from all such investigations.

6. Have there been investigations of the adequacy of the **cut or placement score(s)** used with this test?

- ☐ YES. A brief narrative is attached summarizing the procedures and findings from all such investigations.
- ☐ NO. Projected completion date: _____

7. Have there been investigations planned (for first-time submissions) or conducted (for renewal) of **disproportionate impact** in those courses that rely on this test to assist in placement decisions?

- ☐ YES. A brief narrative is attached summarizing the procedures and findings from all such investigations.
- ☐ NO. Projected completion date: _____

Documented evidence maintained at the appropriate college or district office supports the suitability of this assessment instrument to provide fair and equitable student placement information, as described in the California Community College validation standards. At a minimum, evidence from at least one validity study (content, criterion-related or consequential) and a bias study supports the continued use of the instrument.

College Assessment Officer	Signature	Date
College Research Officer	Signature	Date
College Subject Discipline Faculty/Chair	Signature	Date
College Superintendent/President	Signature	Date

Appendix C: Flowchart for Determining Application Category and Tables of Criteria

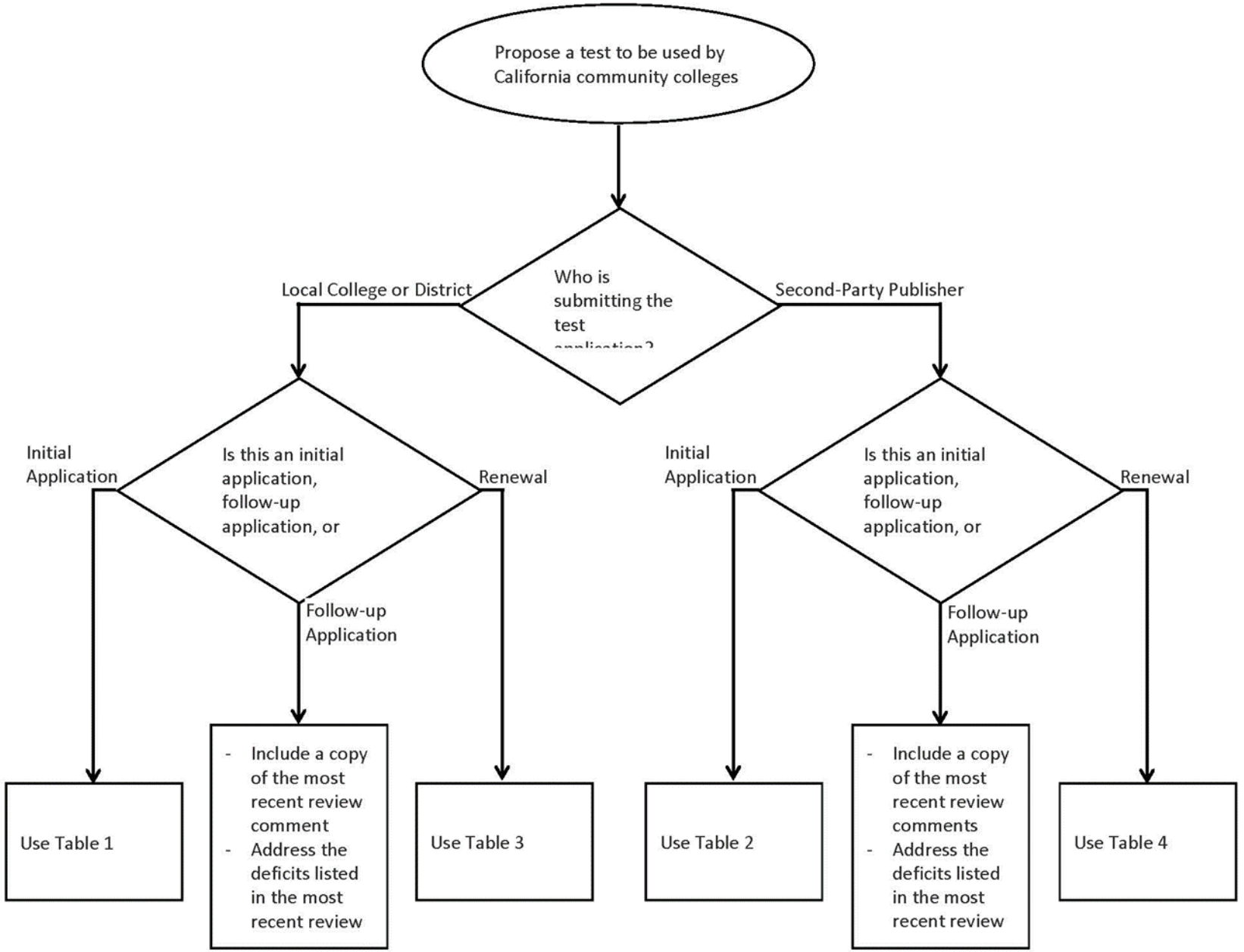

Propose a test to be used by California community colleges
Who is submitting the test application?
Local College or District
Second-Party Publisher
Is this an initial application, follow-up application, or
Initial Application
Renewal
Follow-up Application
Use Table 1
- Include a copy of the most recent review comment
- Address the deficits listed in the most recent review
Use Table 3
Is this an initial application, follow-up application, or
Initial Application
Renewal
Follow-up Application
Use Table 2
- Include a copy of the most recent review comments
- Address the deficits listed in the most recent review
Use Table 4

Table 1. Initial Application for Locally Developed/Managed Tests	
	Description and Requirements
Fairness/Test Bias	Logical review conducted at the local college or Citing logical review results conducted by other community college(s) with similar demographics or Citing logical review results conducted by the test publisher if appropriate See Section FOUR, Subsection I.1.a.
Content Validity	Item-by-item analysis comparing each test item to each course prerequisite See Section FOUR, Subsection I.2.a.
Consequential or Criterion Validity	Only need if providing local cut-score validation
Cut-Score Validation	Judgmental approach or Empirical approach (e.g., consequential/criterion validity study) See Section FOUR, Subsection I.2.c.
Reliability	At least one reliability estimate and the standard error of measurement associated with it and Direct performance assessments: Inter-scorer reliability See Section FOUR, Subsection I.3.
Disproportionate Impact	A plan describing how data on disproportionate impact will be monitored and evaluated locally See Section FOUR, Subsection I.1.b.
ADA Accommodations	Identify accommodations that are provided and Altered test forms need documentation and justification for all changes See Section FOUR, Subsection I.1.c.

Note. Additional information may be needed for computer-based testing. See Section FOUR, Subsection IV. for details.

Table 2. Initial Application for Second-Party Tests	
	Description and Requirements
Fairness/Test Bias	Logical review of bias, insensitivity, and offensiveness by a panel and Empirical review and justifications for item retention, revision, or removal See Section FOUR, Subsection II.1.a.
Content Validity	Test objective and table of specifications and Test booklets (or a representative sample of operational test items) and Contact information for requesting operational booklets for local colleges to conduct item-by-item content review See Section FOUR, Subsection II.2.a.
Consequential or Criterion Validity	Criterion or consequential validity study See Section FOUR, Subsection II.2.b.
Cut-Score Validation	Not required
Reliability	Reliability estimates for all relevant sources of measurement errors, including coefficient of stability and Standard errors of measurement and Direct performance assessments: Inter-scorer reliability See Section FOUR, Subsection II.3.
Disproportionate Impact	Not required
ADA Accommodations	Alternative test forms and/or appropriate accommodations for students with disabilities See Section FOUR, Subsection II.1.b.

Note. Additional information may be needed for computer-based testing. See Section FOUR, Subsection IV. for details.

Table 3. Renewal Application for Locally Developed/Managed Tests	
	Description and Requirements
Fairness Tests **/Test Bias**	Logical review conducted at the local college or Citing logical review results conducted by other community college(s) with similar demographics or Citing logical review results conducted by the test publisher if appropriate See Section FOUR, Subsection I.1.a.
Content Validity	Item-by-item analysis comparing each test item to each course prerequisite See Section FOUR, Subsection I.2.a.
Cut-Score Validation	Empirical approach (e.g., consequential validity study) See Section FOUR, Subsection I.2.c.
Reliability	Not required
Disproportionate Impact	Disproportionate impact analysis and A detailed action plan if disproportionate impact is found See Section FOUR, Subsection I.1.b.
ADA Accommodations	Identify accommodations that are provided and Altered test forms need documentation and justification for all changes See Section FOUR, Subsection I.1.c.

Note. Additional information may be needed for computer-based testing. See Section FOUR, Subsection IV. for details.

Note. All data and analyses used to support renewal must be recent (within the last three years).

Table 4. Renewal Application for Second-Party	
	Description and Requirements
Fairness/Test Bias	Logical review of bias, insensitivity, and offensiveness by a panel and Empirical review and justifications for item retention, revision, or removal See Section FOUR, Subsection II.1.a.
Content Validity	Contact information for requesting operational booklets for local colleges to conduct item-by-item content review See Section FOUR, Subsection II.2.a.
Consequential or Criterion Validity	Criterion or consequential validity study See Section FOUR, Subsection II.2.b.
Cut-Score Validation	Not required
Reliability	Not required
Disproportionate Impact	Not required
ADA Accommodations	Alternative test forms and/or appropriate accommodations for students with disabilities See Section FOUR, Subsection II.1.b.

Note. Additional information may be needed for computer-based testing. See Section FOUR, Subsection IV. for details.

Note. All data and analyses used to support renewal must be recent (within the last three years).

Appendix D: Assessment Advisory Committee Charter (September 2021)

STATE OF CALIFORNIA

California Community Colleges
Chancellor's Office

Assessment Advisory Committee
Charter
September 2021

The Assessment Advisory Committee conducts the review of assessment instruments submitted by colleges and test publishers for Chancellor's Office approval. The committee works with the Chancellor's Office and psychometric consultants for the Chancellor's Office, who conduct the psychometric review of assessment instruments and provide other technical expertise as required. The committee then advises the Chancellor's Office on assessments presented for approval and provides recommendations regarding those approvals.

Establishment and Authority

Per Education Code 78213, a community college district or college shall not use any assessment instrument related to Education Code 78213 without the authorization of the board of governors. The board of governors may adopt a list of authorized assessment instruments and shall establish an advisory committee to review and make recommendations concerning all assessment instruments used by districts and colleges related to Education Code 78213. See the full text of Education Code 78213 below.

Membership

Assessment Advisory Committee members are appointed by a representative set of stakeholder groups and associations across the CC system. Each member serves a one-year term and is eligible for a second year, as determined by their appointing group/association. The Assessment Advisory Committee members consists of the following voting members:

- Three representatives from the AB 705 ESL Implementation Work Group.
- Two representatives from the Educational Services and Support Division of the CCCCO.
- Four representatives from the Academic Senate for California Community Colleges (one each for English, math, ESL, and non-credit).
- One representative from the California Association of Community College Registrars and Admissions Officers.
- Two representatives from the California Community Colleges Assessment Association.
- One representative from the California Community Colleges Classified Senate (involved in assessment).
- One representative from the Chief Instruction Officers.

- One representative from the Chief Student Services Officers.
- Two representatives from the Research & Planning (RP) Group (with one preferably having experience with multiple measures).
- One representative from the Research and Data Division of the CCCCO.
- One representative from the Student Equity and Achievement Program (SEAP) Advisory Committee (with background in assessment).
- Two representatives (and one alternate) from the Student Senate for California Community Colleges.
- One representative from the Workforce and Economic Development (WED) Division of the CCCCO.

Resource Members

- One representative from the Office of General Counsel of the CCCCO.

Leadership

The Assessment Advisory Committee is overseen by a Vice Chancellor of Educational Services and Support and is co-chaired with an Educational Services and Support Dean.

Purpose and Responsibility

The Assessment Advisory Committee's responsibilities are as follows:

- Review and evaluate assessment validation submissions as needed and provide recommended levels of approval to the Chancellor's Office based on guidance from the psychometric consultants.
- Review and provide feedback on technical assistance materials on assessment-related topics.
- Provide guidance on assessment issues in the CCC.
- Assist with planning assessment trainings, webinars, and workshops as needed.

Per Education Code 78213, as the committee reviews assessment instruments, reviews should be conducted to ensure the following requirements are fulfilled:

(1) Assessment instruments shall be sensitive to cultural and language differences between students, and shall be adapted as necessary to accommodate students with disabilities.
(2) Assessment instruments shall be used as an advisory tool to assist students in the selection of appropriate courses.
(3) Assessment instruments shall not be used to exclude students from admission to community colleges.

For purposes of Education Code 78213 and therefore the work of this committee, "assessment" means the process of gathering information about a student regarding the student's study skills, English language proficiency, computational skills, aptitudes, goals, learning skills, career aspirations, academic performance, and need for special services. Assessment methods may include, but not necessarily be limited to, interviews, standardized tests, attitude surveys, vocational or career aptitude and interest inventories, high school or

postsecondary transcripts, specialized certificates or licenses, educational histories, and other measures of performance.

As a Bagley-Keene committee, the Assessment Advisory Committee will typically meet in-person 2-4 times per year as needed (with exceptions made for virtual meetings pending emergencies). The Chancellor's Office will cover travel costs for in-person meetings according to state travel policy and rates.

Members are expected to review materials in advance of the meetings, actively engage in discussions during meetings, and to participate in work groups as needed.

Decision Making and Recommendations

To establish quorum for decision-making, 50% plus one of the voting members must be present. Vacancies do not count towards the determination of the quorum. The committee shall make every effort to reach consensus when determining recommendations. If consensus cannot be reached, then recommendations shall be made by vote of the voting membership.

Committee recommendations will be received by the presiding Vice Chancellor of Educational Services and Support and taken to the Chancellor for review. Final recommendations will be presented to the Board of Governors for approval.

GLOSSARY

Consequential-related validity evidence addressing desired or undesired outcomes that follow from the use of test scores to advise placement of students into courses.

Content-related validity evidence addressing the extent to which course pre-requisite knowledge and skills are being measured by the items on a test for all courses into which the test scores are being used to place students

Corrected validity coefficients psychometric procedures that estimate the relationship between two sets of scores if the test scores were measured with perfect reliability (corrected for attenuation) or full variability (corrected for restriction of range).

Correlation coefficient a statistical index that summarizes the magnitude of the relationship between two sets of scores for the same group of individuals. This index takes on values ranging from –1.00 to 1.00 with values around zero (.00) representing no relationship.

Criterion-related validity evidence addressing the extent to which scores on the placement test are related to scores on an appropriate criterion measure of student ability to meet different course requirements into which the students are being placed or an appropriate measure of student success in different courses.

Critical mass the accumulation of evidence across a diverse set of colleges, which can be used to gain approval for the use of a test instrument by all colleges in the system.

Differential prediction evidence addressing the extent to which scores on a placement test are equally predictive of an outcome measure for all subgroup classifications, e. g., gender, ethnicity, age, etc.

Direct performance assessments that require an open-ended response from the test taker to a task, set of tasks or set of defined stimulus conditions. Responses then are scored using a standardized scoring rubric that has defined scale values indicating the adequacy of performance at different levels of proficiency.

Empirical approach to setting cut-scores procedures to identify cut-score values based on differential test taker test performance under certain design conditions.

Evidence Based refers to any practice or strategy informed by objective evidence, e.g., research that conforms to explicit criteria.

Internal consistency a method of estimating test score reliability based on the consistency or relationship of responses to test items across test takers for a single administration of the test.

Examples of methods or indices include Kuder-Richardson formula 20 or 21, coefficient alpha and split-half procedures.

Interscorer reliability coefficient an index of reliability indicating the consistency of ratings assigned to test taker responses (usually from performance assessment data) by two or more raters.

Judgmental approach to setting cut-scores procedures to identify cut-score values based on expert panel review, evaluation and judgments about the appropriateness and difficulty of test and test item content, and expected performance for identified populations of test takers.

Norms reported score distributional characteristics for samples of test takers that are intended to represent a population of test takers with described characteristics such that the performance of the norm group can offer relative interpretation of a person's test score with reference to the performance of test takers in the norm group.

Reliability evidence addressing the degree of consistency of measurements when the procedures producing test scores are repeated on a population of individuals or groups.

Stability coefficient an estimate of the reliability of test scores using a procedure requiring that data be collected from the same group of individuals on two separate occasions with an intervening period of at least two weeks between administrations.

Standard error of measurement an index related to the reliability of test scores, which provides information addressing the degree of inaccuracy for specific test score values.

Test the Standards for Educational and Psychological Testing (2014) defines a test as "an evaluative device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process."

Transformed scale scores scores that are reported on a scale other than that produced by raw scores, e.g., percentile ranks or scores reported on a scale with a different mean and standard deviation than those of the raw scores.

Validity evidence addressing the extent to which the interpretation of scores from a test is meaningful, appropriate and useful to serve the purpose of placement of students into different courses.