A GUIDE TO EVALUATING COLLEGE- AND CAREER-READY ASSESSMENTS:

Focus on Test Characteristics

Executive Summary

MARCH 2016

ERIKA HALL, PH.D.
Center for Assessment

SUSAN LYONS, PH.D.
Center for Assessment

Center for Assessment

TABLE OF CONTENTS

TABLE OF CONTENTS

INTRODUCTION

The Center for Assessment was funded by the High Quality Assessment Project[1] to develop methodologies and procedures to help guide persons and organizations who want to apply the Council of Chief State School Officers (CCSSO) Criteria for High Quality Assessments[2] in evaluating summative assessments designed to measure college and career readiness standards[3]. This includes assessments that do not specifically define "college-ready" cut-scores or benchmarks, but were developed to address knowledge and skills defined as necessary for entrance into college, careers and technical education[4].

Center for Assessment researchers have grouped the CCSSO criteria into two components—those dealing with test content and those dealing with test characteristics and program implementation. The criteria associated with test content focus primarily on the quality of items, the accessibility of item and test content, and the alignment of test content to the priority content of college and career ready standards. The criteria associated with test characteristics focus on the psychometric and statistical properties of assessment instruments and the quality of test administration, reports and supplemental information provided to aid in the interpretation and use of test results.

This document summarizes procedures for evaluating assessments against the CCSSO Criteria related specifically to test characteristics[5]. Language that is drawn directly from CCSSO's Criteria for Procuring and Evaluating High Quality Assessments document is shaded in a light green box in order to help differentiate it from the text and materials developed by the Center for Assessment.

> Language that is drawn directly from CCSSO's Criteria for Procuring and Evaluating High Quality Assessments document is shaded in a **light green box** in order to help differentiate it from the text and materials developed by the Center for Assessment.

ABOUT THIS DOCUMENT

This Executive Summary is one of three documents associated with The Guide to Evaluating College- and Career-Ready Assessments: Focus on Test Characteristics. The other two documents include the Evaluation Methodology and the Criteria Evaluation Framework (CEF). The Executive Summary is intended to provide a quick overview of each part of the Evaluation Methodology and introduce the format/structure of the CEF. The intended audience for the Executive Summary includes policy makers, potential funders and requesters, and anyone interested in getting a quick snapshot of what the test characteristics methodology entails.

The Evaluation Methodology outlines a comprehensive set of procedures and guidelines to support the evaluation of assessments against the CCSSO criteria related to test characteristics. The Evaluation Methodology is intended to be used primarily by those implementing an evaluation, but it serves as invaluable reference to anyone interested in the understanding the evaluation process and materials in more detail. The Criteria Evaluation Framework provides an operational definition of each test characteristics criterion, making it a necessary companion to the Evaluation Methodology. The CEF is intended to be used by: the evaluation team (to support the review and evaluation of evidence provided); those charged with supplying evidence to support evaluation (to better understand the type and scope of documentation expected); and anyone interested in better understanding how the CCSSO criteria were operationalized to support evaluation (e.g., assessment developers, measurement professionals).

---

[1] The High-Quality Assessment Project (HQAP) supports state-based advocacy, communications and policy work to help ensure successful transitions to new assessments that measure K–12 college- and career-readiness standards. HQAP's work is funded by a coalition of national foundations, including the Bill and Melinda Gates Foundation, the Lumina Foundation, Helmsley Charitable Trust, the Charles and Lynn Schusterman Foundation and the William and Flora Hewlett Foundation.

[2] See the Criteria for Procuring and Evaluating High Quality Assessments at the link: http://www.ccsso.org/Documents/2014/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf

[3] It is important to note that the test characteristics methodology as well as the test content methodology were developed independent from the U.S. Department of Education's Peer Review Guidance released in September of 2015 https://www2.ed.gov/policy/elsec/guid/assessguid15.pdf. However, evaluations generated using those methodologies could provide important evidence for submission to peer review.

[4] Or in earlier grades, to address the pre-requisite skills necessary to meet these expectations by the end of high school.

[5] Separate documentation for evaluating the criteria dealing with test content, "Guide to Evaluating Assessment Using the CCSSO Criteria for High Quality Assessments: Focus on Test Content" is available at the Center for Assessment's website, www.nciea.org.

SUMMARY OF PART 1:
DEVELOPMENT OF THE TEST CHARACTERISTICS METHODOLOGY

The CCSSO Assessment Quality Criteria

The CCSSO Criteria for High Quality Assessments, which serve as the foundation for this methodology, were developed by the Council of Chief State School Officers to support states as they "develop procurements and evaluate options for high-quality state summative assessments aligned to their college- and career readiness standards." The CCSSO criteria are grouped into five broad categories each containing one or more associated criteria.

> A. Meet Overall Assessment Goals and Ensure Technical Quality
> B. Align to Standards – English Language Arts/Literacy
> C. Align to Standards – Mathematics
> D. Yield Valuable Reports on Student Progress and Performance
> E. Adhere to Best Practices in Test Administration
> F. State Specific Criteria

Each category is identified by a letter, and a criterion is identified by a letter and number[6]. The test content evaluation process includes specific elements of criterion A.5 related to accessibility, criterion A.6 related to transparency of test design and expectations, and all of the criteria associated with categories B and C, related to alignment to standards for ELA and mathematics respectively. This guide, addressing the evaluation of test characteristics, includes criteria A.1-A.4, A.7, technical aspects of criterion A.5, and all of the criteria associated with categories D and E covering reporting and test administration. Criterion category F, state specific criteria, does not include specific criteria and is, therefore, not covered in either document. The full list of criteria addressed by each methodology is provided in Table 1.

TABLE 1. ORGANIZATION OF CCSSO CRITERIA INTO TEST CONTENT AND TEST CHARACTERISTICS

TEST CONTENT	TEST CHARACTERISTICS

A. Meet Overall Assessment goals and Ensure Technical Quality
   - A.5 Providing accessibility to all students, including English learners and students with disabilities (partial)
   - A.6. Ensuring transparency of test design expectations.
 	
B. Align to Standards - English Language Arts/Literacy
   - B.1 Assessing student reading and writing achievement in both ELA and literacy
   - B.2 Focusing on complexity of texts
   - B.3 Requiring students to read closely and use evidence from texts
   - B.4 Requiring and range of cognitive demand
   - B.5 Assessing writing
   - B.6 Emphasizing vocabulary and language skills
   - B.7 Assessing research and inquiry
   - B.8 Assessing speaking and listening
   - B.9 Ensuring high-quality items and a variety of item types
 	
C. Align to Standards - Mathematics
   - C.1 Focusing strongly on the content most needed for success in later mathematics
   - C.2 Assessing a balance of concepts, procedures, and applications
   - C.3 Connecting practice to content
   - C.4 Requiring a range of cognitive demand
   - C.5 Ensuring high-quality items and a variety of item types
	A. Meet Overall Assessment goals and Ensure Technical Quality
	   - A.1 Indicating progress toward college and career readiness
	   - A.2 Ensuring that assessments are valid and required for intended purposes
	   - A.3 Ensuring that assessments are reliable
	   - A.4 Ensuring that assessments are designed and implemented to yield valid and consistent test score interpretations within and across years
	   - A.5 Providing accessibility to all students, including English learners and students with disabilities (partial)
	   - A.7 Meeting all requirements for data privacy and ownership
	
	D. Yield Valuable Reports on Student Progress and Performance
	   - D.1 Focusing on student achievement and progress to readiness
	   - D.2 Providing timely data that inform instruction
	
	E. Adhere to Best Practices in Test Administration
	   - E.1 Maintaining necessary standardization and ensuring test security

While the criteria could have been divided in a number of defensible manners, the researchers at the Center for Assessment believe this division is appropriate given key differences in the skills and expertise necessary to evaluate these two groups of criteria (i.e., content vs. technical), and expectations about when the data/information necessary to support evaluation will be available.[7]

Development of a Criterion Evaluation Framework (CEF) to Support the Evaluation of Assessment Quality

The CCSSO criteria provide a strong foundation upon which to build an evaluation of assessment quality; however, additional detail and structure was necessary to develop a comprehensive evaluation methodology.

The Criteria Evaluation Framework (CEF) expands upon the CCSSO Criteria by: 1) specifying the claims underlying each criterion, 2) describing what sufficient evidence should look like, 3) providing comments and examples that inform the evaluation process, 4) highlighting key connections among claims and criteria, and 5) supporting the credibility of the evaluation by aligning each criterion to the joint *Standards for Educational and Psychological Testing (2014)*.

Brief Overview of the Development and Review Process

Throughout the development of the methodology and the Criteria Evaluation Framework, the materials associated with this effort underwent a series of internal and external reviews, including the implementation of a test case. Figure 1 below broadly summarizes the steps in the development process.

FIGURE 1. DEVELOPMENT TIMELINE
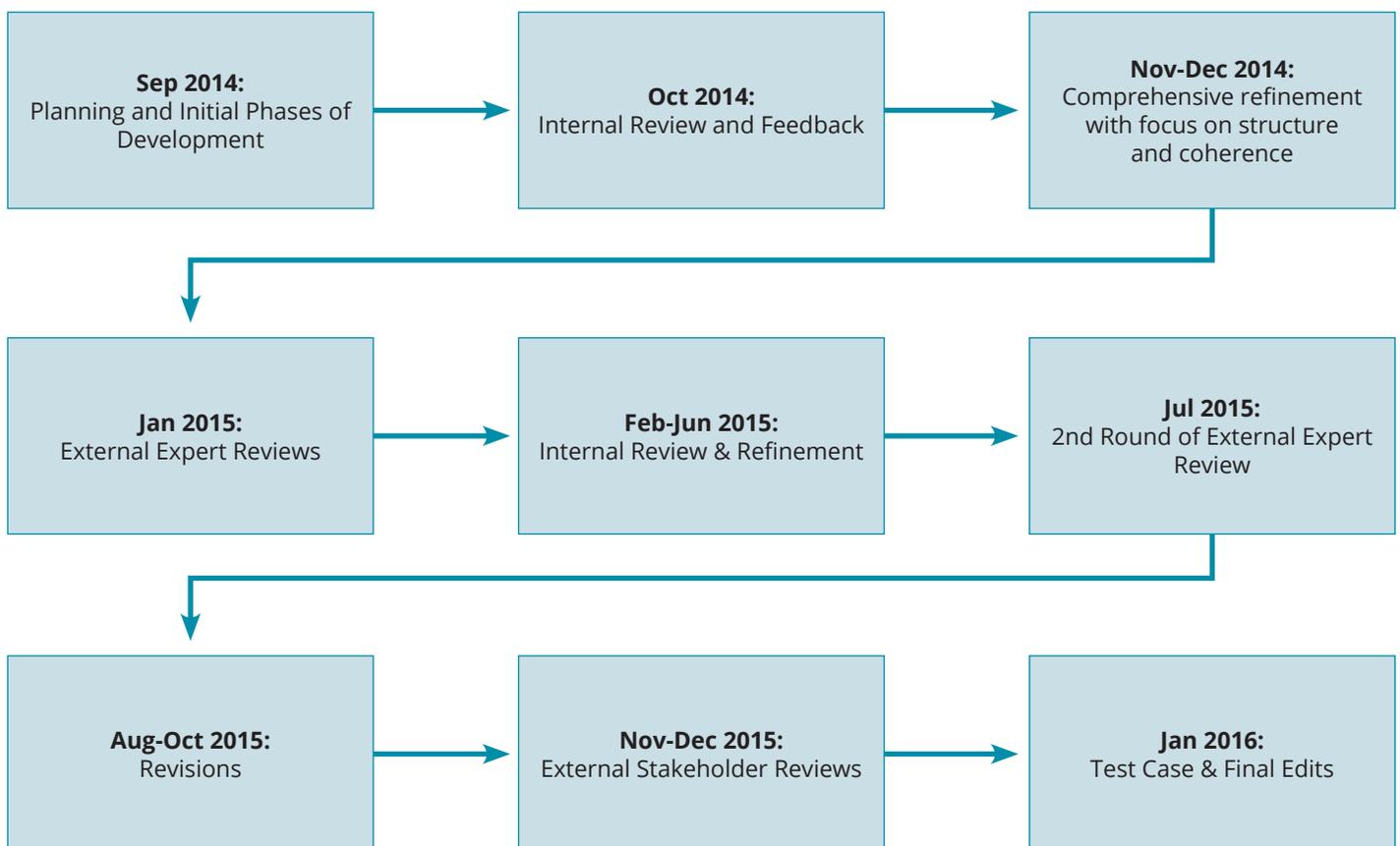
As shown in Figure 1, the development process was iterative in nature allowing for the influence of ideas from a number of different and valuable perspectives. The resulting methodology and associated Criteria Evaluation Framework are much stronger due to the thoughtful and valuable contributions of all those involved.

---

[7] A large portion of the information necessary to support a comprehensive evaluation of the test characteristics criteria will not be available until after an assessment has been administered operationally.

A Framework for Operationalizing the CCSSO Criteria

To guide the evaluation of assessment quality relative to *CCSSO's Assessment Quality Criteria*, one must not only identify the type of evidence necessary to support evaluation, but also decide what specifically about that evidence should be considered in evaluating each criterion. Due to a variety of contextual factors, what constitutes appropriate or high quality evidence is likely to vary across assessment programs. These factors include the manner in which results are to be used, the stakes associated with those uses, and the sponsoring agency[8] or vendor's theory of action as to how the assessment will bring about desired change. For many programs other contextual factors, including the assessment's age or phase of development[9] and degree to which the associated standards have been addressed in the classroom will also be influential.

It is due to these contextual factors that there can never be a "universal" evaluation system that explicitly dictates how evidence should be weighed and evaluated for *all* tests. Instead, the Criteria Evaluation Framework (CEF) provides the structure and content necessary to implement a thoughtful, evidence- and discussion-based evaluation of assessments designed to measure student achievement and progress relative to college- and career-ready standards. While we attempt to operationalize the criteria in a manner that supports a fair, transparent and reliable evaluation, due to the contextual factors previously discussed, expert discussion and judgment plays a significant role in the process and results.

The CEF was organized around the following hierarchical structure.

FIGURE 2. STRUCTURE OF THE CRITERIA EVALUATION FRAMEWORK (CEF)
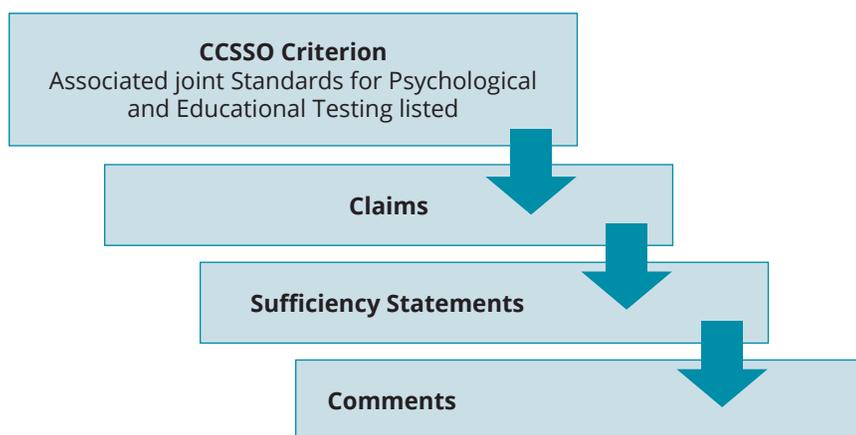
In establishing this structure the first step was to articulate a set of **Claims** supporting each CCSSO criterion. **Claims are statements we want to make about procedures, materials, reports, and/or data given the evidence provided for review.** For each criterion there may be any number of associated claims. As a set, claims suggest not only the type/range of evidence expected but what features of that evidence are important relative to a given criterion. Claims that were developed in direct consideration of a particular criterion are referred to as *primary claims*. Claims that are primary for one criterion, but relevant to the evaluation of another are listed as *secondary claims* for that secondary criterion. For example, there are 11 primary claims associated with A.1 as shown on page 1 of Appendix A. The first five primary claims relate to the process used to develop and evaluate the performance level descriptors and the last six reflect the claims that are associated with setting standards. In addition to these 11 primary claims there are five secondary claims, A.4.6 – A.4.10. These claims are primary claims for A.4, but reference evidence and considerations that will also be relevant to a comprehensive evaluation of A.1. A complete list of the claims associated with each CCSSO criterion is provided in Appendix A.

While the claims define what must be reviewed to evaluate each of the CCSSO criteria, they do not dictate how those materials should be reviewed or the means by which decisions about the quality, appropriateness, and sufficiency of that evidence should be determined. Therefore, for each claim we provide examples of what evidence for high quality assessments should look like and suggestions to inform the evaluation of these examples in different contexts. These elements, represented by the bottom two levels of the framework, are referred to as **sufficiency statements** and **comments**, respectively.

---

[8] The sponsoring agency is the sponsor of the evaluation, and may be the state, a school district, or some other organization.

[9] See page 3 of the *Criteria for Procuring and Evaluating High Quality Assessments* for a discussion about the interaction between an assessment's phase of development and the evidence expected to support evaluation. http://www.ccsso.org/Documents/2014/CCSSO%20Criteria%20for%20High%20Quality%20Assessments%2003242014.pdf

Sufficiency statements describe those features/characteristics we believe should be reflected in a particular type of evidence in order for it to lend useful and adequate support to a given claim. Those involved in conducting the evaluation will be asked to consider these features in addition to the assessment's current phase of development to determine the degree to which each claim and criterion are supported. To inform this process, comments and examples are provided to highlight how contextual factors may influence one's thinking about the quality of evidence submitted for a given test. **Comments are included as additional notes to aid reviewers in judging the quality of evidence within the context of an assessment program.**
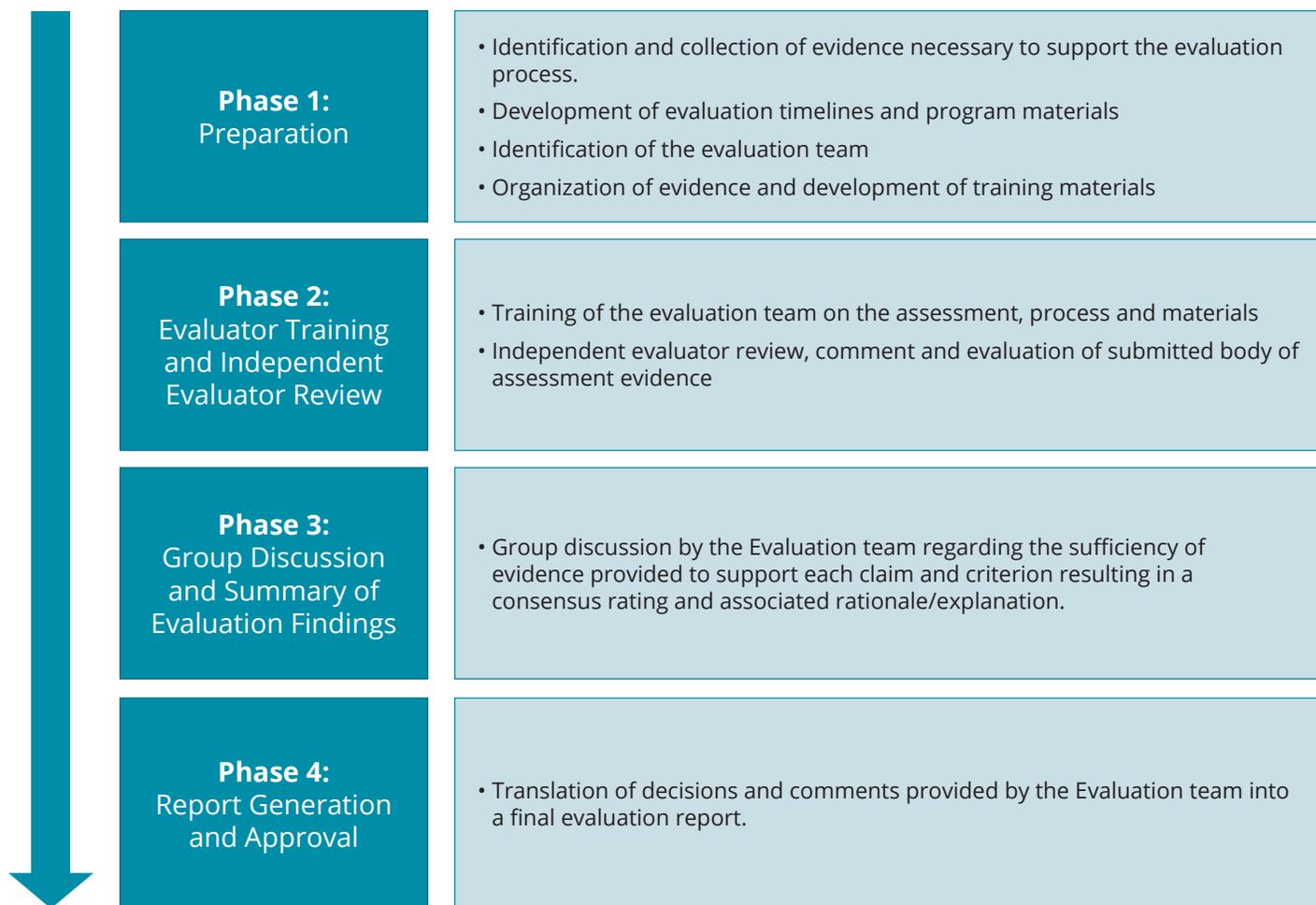
SUMMARY OF PART 2:
GUIDE FOR IMPLEMENTATION

Throughout the discussion which follows, we consistently refer to four different entities, each of which has a clearly defined role in the overall evaluation process. For clarity, each group and its expected role is outlined in Table 2.

TABLE 2. GROUPS INVOLVED IN THE TEST CHARACTERISTIC CRITERIA EVALUATION METHODOLOGY

GROUP	ROLE
Requester: The group or organization requesting the evaluation.	- Identify the test(s)/assessment programs to be evaluated, the purpose of evaluation and the intended user of the evaluation results. - Outline the manner in which evaluation results are (or are intended to be) used.
Implementer (a.k.a. Implementation Team): The group or organization responsible for conducting the evaluation in adherence to this methodology	- Coordinate and schedule the evaluation activities - Identify and contact appropriate technical experts to act as evaluators - Work with the Provider to identify all vendors/parties that will need to supply evidence to support evaluation - Obtain evidence from the Provider and ensure it is organized in the manner specified. - Make sure the Provider understands the rules related to the provision of evidence (e.g., meeting all requirements the first time around). - Create secure repository for materials and provide expert reviewers with access - Provide training to experts on the Independent Evaluation process and the format/location of provided materials - Facilitate the Evaluation Team meeting - Generate final evaluation report based on Evaluator decisions and feedback.
Provider: The primary group/organization responsible for providing information to the Implementation Team. The Provider may need to contact multiple groups (state agency, companies, university, organizations, technical teams) - those who are, or will be, responsible for the design, development, scoring, reporting, security and administration of the test – to acquire all of the materials necessary for evaluation.	- Identify, gather and organize appropriate evidence to inform the evaluation process - Support the development of a general overview/summary of the assessment that outlines, phase of development or implementation, history of the assessment, purpose for which it was developed, etc. - Address questions posed by Implementers and requests for clarification (as necessary) - Provide brief comments in response to the final report (if desired)
Evaluators (a.k.a., Evaluation Team): The technical experts charged with reviewing and evaluating the submitted evidence	- Review the evidence provided to support evaluation - Make comments and ratings about the adequacy of that evidence relative to particular claims and the CCSSO criteria - Comment on and approve the final report

The evaluation methodology includes four phases, as summarized in Figure 3. Each phase is briefly introduced in the section which follows and detailed guidelines for implementation are provided in the Evaluation Methodology. It is important to note that the methodology is written to address a situation where one assessment program is selected or identified for evaluation. Methodological considerations specific to the evaluation of multiple assessment programs simultaneously (for comparative purposes) and assessments developed to serve multiple states are addressed in the complete Evaluation Methodology.

FIGURE 3. PHASES OF THE TEST CHARACTERISTICS EVALUATION METHODOLOGY

Phase 1: Preparation	• Identification and collection of evidence necessary to support the evaluation process. • Development of evaluation timelines and program materials • Identification of the evaluation team • Organization of evidence and development of training materials
Phase 2: Evaluator Training and Independent Evaluator Review	• Training of the evaluation team on the assessment, process and materials • Independent evaluator review, comment and evaluation of submitted body of assessment evidence
Phase 3: Group Discussion and Summary of Evaluation Findings	• Group discussion by the Evaluation team regarding the sufficiency of evidence provided to support each claim and criterion resulting in a consensus rating and associated rationale/explanation.
Phase 4: Report Generation and Approval	• Translation of decisions and comments provided by the Evaluation team into a final evaluation report.

Phase 1: Preparation

There are three major activities that must occur before an assessment evaluation can be initiated: the evidence necessary to support review must be acquired, an appropriate Evaluation Team must be identified, and the review materials must be assembled in a secure repository.

When identifying evidence to support evaluation the Provider is encouraged to *identify the minimum amount of documentation necessary to allow for qualified technical experts to make consistent determinations regarding the extent to which a given claim has been met.* In other words, the evidence provided should be detailed and comprehensive, but it must not be a "data dump" that requires the evaluators to sift through piles of marginally relevant materials and determine what is important. Furthermore, because the evidence necessary to support the evaluation of a given assessment depends on a variety of interacting factors, those selected to conduct the evaluation must have a deep understanding of applied psychometric issues and the manner in which they interact with contextual factors to influence decisions regarding the quality, relevance and sufficiency of evidence. While some of these interactions are referenced in the CEF, the unique interplay of all possible influences can never be fully addressed.

Finally, since much of the evidence provided to support assessment review may be confidential or proprietary (e.g., test items, forms, draft procedural documentation, security protocols, etc.), detailed procedures related to the storage, delivery, and removal of submitted evidence should developed by the Implementation Team and all parties involved in the evaluation should be required to sign a non-disclosure agreement.

Phase 2: Evaluator Training and Independent Evaluator Review

The primary activities associated with Phase 2 of the methodology include 1) evaluator training on the assessment, evaluation materials and methodology and 2) independent evaluator review and evaluation of all submitted evidence.

Prior to the scheduled window for independent evaluation, the Implementation Team should conduct training for the Evaluators that serves to introduce the assessment to be reviewed and prepare evaluators for the independent review process. The purpose of the assessment overview is to provide background and contextual information relevant to the evaluation process. This would include such things as the "age" of the assessment, the target population, the standards the assessment was developed to address, the extent to which the standards have been taught/assessed in the state, the purpose of the assessment and intended uses of assessment results, and any other *relevant* factors that should be considered when conducting the evaluation.

Once the assessment has been introduced, the Implementers should provide a brief overview of the evaluation process and the materials that will be used to support it (e.g., the Criteria Evaluation Framework, evidence logs, rating sheets, etc.). Finally, a strategy for implementing the independent evaluator review should be discussed. The strategy should include a proposed process for organizing and reviewing evidence and guidelines for making claim and criterion-level ratings. Evaluators should be given be given ample time to ask questions, discuss the process and practice rating samples of evidence.

The independent evaluator review consists of the following six steps:
1. Read through entire CEF
2. Review the Evidence Log
3. Review evidence submitted for Criterion A.1 to make informal ratings at the claim-level
4. Review the evidence for the remaining Criteria
5. Based on all evidence reviewed, make preliminary holistic ratings at the criteria-level
6. Submit forms with ratings and comments to Implementation Team

While detailed procedures for conducting Phase 2 of the methodology are provided in the Evaluation Methodology, it is important to note that they represent just one of many different ways in which these activities may occur.   Specifics related to the manner in which panelists are convened (e.g., webinar vs. face-to-face) and ratings are collected (e.g., on paper or electronically), for example, may vary depending on the scope of the evaluation effort and timeline for implementation.

Phase 3: Group Discussion and Summary of Evaluation Findings

Shortly after the completion of Phase 2, the Evaluation Team should be convened for group discussion. The goal of this meeting is two-fold: 1) to establish a consensus rating on the strength of the evidence presented for each criterion and 2) articulate the components of an evidence-based argument for each rating that references the evidence reviewed in relation to the specified claims and relevant contextual factors.

During Phase 3 evaluators share their initial criterion and claim-level ratings with the goal of gauging group consensus and understanding where variability exits. Discussions should focus on identifying those claims and factors that had the greatest impact on the evaluators' ratings and areas where evidence was missing or considered inappropriate. In the end evaluators should come to consensus on: 1) the overall rating to be assigned to each criterion, and 2) influential factors that should be highlighted in the final report (i.e., included in the rationale for each rating). To the extent possible, contextual factors and relationships between criteria that were critical to the evaluation should be noted.

Phase 4:   Report Generation and Approval

Two evaluation reports will serve as the final products of the Test Characteristics evaluation: 1) a high-level **executive report**, and 2) a detailed **claim-level evaluation report**. The first report documents the consensus rating for each criterion along with a brief summary of the rationale for that rating. The executive report should also clearly indicate the assessment phase of development/implementation at the time of evaluation. This report is should be created by the Implementation Team after the completion of Phase 3. A sample template for this executive report can be found in Appendix B. The format of this template is analogous to one of the final reporting products produced by the Test Content evaluation and can be put side-by-side to create a comprehensive evaluation report for the assessment that spans all of the CCSSO Criteria.

CONTACT

Though this methodology has undergone a number of internal and external reviews including a test case, we acknowledge that there may be errors or oversights. As you review the documentation and hopefully use it to support an operational evaluation of an assessment program we encourage you to reach out to us with your feedback and commentary. For this purpose we have included the contact information of the primary developers below. Thank you.

Erika Hall, Ph.D.
ehall@nciea.org

Susan Lyons, Ph.D.
slyons@nciea.org

APPENDICES

A. Complete Structure of Claims for Evaluation of Test Characteristics
B. Sample Test Characteristics Summary Report Template

APPENDIX A:
COMPLETE STRUCTURE OF CLAIMS FOR EVALUATION OF TEST CHARACTERISTICS

A.1 Indicating progress toward college and career readiness: Scores and performance levels on assessments are mapped to determinations of college and career readiness at the high school level and for other grades being on track to college and career readiness by the time of high school graduation.

Primary claims related to the definition of CCR:
A.1.1.   College- and career readiness has been clearly defined for operational use.

Primary claims related to performance level descriptors:
A.1.2.   The process for developing performance level descriptors (PLDs) provides for PLDs that accurately represent the expectations defined by the CCR content standards within and across grades.
A.1.3.   Knowledgeable experts were involved in the process of developing and reviewing the PLDs.
A.1.4.   The process used for developing performance level descriptors (PLDs) supports their intended use(s).
A.1.5   The process for developing performance level descriptors (PLDs) includes an evaluation of alignment of the PLDs to the content of the test questions that differentiate performance at each level, and, as needed, re-writing based on new evidence concerning skills needed for success in college and careers.

Primary claims related to standard setting:
A.1.6.   A description and coherent rationale are provided for how the proposed and/or implemented standard setting methodology yields valid determinations of progress toward, or attainment of, college and career readiness.
A.1.7.   A coherent rationale accompanies methodological decisions regarding the level of involvement of grade-level educators, higher education, industry, and career technical experts (CTEs) in the standard setting process.
A.1.8.   Appropriate external CCR benchmarks and research studies are/were used in the standard setting process.
A.1.9.   Procedures and rationales for any adjustments made to proposed cut scores after the standard setting meeting are based on a defensible rationale and method.
A.1.10.   Studies planned or conducted to evaluate the validity of CCR performance standards over time are appropriate given the inferences they are intended to support.
A.1.11.   The standard setting procedures were followed as specified, and the final cut scores and the results of validity studies have been reviewed by technical experts.

Secondary claims from A.4 related to scaling: A.4.6 – A.4.10

A.2 Ensuring that assessment results are valid for required and intended purposes: Assessments produce student achievement and student growth data, as required under Title 1 of the Elementary and Secondary Education Act (ESEA) and ESEA Flexibility that provide for valid inferences that support the intended uses, such as informing:
- School effectiveness and improvement;
- Individual principal and teacher effectiveness for purposes of evaluation and identification of professional development and support needs;
- Individual student gains and performance; and
- Other purposes defined by the state.

Primary claims related to assessment design:

A.2.1   The purposes of the assessment, the target population, and each of the intended interpretations and uses of assessment results are clearly articulated.

A.2.2.   The construct or content domain of interest, how it is defined, and the rationale for that specification are clearly articulated.

A.2.3.   The assessment design reflects the construct definition and supports the intended interpretations and uses.

A.2.4.   Documentation is provided that clearly specifies the inferences and assumptions underlying the design of the assessment.

Primary claims related to validity evaluation:

A.2.5.   An outline, framework or plan summarizes those studies that have been or will be conducted to collect evidence to support the interpretive argument or validity evaluation plan, including the three primary uses as stated below. (Note: Evidence provided should include both descriptions of planned studies, and documentation /results from completed studies.)

    A.2.5.a. Evidence is provided to support the use of assessment results for making valid inferences about student performance and readiness for college and career (or on-track to CCR).

    A.2.5.b. Evidence is provided to support the use of assessment results for making valid inferences about student growth over time.

    A.2.5.c. Evidence is provided to support the use of assessment results for making valid inferences about school, principal, and teacher effectiveness (if such a use is intended) and informing improvement activities.

A.2.6.   The planned or completed validity evaluation considers the fairness of the assessment program for all examinees with respect to both intended and unintended consequences.

A.2.7.   The design and/or results of planned and/or completed validation studies were reviewed and endorsed by an independent, expert review panel (e.g., technical advisory committee).

While no secondary claims are included directly in the text, criterion A.2 is a special case in that the quality of evidence presented in support of the other criteria will directly influence judgments regarding the validity of score interpretation and use. Therefore, it is essential that when making holistic judgments regarding criterion A.2, consideration be given to the strength of support provided for all other criteria.

A.3 Ensuring that assessments are reliable: Assessments minimize error that may distort interpretations of results, estimate the magnitude of error, and inform users of its magnitude.

Primary claims related to reliability:

A.3.1.   Procedures for quantifying/ calculating reliability indices (e.g. Coefficient alpha, inter-rater reliability, classification accuracy and consistency, generalizability coefficient) and precision (e.g., standard error of measurement with associated confidence bounds, including both overall and conditional SEM, decision-accuracy indices) for each reported score are comprehensive, defensible, and well documented.

A.3.2.   Clear criteria are in place for evaluating the appropriateness of obtained reliability indices and estimates of precision.

A.3.3.   The pre-specified reliability and precision indices were estimated and the results indicate adequate support for intended uses.

Secondary claim related to informing users of reliability: D.1.2

A.4 Ensuring that assessments are designed and implemented to yield valid and consistent test score interpretations within and across years:
• Assessment forms yield consistent score meanings within and across years, as well as for various student groups, and delivery mechanisms (e.g., paper, computer, including multiple computer platforms).
• The score scales facilitate accurate and meaningful inferences about test performance.

Primary claims related to assessment development:

A.4.1.   Item design/ development materials are written at a level of detail that supports appropriate construct coverage and consistency over forms within and across years.

A.4.2.   Items undergo a comprehensive review to ensure they are appropriate, fair, accessible and likely to be interpreted by students in a consistent, accurate manner regardless of group membership or delivery mechanism.

A.4.3.   Item pilot testing and psychometric review procedures are designed to ensure items are fair for all students and provide for valid measures of student performance relative to the construct of interest.

A.4.4.   Test specifications clearly articulate what "equivalence" means from content (KSAs), format and statistical perspectives.

A.4.5.   A comprehensive test review process is in place to ensure test forms meet the content and statistical requirements outlined in the test specifications.

Primary claims related to scaling and equating:

A.4.6.   The design of the scale accounts for the design of the assessment and the manner in which results are intended to be interpreted and used.

A.4.7.   The procedures used to estimate student performance and translate these estimates to a different scale are transparent, fair, and consistent with the reported meaning of the scale scores.

A.4.8.   Procedures for scoring items or sections that involve human judgment (e.g. performance tasks, essays) support accurate and consistent scoring within and across items, forms, administrations, and sub-groups by minimizing construct-irrelevant score variance within and across scorers.

A.4.9.   Linking and/or equating procedures are clearly specified, comprehensive, and demonstratively appropriate.

A.4.10.   The scaling and linking/equating procedures were followed as specified, and the results have been reviewed and accepted by technical experts.

Secondary claims from E.1 related to assessment standardization: E.1.1, E.1.2

A.5 Providing accessibility to all students, including English learners and students with disabilities.
   • Assessments provide for reliable scores and valid score interpretations related to intended use for English learners.
   • Assessments provide for reliable scores and valid score interpretations related to intended use for students with
      disabilities.

Primary claims relating to the test characteristics associated with accessibility
   A.5.1.   The testing user interface and item format does not introduce construct-irrelevant variance that impedes student
            performance.
   A.5.2.   Students are matched with appropriate accommodations/ accessibility features.
   A.5.3   Score reliability is appropriately estimated and evaluated for English learners and students with disabilities (SWD).
   A.5.4   Validity evidence supports the intended use and interpretation of scores for English learners and students with
            disabilities (SWD).

Secondary claims from A.4 related to item development & review: A.4.2 – A.4.3

A.7   Meeting all requirements for data privacy and access: All assessments must meet federal and state requirements for
student privacy, and all data must be readily accessible by the state.

Primary claims related to student privacy:
   A.7.1.   Adequate steps have been taken to ensure compliance with Federal Educational Rights and Privacy Act (FERPA) and
            any additional state regulations related to maintaining student privacy.
   A.7.2.   Comprehensive procedures are in place to protect personally identifiable information (PII) from unauthorized
            access or use.
   A.7.3.   Procedures are in place to ensure all data is managed securely.

Primary claims related to data access:
   A.7.4.   An assurance is provided of state ownership of all required data reflecting compliance with state laws.
   A.7.5.   Procedures and timelines are in place to ensure a state   is provided with all data necessary to support desired
            analyses (e.g., forensics, quality control, accountability calculations) in a timely and useable fashion.
   A.7.6.   Procedures are defined for how data will be securely transferred between vendors and the state, and stored or
            destroyed after administration/reporting.

Secondary claims from E.1 related to security of test materials: E.1.3-E.1.6

D.1 Focusing on student achievement and progress to readiness: Score reports illustrate a student's progress on the
continuum toward college and career readiness, grade by grade and course by course. Reports stress the most important
content skills and processes and how the assessment focuses on them to show whether or not students are on track to
readiness.

Primary claims related to score report content and format:
   D.1.1.   The content and format of the score reports are consistent with and supported by the assessment design, and the
            psychometric procedures for developing the scale(s), and support the intended uses.
   D.1.2.   Score reports support inferences regarding student achievement relative to key content and performance
            standards.
   D.1.3.   Score reports provide for valid inferences regarding career and college readiness, or on-track to CCR.

D.2 Providing timely data that inform instruction: Reports are instructionally valuable, easy to understand by all audiences and delivered in time to provide useful, actionable data to students, parents and teachers.

Primary claims related to timeliness of score reports:
- D.2.1.   Directions for accessing and viewing score reports (when necessary) are broadly distributed and clear to end-users.
- D.2.2.   Reporting timelines, procedures and technology provide for the dissemination of test results in a timely fashion.

Primary claims related to instructional utility of score reports:
- D.2.3.   The content and structure of score reports provide useful and actionable information for making instructional decisions.

E.1 Maintaining necessary standardization and ensuring test security: in order to ensure the validity, fairness and integrity of state test results, the assessment systems maintain the security of the items and tests as well as the answer documents and related ancillary materials that result from test administration.

Primary claims related to standardization:
- E.1.1.   Test distribution and administration directions are clear and sufficiently scripted to provide for standardization.
- E.1.2.   Procedures for training and monitoring test administrators are effective and well documented.

Primary claims related to security:
- E.1.3.   Comprehensive procedures are in place to ensure the security of assessment materials.
- E.1.4.   Effective test security training is provided for all personnel who come into contact with test materials.
- E.1.5.   Procedures are in place to test and validate the effectiveness of security safeguards.
- E.1.6.   Activities construed as cheating or other breaches of test security are clearly defined and transparent.
- E.1.7.   Detailed procedures are in place to support the detection of testing irregularities.
- E.1.8.   Clearly documented procedures and specifications are provided for responding to breaches in test security.

APPENDIX B:
SAMPLE TEST CHARACTERISTICS SUMMARY REPORT TEMPLATE

Assessment Phase of Development: _____

	Degree of Match with CCSSO Criteria			
	Weak	Limited	Good	Excellent
A.1 Indicating progress toward college and career readiness: Scores and performance levels on assessments are mapped to determinations of college and career readiness at the high school level and for other grades to being on track to college and career readiness by the time of high school graduation.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
A.2 Ensuring that assessments are valid for required and intended purposes: Assessments produce data, including student achievement data and student growth data required under Title I of the Elementary and Secondary Education Act (ESEA) and ESEA Flexibility, that can be used to validly inform the following: • School effectiveness and improvement; • Individual principal and teacher effectiveness for purposes of evaluation and identification of professional development and support needs; • Individual student gains and performance; and • Other purposes defined by the state.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
A.3 Ensuring that assessments are reliable: Assessments minimize error that may distort interpretations of results, estimate the magnitude of error, and inform users of its magnitude.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
A.4 Ensuring that assessments are designed and implemented to yield valid and consistent test score interpretations within and across years: • Assessment forms yield consistent score meanings within and across years, as well as for various student groups, and delivery mechanisms (e.g., paper, computer, including multiple computer platforms). • The score scales facilitate accurate and meaningful inferences about test performance.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
A.5 Providing accessibility to all students, including English learners and students with disabilities: • Assessments produce valid and reliable scores for English learners • Assessments produce valid and reliable scores for students with disabilities.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				

	Degree of Match with CCSSO Criteria			
	Weak	**Limited**	**Good**	**Excellent**
A.7 Meeting all requirements for data privacy and ownership: All assessments must meet federal and state requirements for student privacy, and all data is owned exclusively by the state.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
D.1 Focusing on student achievement and progress to readiness: Score reports illustrate a student's progress on the continuum toward college and career readiness, grade by grade, and course by course. Reports stress the most important content, skills, and processes, and how the assessment focuses on them, to show whether or not students are on track to readiness.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
D.2 Providing timely data that inform instruction: Reports are instructionally valuable, easy to understand by all audiences and delivered in time to provide useful, actionable data to students, parents and teachers.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				
E.1 Maintaining necessary standardization and ensuring test security: in order to ensure the validity, fairness and integrity of state test results, the assessment systems maintain the security of the items and tests as well as the answer documents and related ancillary materials that result from test administration.	🔴	🟡	🟢	🟢
[[summary of rationale and other comments]]				