# MEMBER PROCEDURES MANUAL

## Smarter Balanced Assessment Consortium

October 31, 2023

# TABLE OF CONTENTS

Note: Sections highlighted in gray designate a change from the 2022-23 Member Procedures Manual.

Additional information about the changes is available in the Revision Log at the end of this document.

# 1.0 INTRODUCTION

The *Member Procedures Manual* is designed to help member leadership prepare for the administration of the Smarter Balanced assessments. It includes information for the summative assessments, interim assessments, data warehouse/reporting system, and Tools for Teachers.

The manual provides a general overview of Smarter Balanced policy topics such as test security, test scheduling, and general administration as it relates to the summative and interim assessments. In addition, this manual provides a high-level overview of member and/or district responsibilities, services provided by Smarter Balanced, and examples of services that members are responsible for securing. For each applicable section, the manual will clearly indicate member responsibility through the following three distinct roles:

- The role of the member in making key policy decisions and setting the stage for valid test administration;

- The role of the member in monitoring the implementation of member and Smarter Balanced policies and procedures during test administration; and

- The role of the member in gathering and documenting evidence of valid test administration, scoring, and other key information for internal use, Consortium validity evidence, and for future U.S. Department of Education peer review of state assessment systems, as applicable.

In order for Smarter Balanced to be sustained and to afford members the opportunity to derive scores that are comparable, it is essential that members understand the critical nature of their responsibility in light of these roles. In addition, this manual is intended to support members in the key conversations they must have with their service providers to facilitate implementation. Toward that end, each section will include links to relevant resources with more detailed information to support implementation.

The Implementation Guide for States and Service Providers complements this manual. This document is intended for use by new states who plan to procure Smarter Balanced assessment system solutions. It may also be used by service providers who respond to state Request for Proposals (RFPs) or who contract directly with districts to provide interim assessments and Tools for Teachers services.

Most materials referenced in this manual are posted in the Smarter Balanced Secure File Transfer Server, the Smarter Balanced public website, the Smarter Content Explorer,

SmarterApp.org or the Smarter Balanced Member Portal. The location of online resources is included throughout this manual.

## 1.1 UC SANTA CRUZ SILICON VALLEY EXTENSION

The Smarter Balanced Assessment Consortium is part of the University of California, Santa Cruz (UCSC) as a unit operating under the Silicon Valley College of Extension. Each member's Smarter Balanced *Memorandum of Understanding and Agreement* (MOU) outlines the various contractual, fiscal, and procedural responsibilities of both member states and UCSC.

## 1.2 SMARTER BALANCED TECHNICAL ADVISORY COMMITTEE

The Smarter Balanced Technical Advisory Committee (TAC) provides guidance on technical assessment matters pertaining to validity and reliability, accuracy, and fairness. Members of the TAC are highly regarded national experts who have been widely published in their fields. Areas of expertise include assessment design; computer adaptive testing (CAT); assessment accessibility resources; uses of tests; mathematics and English language arts/literacy. Information about current TAC members is posted on the Governance Structure page of the Smarter Balanced website. The TAC meets in person twice each year at the fall and spring TAC Meeting and Collaboration Conference and virtually twice each year in summer and winter.

Smarter Balanced staff is also available to support member TAC meetings. Upon advanced notice, the Executive Director may be available in person or virtually to address important questions or issues that may have political implications. With advanced notice, the Smarter Balanced Director of Psychometrics may be available to participate in member TAC meetings virtually to assist with Smarter Balanced psychometric topics.

# 2.0 POST-ADMINISTRATION EVIDENCE COLLECTION AND DOCUMENTATION

Smarter Balanced Governing Members that administer any part of the Smarter Balanced summative assessment should document and maintain evidence of implementation and test administration policies for internal use, Consortium validity evidence, and for future U.S. Department of Education peer review of state assessment systems. This evidence includes:

- Documented scoring procedures and evidence of inter-rater reliability
- Bring Your Own Device policy
- Omission of any accommodation or designated support
- Private School/Home-Instructed Student policy
- Test Security policy

Smarter Balanced maintains some implementation and test administration policies in the *Member Test* Administration *Policies* document available on the [K-12 Leads page](#) of the Member Portal. This information is updated annually.

# 3.0 TEST REGISTRATION TOOL AND SINGLE SIGN-ON (SSO)

## 3.1 TEST REGISTRATION TOOL

During the development phase, Smarter Balanced developed an open-source Administration and Registration Tool (ART) for use with other Smarter Balanced open-source systems (e.g., test delivery system, Data Warehouse). ART allowed authorized users to create and manage records for students, users and entities, such as states, districts and institutions. Users could upload records in groups or manage them individually and also manage the assessment settings and accessibility resource information for students.

The test registration tool is the user role–driven application that provides administrators with the tools to add and manage users and students participating in the Smarter Balanced assessments. This system should use a role-specific design to restrict access to certain tools and applications based on the user's designated role and responsibilities.

States will need to procure a test registration tool to register students for the summative and interim assessments. Important registration tool features are described below.

### 3.1.1 Test Registration Tool Features

This portion of the tool should capture

- Test information, including testing windows, test opportunities, delay rules, and implicit eligibility rules;
- Institution information, such as information on schools, districts, and states;
- User and user role information;
- Student and student accessibility resources information; and
- Student group information.

Test administrator or teacher information may also be captured during test registration.

States should work with their service providers to configure their systems to reflect the roles/permissions they want implemented in their organization for the summative and interim assessments. Roles such as adding/deleting students, adding/editing designated supports and accommodations, and editing student demographics should be configured to meet the state's preferences. States will need to monitor their configuration options to validate that the

roles/permissions are appropriate and do not pose risk to the validity of the test. States should describe their requirements or their process for monitoring appropriate user role access in their procurement.

### 3.1.2 Member/State User Roles

Member/State users will be responsible for some or all of the tasks listed in Table 1 (responsibilities vary from member to member). In addition, members will be responsible for defining the responsibilities of other roles in their state/territory, both within their student and teacher administration and registration systems and for general test administration responsibilities.

Table 1: Example Member User Role

| User Role | Description |
|---|---|
| Member/State Users | Member/State Users will be responsible for:<br><br>▶ uploading, adding, and editing student information into the test registration tool unless delegated to local education agencies (LEAs);<br><br>▶ adding other member users, such as district administrators (DAs), district test coordinators (DCs), school test coordinators (SCs), and test administrators (TAs) into the student and teacher registration system;<br><br>▶ reviewing and resolving test security incidents as submitted by LEAs;<br><br>▶ approving appeals (e.g., resets, reopens, restores, invalidations); and<br><br>▶ coordinating with Smarter Balanced for paper administration files.<br><br>The initial member user assignment is managed by the member Help Desk. Member users can then add other users. |

### 3.1.3 Accessibility Resources

Student accessibility resource information should also be captured in the test registration system. To facilitate the selection of specific student designated supports and accommodation needs, Smarter Balanced created an Individual Student Assessment Accessibility Profile (ISAAP) process and tool that states can use. For students requiring one or more designated supports and/or accommodations, schools will be able to document this need prior to test administration. Furthermore, the ISAAP can include information about universal tools that may need to be turned off for an individual student. The ISAAP is available to states for their use, but its use is not required. The Understanding Individual Student Assessment Accessibility Profile video and Individual Student Assessment Accessibility Profile: A Tool to Support Students video are two resources that explain the ISAAP tool and how it may be used. States may use their own processes or tools for this purpose. See *Section 7.2* for additional information about accessibility resources for students.

### Identifying Needs Prior to Test Administration

By documenting individual student accessibility needs prior to test administration, the state's test administration delivery system will be able to activate the specified options when the student logs in to an assessment. In this way, the ISAAP allows educators and schools to focus on each individual student to document the designated supports and accommodations required to ensure that the assessed content is accessible for participating students. Whether the state uses ISAAP or its own system, the state and its service provider will still need to determine how the information identifying the accessibility resources needed by individual students will be entered into the state's test registration system.

The test registration tool is the user role-driven application that provides administrators with the tools to add and manage designated supports and accommodations for students participating in the Smarter Balanced assessments. This system should use a role-specific design to restrict access to certain tools and applications based on the user's designated role.

### Member Responsibility

Members will work with their service providers to configure their systems to reflect the roles/permissions they want implemented in their organization for the summative (and interim) assessments. Roles such as adding/deleting student profiles, adding/editing universal tools, designated supports, and accommodations, and editing student demographics should be configured to meet the member's preferences.

### Monitoring

Members should monitor their configuration options to validate that the roles/permissions are appropriate and do not pose risk to the validity of the test. For more information, see the *Online Summative TAM* and the *Smarter Balanced Test Registration Tool User Guide Template.* Both documents are available in the Document Library of the Smarter Balanced Member Portal and in the Assessment and Training Operations folder of the Secure File Transfer Server.

### RECOMMENDATION REGARDING DOCUMENTATION OF EVIDENCE

Due to varying member structures and capacity, some members may choose to configure user roles differently from the default user-role settings defined in the *Test Registration Tool User Guide Template.*

In their evidence documentation, members that take advantage of this opportunity, should plan to include their process for monitoring appropriate user role access. See section 1.3 of the Online Summative TAM for sample user roles.

## 3.2 SINGLE SIGN-ON (SSO)

A Single Sign-On (SSO) and Access Management system provides central management of educator accounts and permissions across multiple applications. Centralizing access management services allows educators to use one login to access multiple services and lets states manage only one account per educator. In many cases, SSO services are among the services that states receive under contract from their assessment service provider. However, some states contract for SSO separately from their assessment services. This arrangement enables them to use the same SSO system for all their educator services and across student information systems, learning object repositories, and assessment platforms.

Smarter Balanced uses a federated authentication system in which its Tools for Teachers and Reporting services integrate with state provided SSO solutions. Service providers pass authorization information to Smarter Balanced for applications (Smarter Balanced Reporting System, Tools for Teachers) used by the state or district.

**Other Documentation**

- *Smarter Balanced Assessment Consortium: Accessibility and Accommodations Framework*
- *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines*
- *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Implementation Guide*
- *Understanding the Individual Student Assessment Accessibility Profile*
- *Individual Student Assessment Accessibility Profile: A Tool to Support Students*
- Smarter Balanced Assessment Consortium: *Test Registration Tool User Guide Template*

# 4.0 ONLINE TEST DELIVERY SYSTEM

Members must select a service provider with an online test delivery system capable of delivering and scoring the Smarter Balanced assessment.

**Member Responsibility**

Members are responsible for *reviewing evidence* that the selected service provider's test delivery systems do the following:

- Render test items correctly: They should look the same as the items hosted on the Smarter Balanced Sample Items Website.

- Render accessibility resources correctly: Service providers should consult the [Usability Accessibility and Accommodations Guidelines](#) and the [Item Types and Accessibility Features](#) document available in SmarterApp.

- Collect and score the student's response correctly.

- Implement an effective computer-adaptive testing algorithm.

- Score the test correctly.

- Require test-taking devices that can be secured and meet Smarter Balanced requirements. See the [Student Device and Secure Browser Requirements](#).

## 4.1 CONFIGURATION OF THE TEST DELIVERY SYSTEM

Service providers typically support client-specific customizations of their test delivery systems. Members should consider how configuring certain aspects of test delivery may impact security of the test and comparability with other members in the Consortium.

Configuration settings that may affect security:

- Which user roles can add or edit student registration and eligibility?

- Which user roles can enter appeals?

- Must a proctor be associated with the school where he or she is administering the test?

- Who assigns roles and to whom are they assigned?

- When can students take breaks? How long do those breaks last? Under what conditions can students review questions presented before the break?

Configuration settings that may affect comparability:

- How is the computer-adaptive engine configured? How is it constrained to the blueprint?

- How are accessibility resources managed? Who can assign them to students?

- How are machine and human scoring types managed? What processes ensure inter-rater reliability?

**Member Responsibility**

Members are responsible for reviewing these choices and determining what changes need to be made based on Smarter Balanced recommendations and each member's policies. Members should review their customized configurations on an annual basis to determine if they are still in accord with member policy and Smarter Balanced guidelines.

**Recommendation Regarding Documentation of Evidence**

In their evidence documentation, members should plan to include a list of user role configurations that describe:

- What user roles can add/edit student profiles in the systems?

- What user roles can enter appeals?

- What user roles can add/edit test settings (designated supports and accommodations)?

## 4.2 TEST DELIVERY SYSTEM DEPLOYMENT

In support of members' implementation efforts, Smarter Balanced is willing to assist members in the design of a process and in the reviewing of any chosen platform to determine if it meets a core set of implementation protocols and requirements. This voluntary review would include an evaluation of correct item rendering, accommodations functionality, item scoring procedures, claim and overall score generation, the delivery of results to the data warehouse, and other essential functions. If the evaluation is associated with an RFP or other competitive bidding process, Smarter Balanced and the member should first determine how to ensure this assistance is in accordance with state procurement requirements.

## 4.3 TEST AND ITEM MANAGEMENT SYSTEM (TIMS)

Smarter Balanced releases content and test packages to member service providers for use in a given school year. During the ingest and review of the test packages, service providers and members may provide feedback regarding the content of the packages and suggest revisions. Smarter Balanced receives and reviews requests for revisions via a process called Content Updates.

Feedback about the content of test packages is provided through the Test and Item Management System (TIMS). Smarter Balanced reviews the feedback and makes updates:

- as appropriate based on industry standards, and psychometric and content expert analyses; and

- when the changes can be made at a time that maintains the integrity of all state testing windows.

Updates to the content of packages are released on Thursdays.

Consistent with the practice of adhering to industry standards and for maintaining the integrity of testing windows, Smarter Balanced members set milestone dates that establish the period in which testing windows are not considered as a decision factor. After these milestone dates, only urgent and essential updates may be made for the current school year's packages. These dates are anticipated to be:

- Practice & Training Tests: September 1, 2023

- Interim Assessments: September 1, 2023

- Summative Assessments: January 11, 2024

After the dates above have passed, Smarter Balanced will review any urgent feedback about items with member representatives to identify if the content updates should be made. All other non-urgent updates to content will be included in a future year's test package.

The TIMS system is also used to author Instructional Resources. For more information, see *Section 15.0 Tools for Teachers* in this guide.

## 4.4 BRING YOUR OWN DEVICE (BYOD) LOCAL OPTION

Members may allow schools and districts to adopt and implement a policy regarding Bring Your Own Device (BYOD).

See the Implementation Guide for States and Service Providers for additional information regarding the components, functionality, and configuration of a test delivery system. This document is available to members in the Manuals and User Guide section of the Member Portal Document Library. It is also available to service providers in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer Server.

**Member Responsibility**

Members should determine if they will allow this option and define the member-specific policy for their districts and schools. The policy must include an explicit requirement that all devices used for testing must meet the Smarter Balanced Student Device and Secure Browser Requirements including the installation and use of a service provider-supported secure browser.

**Recommendation Regarding Documentation of Evidence**

In their evidence documentation, members should plan to include their "Bring Your Own Device" policy, if applicable.

# 5.0 RESOURCES AVAILABLE TO SUPPORT MEMBERS

Smarter Balanced has developed a variety of materials to support members.

## 5.1 TEST ADMINISTRATION RESOURCES

### 5.1.1 Online Summative Test Administration Manual

The Smarter Balanced Online Summative Test Administration Manual (TAM) is a customizable document designed to support the administration of the Smarter Balanced online summative assessment. The Online Summative TAM provides test coordinators and test administrators with information about help desk services, test security, testing irregularities and breaches, accessibility supports for students, recommendations for scheduling test sessions for the computer adaptive test (CAT) and performance task (PT) portions of the test, tasks to complete prior to, during, and after testing, and test directions to read during test administration. The *Braille Test Administration Manual Insert* provides information about administering a braille version of the test.

### 5.1.2 Supplement: Online Summative Test Administration Manual for Remote Test Administration

The *Supplement: Online Summative TAM Supplement for Remote Test Administration* includes additional protocols and procedures specific to the remote administration of the Smarter Balanced English language arts/literacy and mathematics summative assessments. The addendum is intended to be used as a supplement to the member-customized Online Summative Test Administration Manual.

The Online Summative TAM and the Supplement Online Summative TAM for Remote Test Administration are available in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer Server and in the Document Library of the Member Portal.

### 5.1.3 Test Administrator User Guide Template

The Test Administrator User Guide template was developed for use with the Smarter Balanced open-source test delivery system (TDS). The guide describes the functionality of the TDS including the Test Administrator (TA) Interface and Student Interface. Service providers may use this template to develop a customized Test Administrator User Guide for their proprietary TDS. The Test Administrator User Guide Template is available in SmarterApp.org.

### 5.1.4 Interim Assessment Guide for Administration

The Interim Assessment Guide for Administration is a customizable document that provides information about the Smarter Balanced interim assessments in English language arts/literacy and mathematics. The document includes information about interim test administration

options, the Smarter Balanced Reporting System, and Tools for Teachers. States and their service providers may use this document to update their existing interim test administration manuals, as needed. This guide may also be used for district-level implementation of interim assessments in member states that have not purchased the complete package and for statewide implementation of the interim assessments in non-member states (potentially by service providers that have not previously delivered Smarter Balanced assessments). The Interim Assessment Guide for Administration is available in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer Server and in the Document Library of the Member Portal.

### 5.1.5 Online Training Modules

The Online Test Administration Training Modules are customizable training resources available in the Assessment Training and Operations/Training Modules folder of the Secure File Transfer Server. The modules are available in non-narrated PowerPoint format to allow for customization by members and service providers. Some modules are available in narrated format if customization is not needed, or if there are only slight customizations that do not affect the existing narration.

- Module 1: What is a CAT?

- Module 2: Technology Requirements for Online Testing

- Module 3: Performance Task Overview

- Module 4: Student Interface for Online Testing

- Module 5: Test Administrator Interface for Online Testing

- Module 6: Administration and Registration Tools (ART) [This module may be used as a template for the creation of a training module for the service provider's test registration tool.]

- Module 7: Accessibility and Accommodations

- Module 8: Embedded Universal Tools and Online Features

- Module 9: Test Administration Overview

# 5.2 POLICY AND IMPLEMENTATION RESOURCES

### 5.2.1 Member Procedures Manual

This document, the Member Procedures Manual, is designed to help member leadership prepare for the administration of the Smarter Balanced assessments. It includes information for the summative assessments, interim assessments, data warehouse/reporting system, and Tools for Teachers. The manual provides a general overview of Smarter Balanced policy topics such as test security, test scheduling, and general administration as it relates to the summative and interim assessments. In addition, this manual provides a high-level overview of member and/or district responsibilities, services provided by Smarter Balanced, and examples of services that members are responsible for securing. This manual is available to members in the Document Library of the Member Portal.

### 5.2.2 Usability, Accessibility, and Accommodations Resources

Smarter Balanced strives to provide every student with a positive and productive assessment experience, generating results that are a fair and accurate estimate of each student's achievement. To this end, Smarter Balanced provides members with several resources that describe the universal tools, designated supports and accommodations for the Smarter Balanced English language arts (ELA)/literacy and mathematics. These tools, supports, and accommodations are available in embedded and non-embedded versions depending on whether they are provided as digitally delivered components of the test administration system (embedded) or separate from it (non-embedded). These resources include:

- The Smarter Balanced Assessment Consortium's *Usability, Accessibility, and Accommodations Guidelines (*UAAG), available on the Accessibility and Accommodations page of the Smarter Balanced Website, focus on universal tools, designated supports, and accommodations for the Smarter Balanced assessments of English language arts (ELA)/literacy and mathematics.

- The *Usability, Accessibility, and Accommodations Implementation Guidelines*, available on the Accessibility and Accommodations page of the Smarter Balanced Website, provides members with both (a) clarification of the critical information to convey to districts and schools about the Smarter Balanced *UAAG* document, and (b) strategies for ensuring that the information in that document is conveyed to districts and schools.

- Scribing and Read Aloud Guidelines are included as appendices in the *UAAG* and are also posted on the Accessibility and Accommodations page of the Smarter Balanced website.

- The Smarter Accessibility page helps members identify similarities between the assessment resources and classroom practices. The accessibility strategies correspond to the resources that are currently included in the UAAG and assist educators and decision-makers by providing a description of both the resource and its classroom equivalent.

Members should distribute/communicate the *UAAG* to their schools and districts and provide additional support as needed. For students completing a paper-pencil assessment, members will need to work with schools to ensure all applicable universal tools, designated supports, and accommodations are available.

### 5.2.3 Implementation Guide for States and Service Providers

The *Implementation Guide for States and Service Providers* describes critical elements of Consortium policies and practices that states and service providers may consider in developing or responding to Requests for Proposals (RFP) regarding the Smarter Balanced Assessment System. This document is available to members in the Manuals and User Guide section of the Member Portal Document Library. It is also available to service providers in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer Server.

**Member Responsibility**

Some of the materials, such as the *Online, Summative, Test Administration Manual* (*Online TAM*), the *Interim Assessment Guide for Administration* and the test administration training modules will require customization by each member to meet their unique needs. Other materials, such as the *Usability, Accessibility, and Accommodations Guidelines,* the *Member Procedures Manual* and the *Implementation Guide for States and Service Providers* are policy documents and should not be customized unless an applicable law, regulation, or policy requires the omission of an accessibility feature or other change.

For all materials developed to support the Smarter Balanced assessments, members should determine the appropriate mechanism for posting and disseminating the information to their districts and schools.

Members may identify a need for additional materials beyond those developed by Smarter Balanced (e.g., scoring guides, released item booklets, additional administration manuals by grade/content area, assistive technology manuals) and will be responsible for developing or working with their service providers on any additional materials the member may desire.

## 5.3 PRACTICE TESTS, TRAINING TESTS, AND RELEASED ITEMS

Practice and Training Tests are available in grades 3-8 and high school. Each Smarter Balanced state and territory hosts the practice and training tests on their own test delivery system. This allows students and educators to become familiar with how the tests are presented on their test delivery systems. The content for the Practice and Training tests is also available in the Test Packages/2023-24 and the Test Packages/2023-24 Practice Training Test Packages folder of the Secure File Transfer Server.

| | Practice Test and Training Test |
|---|---|
| Practice Test | Practice Tests include test items and performance tasks for each grade level (3–8 and high school). The Practice Tests provide a preview of the item types included in the Smarter Balanced assessments. |
| Training Test | Training Tests allow students to become familiar with the format and functionality of the online test. They include an example of every interaction type that a student will encounter in their particular grade band. This resource is available by grade band (3–5, 6–8, and high school) and is intentionally brief, consisting of fewer than ten items per content area. |

Users (educators, students and stakeholders) can log into the Practice and Training Tests as a "guest" and do not require login credentials.

Practice and Training Tests do not require the use of the secure browser except for the use of some accommodations which require use of assistive technology. Secure browsers may be used to become familiar with the functionality of the test delivery system with the student's assistive technology software or device. Test administrators may also access the Practice Test using the Test Administrator (TA) Interface to practice setting up test sessions.

Smarter Balanced hosts the Sample Items Website for public use that includes released, formerly operational test questions from the summative assessment. The sample questions illustrate the various item types available on the Smarter Balanced assessments and feature the full suite of accessibility resources, including universal tools, designated supports and accommodations to support students' needs.

The Smarter Annotated Response Tool (SmART) helps educators understand how student writing is scored on Smarter Balanced assessments. It provides annotated anchors of sample student responses for released Smarter Balanced practice test items that may be used to support student understanding of scoring rubrics.

**Member Responsibility**

The Practice and Training tests should be hosted on each Consortium member's public website or assessment portal to ensure that students' experiences mirror the operational test administration experience. The Practice and Training tests use guest credentials; however, the Test Administrator (TA) Interface is required to simulate print-on-demand and braille embossing in a training context. Members' service providers should load the practice and training tests into their testing environment so that TAs can perform training on setting up a test session using the TA Interface and the Practice Test. Members will need to customize the Online TAM with their specific process for getting user logins to the TA Interface.

### 5.3.1 Member Use of Released Items

This policy clarifies member's authority to import released items into a test delivery platform or learning management system.

Members may import released items or sample items into a contracted vendor managed or state managed:

- test delivery system;
- learning management system; or
- alternate web application or portal.

The system provides either authorized adults or eligible students access to the assessment items and associated resources as appropriate when:

- The state is a member in good standing
- The member executes a contract and non-disclosure agreement with the vendor as necessary to protect the intellectual property of the Consortium
- The member notifies Smarter Balanced of the system they are using
- If the system allows items and stimuli to be printed or rendered as a PDF, and the copies attribute passage copyright to passage authors and item ownership copyright to the Consortium's fiscal agent.
- The system has terms and conditions that prevent users from using their system to import content into a new system
- Items may be available to teachers and students in tests and as individual items
- Items are not amended or altered
- Items are not added to stimuli or passages that are included in the released item set. If members seek to amend passages with additional items, they may acquire their own copyright arrangement for the relevant passage.
- At least annually, Smarter Balanced will identify changes to the released items. Changes may include removing items when the Copyright Clearance Center removes access to specific content. Except in extreme situations, members will have at least 1-year's notice of such changes.

**Member Responsibility**

Members who opt to use released items in a contracted vendor managed or state managed test delivery system, learning management system, or alternate web application or portal must abide by the above policy.

# 6.0 TEST SECURITY

## 6.1 SUMMATIVE ASSESSMENTS TEST SECURITY

Members of the Smarter Balanced Assessment Consortium have a vested interest in ensuring that assessments are supported by security protocols that establish both fairness for student engagement and validity in the interpretation of results. With regard to the Smarter Balanced assessment system, security needs are amplified due to the increased consequences associated with broader membership.

The action or inaction of one Consortium member may have consequences for the Consortium as a whole. Disclosure or dissemination of secure assessment items or other secure materials associated with the test will adversely affect the validity of the assessment results across the Consortium. The confidentiality of assessment items and corresponding student responses is paramount in maintaining the integrity and validity of any assessment. For these reasons, Consortium members administering assessments provided by Smarter Balanced are expected to take appropriate steps to assure the security of the assessments.

**Member Responsibility**

Prior to beginning any secure Smarter Balanced assessment, Consortium members must have processes in place to support test security for both online and paper-and-pencil (P&P) administrations of the summative assessments (as applicable).

With regard to overall security of the Consortium assessments, each member will be required to have in place:

- Comprehensive protocols to respond to possible security breaches (including test and/or item exposure). Minimum standards describing how to distinguish security breaches from other testing incidents are included in section 4.0 of the Smarter Balanced *Online, Summative, Test Administration Manual (Online TAM)* and *Appendix C Test Security Chart* of this manual).

- Plans supporting adequate training on test security procedures for Test Administrators, Test Coordinators, and other district/school staff such as principals, teachers, proctors, and staff providing non-embedded accessibility resources. Such training should include, but not be limited to, training on item security, adherence to *Online Summative TAM* policies, etc.

With regard to overall test security, the Consortium recommends that each member have in place a process and associated timeline for ensuring that:

- Test Administrators (and any other individuals who will be administering any secure Smarter Balanced assessments) have read and understand the Smarter Balanced *Online Summative TAM*, the Smarter Balanced *Usability, Accessibility, and Accommodations*

*Guidelines*, and viewed the associated Smarter Balanced training modules available in the Assessment Training and Operations folder of the Secure File Transfer Server.

- Test Administrators communicate test security information to students.

- There are clearly defined protocols that describe which individuals (staff or otherwise) may be designated as Test Administrators or in other roles related to administering a secure Smarter Balanced assessment. We recommend verifying the process and how designation are made by those responsible for test administration on behalf of the member.

- There is a process in place for monitoring social media for the posting of any secure assessment materials. See *Appendix D: Guidance for Social Media Summative Assessment Monitoring* for more information.

### Member Monitoring

Members will be required to develop procedures that allow for ongoing monitoring of their test security protocols.

### Resources

See *Appendix D: Guidance for Social Media Summative Assessment Monitoring* for more information.

### Recommendation Regarding Documentation of Evidence

In their evidence documentation, members should plan to include their test security policies and procedures.

## 6.2 INTERIM ASSESSMENTS TEST SECURITY

Interim assessments can serve a variety of educator needs. To better support the range of possible uses consistent with member education agency policies, educators may establish the timeframe, administration policies, and scoring practices for interim assessments. The interim assessments are considered student- and teacher-facing. The student- and teacher-facing designation provides educators with the flexibility to access the test questions and their students' responses to the test questions. Because of this flexibility, the interim assessments are not intended to be used for accountability purposes. **Interim assessments are not for public use, display, or distribution.** This allows educators to use the interim assessments in the intended manner. For this reason, any use, display, or distribution of the interim assessments that results in access to individuals beyond authorized local education agency staff and students is prohibited. The interim assessments also include all the accessibility resources that are available in the summative assessment to provide accurate results for all students. Finally,

interim assessment items must not be copied into a third-party application that is not approved by the member state education agency and in consultation with Smarter Balanced.

Note: A state or territory may establish a policy that allows for interim assessments to be administered without the use of a secure browser provided that the above test security protocols are followed.

In addition, interim assessments may be administered in a non-standardized manner including at grade levels that are different from students' enrolled grades.

# 7.0. SUMMATIVE ASSESSMENT

## 7.1 PARTICIPATION

### 7.1.1 Participation—Who Should Participate?

Consistent with the Smarter Balanced testing plan, all students in eligible grade levels, including students with disabilities, English learners (ELs), ELs with disabilities, and other diverse students should have equal opportunity to participate in the Smarter Balanced assessments. Important exceptions to this plan are noted below.

Smarter Balanced has developed a comprehensive accessibility and accommodations framework, *Usability, Accessibility, and Accommodations Guidelines (UAAG),* for all participating students and those with special assessment needs. The Consortium has also developed a variety of innovative digital accessibility resources that are embedded in the test delivery engine as well as a variety of designated supports and accommodations.

All students enrolled in grades 3–8 and high school are required to participate in the Smarter Balanced **mathematics** test *except:*

- Students with the most significant cognitive disabilities who meet the member's criteria for the mathematics alternate assessment based on alternate achievement standards (approximately 1% or less of the student population).

All students enrolled in grades 3–8 and high school are required to participate in the Smarter Balanced **English language arts/literacy** test *except:*

- Students with the most significant cognitive disabilities who meet the member's criteria for the English language arts/literacy alternate assessment based on alternate achievement standards (approximately 1% or less of the student population).

- ELs who are enrolled for the first year in a U.S. school. These students may instead participate in their required English language proficiency assessment.

State laws should be consulted and followed as they relate to student participation in Smarter Balanced summative assessments.

### 7.1.2 Participation and Performance Calculations

To describe the technical features of the assessments in its technical manuals and to answer Consortium level questions that will be asked by stakeholders, the Consortium must:

- Specify the number of students who participated in the test; and
- Calculate the percent of students who scored at each performance level.

In addition, research and analysis of the test (e.g., differences in mean scale scores, analysis of growth.) need to be based on an explicit population of students' test scores so that differences in results can be explained.

Toward this end, the Smarter Balanced Consortium must adopt procedures that describe how the Consortium will:

- Specify the number of students who participated in the test,
- Calculate the number of students who completed each assessment, and
- Calculate the percent of students who achieved each level of proficiency on the test.

Consortium members have established complex accountability systems, databases, web-applications and reports that use assessment data to calculate participation rates, achievement rates and to make accountability determinations based on these calculations. These systems have evolved over time through agreements with stakeholders and to comply with applicable state and federal laws. Further, members document their approach to calculating participation and achievement rates as part of their Accountability Workbooks and Flex Waivers. For these reasons, changing the manner in which members calculate participation and achievement rates will take time.

**Therefore, states may elect to implement the Consortium rules regarding calculating participation and achievement rates but are not required to do so.**

#### Policy Objectives

The procedures regarding how to calculate participation and achievement rates must be considered in a policy context and must be aligned to the Consortium's goals.

#### Make Valid Claims about what Groups of Students Can Do

The validity of conclusions one can derive about assessment results is partly dependent on the percentage of students who participated in the assessment. For example, one cannot make generalizations about a school's performance if a large number of eligible students did not participate in the test.

### Make Valid Claims about what Individual Students Know and Can Do

The Consortium intends to make claims about what students know and are able to do regarding the Common Core standards. The validity and reliability of these claims are dependent on the number of questions that a student answers. Although students' knowledge and skills in a content area tend to be highly correlated, depending on the correlation to infer knowledge and skills is a risky proposition. For example, if a student were to only answer questions that measure their ability to read, we should limit the degree to which reading data are used to describe the student's achievement in English language arts/literacy given that the student did not demonstrate skills in writing, listening, or research.

### Promote Equity

Assessment can be a powerful tool to help identify gaps in student achievement. Creating incentives for all eligible students to participate helps to make data available that can describe achievement gaps and whether or not we can observe progress in closing the gaps.

### Honor Students' Efforts

The Common Core State Standards require students to be able to demonstrate many challenging skills. In some cases, students will complete most of the test, but for reasons outside of an individual student's control (e.g., illness, or scheduling problems) they are not able to complete an entire assessment. The Consortium seeks to honor students' work to the greatest extent possible while at the same time not compromising the validity of conclusions that stakeholders make based on the results of partially completed assessments.

### Incentivize Desirable Outcomes

The Consortium should create policies that members can use as incentives for adults to encourage students to complete the assessment. These incentives are particularly important as members consider how assessment data can help promote equity in education.

### Procedures

#### Attempted Tests

Attempted tests for the online administration are tests for which a student logs into the computer adaptive test (CAT) session and the performance task (PT) session in a content area. Attempted tests for the paper and pencil administration are tests for which a student provides an answer to one question in the test booklet for a content area.

Once a test is invalidated, the Consortium will not count a test as an attempt.

**Partial Tests**

Partial tests are attempted tests for which a student omitted a response to one or more items.

**Complete Tests**

Complete tests are attempted tests for which students provide an answer to every question on the CAT and PT components.

Note: Those student responses to items that cannot be scored because the student did not provide sufficient information will be treated as an incorrect response and will also contribute to a complete test.

**Valid Tests**

Valid tests are attempted tests that are administered in a manner consistent with the test administration manual. This includes but is not limited to:

- Using only allowable tools and accessibility resources;
- Adhering to test security requirements including appropriate proctoring; and
- Adhering to standardized test administration policies and procedures (e.g., reading test directions, maintaining a suitable testing environment).

## Calculating Participants, Assigning Achievement Levels and Calculating Scores

For purposes of reporting, Consortium level information and conducting Consortium wide research, the following procedures will be used:

**Participants**

The Consortium will count students with attempted English language arts/literacy tests that are valid as participants in English language arts.

The Consortium will count students with attempted mathematics tests that are valid as participants in mathematics.

**Assigning Achievement Levels**

The Consortium will assign achievement levels based on the achievement levels adopted by Consortium members and the scale scores calculated for the tests.

## Calculating Scale Scores

For Complete Tests, the Consortium will score tests based on the students' responses to the items in the CAT and PTs.

For Partial Tests for which students answer at least 10 CAT items, the Consortium will derive a score based on the students' actual responses to items and will consider as incorrect any item

to which students do not provide a response. The difficulty of a student's omitted items from the combination of online CAT and PTs will be estimated based on the average difficulty of the items in the item pool. For fixed form tests, the difficulty of omitted items will be based on the difficulty of the actual omitted items.

For Partial Tests for which students answer 0-9 CAT items, the Consortium will assign the Lowest Obtainable Scale Score (LOSS) for the grade and content area.

A student who answers every question incorrectly will receive the lowest scale score for the grade and content area. A student who answers every question correctly will receive the Highest Obtainable Scale Score (HOSS) for the grade and content area.

### Assigning Achievement Levels

The Consortium will assign Achievement Levels based on the scale scores derived for the assessment.

### Member Responsibility

Members should document their method of scoring tests including partial tests, calculating participation, and performance. Members should attempt to align their process with the adopted processes as described above. This information is also available in the Smarter Balanced Participation and Performance Calculations Policy posted in the [Document Library](#) of the Member Portal.

### Recommendation Regarding Documentation of Evidence

In their evidence documentation, members should plan to include their method of:

1. Calculating Participation
2. Calculating Performance
3. Scoring tests including partial tests

## 7.1.3 Private School/Home-Instructed Student Participation

The inclusion of private schools/home-instructed students in the administration of the Smarter Balanced assessments is subject to state, territory, or agency policy, legislation, and regulations.

### Member Responsibility

Members should develop a policy related to the participation of private schools and home-instructed students.

**RECOMMENDATION REGARDING DOCUMENTATION OF EVIDENCE**

In their evidence documentation, members should plan to include their policies for private school and home-instructed student participation.

Smarter Balanced documents members' private school and home-instructed student participation policies in the Member Test Administration Policy document available on the K-12 Leads page on the Member Portal.

### 7.1.4 Testing in Retained Grades

Students who repeat a grade may be required to participate in the Smarter Balanced assessments in the repeated year, depending upon member policy. Member policy may indicate that students who repeat a grade do not have to take a test in their repeated year in the content area for which they achieved a Level 3 or higher in the previous year. Members may roll the results of those tests forward.

Members may add language around grade retention to the Online Summative TAM.

## 7.2 UNIVERSAL TOOLS, DESIGNATED SUPPORTS, AND ACCOMMODATIONS

### 7.2.1 Universal Tools, Designated Supports, and Accommodations Available to Students

Smarter Balanced assessments are designed to use only the available universal tools, designated supports, and accommodations identified in the Smarter Balanced Usability, Accessibility, and Accommodations Guidelines found via the Accessibility webpage and in the Document Library on the Member Portal. Table 2 describes the universal tools, designated supports, and accommodations available to students during testing.

Table 2: Definitions for Universal Tools, Designated Supports, and Accommodations

| Type | Definition |
|------|-----------|
| Universal Tools | Access features of the assessments that are either provided as digitally delivered components of the test administration system or separate from it. Universal tools are available to all students based on student preference and selection. |
| Designated Supports | Access features of the assessments available for use by any student for whom the need has been indicated by an educator (or team of educators working with the parent/guardian and student). Designated supports are either provided as digitally delivered components of the test administration system or separate from it. |
| Accommodations | Accommodations are changes in procedures or materials that increase equitable access during the Smarter Balanced assessments. Assessment accommodations are not modifications. They generate valid assessment results for students who need them, and they allow these students to show what they know and can do. Embedded and non-embedded accommodations are available for students with documented IEPs or Section 504 Plans. Consortium-approved accommodations do not compromise the learning expectations, construct, grade-level standards, or intended outcome of the assessments. |

A member may elect not to make available to its students any universal tool, designated support, or accommodation that is included in the *UAAG* when the implementation or use of the universal tool, designated support, or accommodation conflicts with the member's state, territory, or agency law, regulation, or policy. Refer to *section 7.2.2 Requests for New Universal Tools, Designated Supports, and Accommodations*, in the event specific student situations require deviation from the UAAG.

**RECOMMENDATION REGARDING DOCUMENTATION OF EVIDENCE**

In their evidence documentation, members should plan to include evidence of universal tools, designated supports, or accommodations permitted by the UAAG that were not made available to students.

Smarter Balanced documents universal tools, designated supports, or accommodations that members elect to not make available to students in the Member Test Administration Policy document available on the K-12 Leads page of the Member Portal.

## 7.2.2 Requests for New Universal Tools, Designated Supports, and Accommodations

Proposed changes to the list of universal tools, designated supports, and accommodations will be brought to Governing Members for review, input, and vote for approval. Furthermore, members may issue temporary approvals (i.e., one summative assessment administration) for individual unique student designated supports or accommodations.

To capture member suggestions and/or school/district requests for improving the Smarter Balanced UAAG, Smarter Balanced has developed a feedback mechanism by which members can request changes to the UAAG each year.

K-12 Leads will evaluate formal requests for temporary, unique student accommodations or designated supports and determine if the request poses a threat to the measurement of the construct. If a member approves a request, they will submit the accommodation or designated support to the Consortium for review to include in the UAAG:

1. The member sends temporary approvals during the UAAG updates process. The Consortium will consider all member-approved temporary approvals for accommodations or designated supports as part of the annual Consortium UAAG review process.

2. If the Consortium determines that additional time to study the issue before the Consortium can engage in a vote is required, a member may notify the Consortium that the member intends to issue temporary approvals for the same accommodation or designated support during the next summative assessment administration. Members should include in their notification to the Consortium the rationale for issuing temporary authorizations for the next summative assessment administration.

The Consortium will provide to members a list of the temporary accommodations/designated supports issued by members that are not Consortium approved accommodations/designated supports and cannot be authorized for the next summative assessment administration. When members review and consider temporary unique designated supports and accommodations, they should engage in a systematic process to examine the accessibility resources relevance, the information available to inform use of the accessibility resource, and the impact of the use of the requested accessibility resource:

- Relevance: Members should evaluate whether the request for the unique accessibility tool serves to support a student's needs in a manner that is not met by other accessibility tools currently offered as a Universal Tool, Designated Support, or Accommodation.

    o Has the student used the accessibility resource regularly during instruction?

    o Has the student used the accessibility resource as part of an assessment of the particular construct and have educators interpreted the results from the testing to be comparable to the performance of the student's peers?

- Information: Members should seek insight from experts and research on whether the requested designated support or accommodation is recommended for use in large-scale assessment. Smarter Balanced provides resources to support this process:

  - Members should consult the Accommodations for ELLs and SWDs: A Research-Based Decision Algorithm written for Smarter Balanced by Abedi and Ewers, which includes research and a decision-making algorithm for validity and effectiveness.

  - Members should also examine the Smarter Balanced *Usability, Accessibility, and Accommodations Guidelines (UAAG)*, which includes all of the Consortium's available accessibility resources and information about resources members have agreed do not yield comparable results.

- Impact: Members should consider the risk of allowing an accessibility resource in terms of validity, feasibility, and potentially, test security:

  - Validity: Would the accessibility resource alter the intended construct measured? If the accessibility resource alters the intended construct, allowing the accessibility resource would provide results to the students, parents and teachers that are not consistent with the achievement level descriptors. In addition, the Consortium members may elect not to make that accessibility resource available in subsequent years.

  - Feasibility: Are there logistical issues related to standardization or validity when implementing the requested accessibility resource?

  - Test Security: Does allowing the use of the requested accessibility resource result in any concerns related to test security?

### 7.2.3 Translated Resources

Smarter Balanced provides translated glossaries as a math designated support. Every five years, the Consortium surveys the membership to understand the most frequently used languages across membership. Smarter Balanced staff review the results with members to determine if any new languages should be added to the current suite of languages that support embedded translated glossaries in mathematics:

- Arabic (Modern Standard & Egyptian)
- Burmese
- Cantonese (traditional & simplified)
- Filipino (Ilokano & Tagalog)
- Hmong (White & Green)
- Korean
- Mandarin (traditional & simplified)

- Punjabi (East & West)
- Russian
- Somali
- Spanish (El Salvadorian, Mexican, & Puerto Rican)
- Ukrainian
- Vietnamese (North & South)

The most recent survey was administered in 2023, and results are being processed.

**Resources**

- [Usability, Accessibility, and Accommodations Guidelines (UAAG)](#)
- [Accommodations for ELLs and SWDs: A Research-Based Decision Algorithm](#)
- [Meeting the Needs of Diverse Learners Smarter Balanced website](#)

# 7.3 TEST ADMINISTRATION WINDOW

Each member will establish a schedule for the administration of the Smarter Balanced summative assessments consistent with the Smarter Balanced standard testing window adopted in September 2013, and the subsequently adopted policy on member flexibility to establish testing windows.

**Grades 3-8**

- Testing shall not begin until at least sixty-six percent (66%) of a school's annual instructional days have been completed, and
- Testing may continue up to and including the last day of school.

**High School**

- Testing shall not begin until at least eighty percent (80%) of a school's annual instructional days have been completed, and
- Testing may continue up to and including the last day of school.

A member may establish more specific windows within the constraints of the Consortium-defined windows described above. A member may alter their test window in keeping within Smarter Balanced standard policies.

*[Note: Sixty-six percent of a school year occurs after the 120th instructional day in a 180-day year, leaving a twelve-week window for grade 3–8 testing; eighty percent of a school year occurs after the 144th instructional day in a 180-day year, leaving a seven-week window for grade 11 testing.]*

**Rationale for the Standardized Window**

- When tests are administered in a standardized window, Smarter Balanced members are better able to argue that student performance and growth of student performance on the tests is comparable from year-to-year.

- With respect to high school testing, the standardized window allows higher education and other post-secondary options that students may choose to more consistently interpret the meaning of the test scores.

- As a generally accepted best practice, end-of-year summative assessments should only be administered after a student has had an opportunity to learn the assessed content standards.

- For most students, consistently administering a summative assessment close to the end of a student's school year will likely provide a better gauge of growth and how much students have progressed from year to year. However, for students receiving accelerated instruction, it may be appropriate to test them earlier in the year after they have received instruction at the full depth and breadth of the assessed content standards.

**Member Flexibility to Establishing Testing Windows**

- Members may elect to shorten testing windows.

- Members may choose to establish year-to-year timeframes for specific grade levels, making estimates of year-to-year growth more stable across districts.

- Members may also elect to provide more flexibility for administration outside the Consortium designated window as long as the change:

    o Does not threaten the security of the Consortium's items or tasks and does not otherwise deviate from the Consortium's test administration manual (e.g., the maximum of a 12-week window; the maximum duration a test may be open, Universal Tools, Designated Supports, and Accommodations).

    o Does not require additional resources from the Consortium (e.g., changing consortium timelines, item development plans, and field testing plans)

- The member establishes its rationale and is prepared to support its approach based on its intended uses of the test.

### 7.3.1 Additional High School Flexibility

**Expanded Testing Windows**

Consortium members may elect to provide flexibility in establishing high school testing schedules that better align to their academic calendar, their students' course enrollment, or as

necessary to address conflicting testing requirements or technology resource constraints. The flexibility may include extending the Grade 11 window to begin after 66% of instruction is completed by course or instructional calendar, as appropriate.

## Assessing Grades Other than Grade 11

States may elect to administer the High School Assessment in grades other than Grade 11.

- Members may administer the high school test in grades 9, 10 or 12. However, members will likely need to collect additional validity evidence in support of decisions made about the knowledge and skills students have mastered for the purpose for which the data will be used.

## Reporting High School Results

The items administered as part of the high school test are aligned to the high school standards. Consortium members established achievement levels based on teachers' expert judgment regarding the level of performance that all students should have by the time they reach the end of their 11th grade. When reporting high school results, members should clearly document if they are administering and scoring Smarter Balanced items in a unique manner, including:

- Grade 9, 10, or 12 as the grade of accountability
- Administering items using a different blueprint and/or a different achievement standard

Consortium members should (1) confer with service providers on whether it is feasible/practicable to increase flexibility in the testing window, and (2) confer with higher education officials on whether flexibility is consistent with the use of Smarter Balanced scores in higher education.

## Recommendation Regarding Documentation of Evidence

In their evidence documentation, members should plan to include their annual summative test administration windows.

Smarter Balanced documents members' annual summative test administration windows in the Member Test Administration Policy document available on the K-12 Leads page of the Member Portal.

# 7.4 REQUIRED TRAININGS

**Member Responsibility**

Members are encouraged to determine what trainings/resources and confidentiality agreements are required by applicable state or local law prior to administration of the online summative assessments. Members may also choose to offer a "certification" to those working with the Smarter Balanced summative assessments.

**Monitoring**

Procedures developed by members related to required training should be monitored regularly and adjusted as needed.

**Recommendation Regarding Documentation of Evidence**

In their evidence documentation, members should plan to include their policies for test administration trainings/certifications.

## 7.4.1 Training Modules

Smarter Balanced developed nine training modules (available in the Assessment Training and Operations folder of the Secure File Transfer Server) that address various topics and support processes that those involved in administration will need to know.

These modules are developed as PowerPoint presentations with scripts for developing narrated versions and allow for member customization. Smarter Balanced will update the training modules as needed.

**Member Responsibility**

Members should plan to customize and incorporate member-specific web links, references to paper-and-pencil tests, policies, user roles, member-specific nomenclature, and system requirements throughout these modules. Members' service providers may develop narrated versions of customizable training modules. The training modules should be made available to the appropriate audiences on the member's website or assessment portal. A description and target audience of each module can be found in section 2.1 of the *Online Summative TAM*.

**Member Monitoring**

Members should review these training modules annually and update them based on new or revised information.

Templates for these training modules are available in the Assessment Training and Operations folder of the Secure File Transfer Server.

## 7.5 IMPROPRIETIES, IRREGULARITIES, AND BREACHES—SUMMATIVE ASSESSMENT

Test security incidents such as improprieties, irregularities, and breaches are behaviors prohibited during test administration (online and paper-and-pencil), either because they may change the meaning of the assessment, give a student an unfair advantage, or because they may compromise the secure administration of the assessment. Whether intentional or by accident, failure to comply with security rules, either by staff or students, constitutes a test security incident.

Improprieties, irregularities, and breaches must be reported in accordance with the instructions in this section for each severity level. For a flow chart of these steps, see section *4.0 Responding to Testing Improprieties, Irregularities, and Breaches* of the *Online Summative TAM*. These steps may be modified according to member preference. For more information about reporting incidents online using the appeals system, see section 4.0 and 5.0 of the *Online Summative TAM* available on the Smarter Balanced Member Portal and the Secure File Transfer Server. Refer to the *Paper-and-Pencil (P&P) TAM* (available on the Smarter Balanced Member Portal and the Secure File Transfer Server for steps on reporting incidents experienced during a paper-and-pencil administration.

| Type | Definition |
|---|---|
| Impropriety | An event that does not align to standardized test administration policy or procedures that has a low impact on the individual or group of students who are testing and has a low risk of potentially affecting student performance on the test, test security, or test validity. These circumstances can be corrected and contained at the local level. An impropriety should be reported to the School Test Coordinator (SC) and District Test Coordinator (DC) immediately and documented in a central location as necessary (should an appeal be required). |
| Irregularity | An event that does not align to standardized test administration policy or procedures that impacts an individual or group of students who are testing and may potentially affect student performance on the test, test security, or test validity. These circumstances can be corrected and contained at the local level and submitted to the member for review |

| Type | Definition |
| --- | --- |
| | and dispensation. An irregularity must be reported to the DC and SC immediately and documented in a central location within 24 hours of the incident (should an appeal be required). An irregularity does not automatically require invalidation. Sometimes an irregularity does not actually change the meaning of a test score. For example, if classroom instructional materials, that do not address the content of the test, remain exposed in a secure testing environment, this would be an example of an irregularity that should be addressed for future testing, but may not change the inferences that can be made about students' scores who tested in that environment. Each unique circumstance of an irregularity must be considered based on the report that addresses the adult and student interactions that differ from the standardized assessment conditions described in the test administration manual. |
| Breach | An event that does not align to standardized test administration policy or procedures that drastically increases the threat to the validity of the test and the test results of individual students who may be exposed to the material included in the breach. Examples may include such situations as a release of secure materials or a security/system risk. These circumstances have external implications for the Consortium and may result in a Consortium decision to remove the test item(s) from the available secure bank. A breach incident must be reported to the SC and DC immediately and documented in a central location (should an appeal be required). A breach does not automatically require invalidation. Sometimes a breach may be addressed quickly enough as to limit the threat to validity. Sometimes an item can be removed from students' test scores to address a breach. Each unique circumstance of a breach must be considered in the context of the breach risk form as described in Appendix F: Item Risk Rubric. |

Members may choose to customize the definitions above to make them *more* restrictive. The definitions and timelines above are the minimum expectations of Smarter Balanced.

## Member Responsibility

Members have the responsibility to:

1. Establish clear policies about monitoring during test administration.

2. Monitor (including social media) and investigate any test security incidents throughout the test administration cycle and determine what, if any, action needs to be taken.

3. Report all breaches to Smarter Balanced immediately (as outlined in Section 4.0 of the *Online TAM*).

4. Review and revise monitoring procedures and/or implement corrective action as needed to avoid these situations in the future.

5. Develop a mechanism for tracking security incidents reported by schools/districts.

6. Develop a process for monitoring security incidents in social media.

As described in section *4.0 of the Online Summative TAM, Responding to Testing Improprieties, Irregularities, and Breaches*, breaches must be reported to Smarter Balanced immediately and irregularities must be reported by the end of the day of the incident to the member via the test security incident log or other mechanism for tracking security incidents. Members should be prepared to provide the following information in the event of a breach:

- Date of incident

- Member/district/school

- Grade/content area/PT or CAT

- Number of students involved in incident

- Description of incident

- District and/or school action taken

- Appeal requested

- Member action taken

## Resources

A customizable Test Security Incident Log template is available in the Assessment Training and Operations/Manuals and Guides folder on the Secure File Transfer Server. Examples of what are considered breaches, improprieties, and irregularities can be found in *Appendix C: Test Security Chart*.

## Recommendation Regarding Documentation of Evidence

In their evidence documentation, members should plan to include their test security policies and procedures.

Members are required to log all breaches and report them to Smarter Balanced, including a description of how the incident was handled. Information about reporting breaches to Smarter Balanced is available in *Appendix D: Guidance for Social Media Summative Assessment Monitoring*. Members should keep logs of all testing irregularities for internal use.

Members should develop a process to ensure the review and investigation of all test security incidents that are logged to determine what, if any, action may need to be taken.

### 7.5.1 Removing an Item from the Item Pool

In certain instances, test security breaches may necessitate items be removed from the item pool. Smarter Balanced will monitor these instances; however, members and their service providers should have procedures in place to inform Smarter Balanced immediately when it appears this type of breach has occurred. Smarter Balanced will investigate and quickly determine if any action is required (see *Appendix E: Item Risk Rubric*). Once a decision has been made, Smarter Balanced will contact all K-12 Leads and service provider program managers to communicate action steps. If Smarter Balanced determines that an action is required, Smarter Balanced will notify all members' service providers, and K-12 Leads of the item ID that must be removed from the pool. Service providers will need to remove the item within 1 business day.

Unless otherwise determined by Smarter Balanced on a case-by-case basis, members do not need to re-score tests for students who respond to an item exposed between the time the item was exposed and when the item was removed from the service providers systems. For breached items included in paper forms, Smarter Balanced and member service providers will develop special scoring tables.

### 7.5.2 Test Reset

Smarter Balanced does not have an established policy on retake opportunities. Under most circumstances, a test reset as described in *Appendix B: Online System – Appeal Types and Conditions for Use* will serve as the mechanism by which students have the opportunity to complete a test in cases in which the original test administration conditions were not completed in a manner consistent with the Online Summative TAM and would result in a score that would not allow for making valid inferences about what a student knows and can do. Members should consult with their service provider if there is an issue that cannot be addressed with a test reset.

# 7.6 PERFORMANCE TASKS

### 7.6.1 Overview

Performance tasks (PTs) measure a student's ability to integrate knowledge and skills across multiple standards — a key component of college-and-career readiness. Performance tasks will be used to better measure capacities such as depth of understanding, research skills, and complex analysis, which cannot be adequately assessed with stand-alone selected or constructed-response items. Many of the items contained in the PTs require students to provide an extended written response that will be hand scored.

The PT is not adaptive and is administered separately from the computer adaptive test (CAT) portion of the Smarter Balanced assessment.

### Paper and Pencil forms

The items associated with single PTs will be administered on the paper and pencil form.

### Accommodations

Refer to the *Smarter Balanced Usability, Accessibility, and Accommodations Guidelines* for information about universal tools, designated supports, and accommodations available to students during testing. With few exceptions, the universal tools, designated supports, and accommodations available during the CAT portion of the assessment are also available during the PT.

### 7.6.2 Performance Tasks Administration

### Assigning Performance Tasks

Performance tasks (PTs) should be randomly distributed by student and grade.

Performance task assignment information must be entered into the member's test registration system.

Refer to *Section 14.2* of the *Member Procedures Manual* for information on assigning PTs when administering the PT embedded field test.

### Performance Task Administration

Performance tasks are individually administered and computer-generated. The member will administer PTs via the same test delivery system as the computer adaptive test (CAT) portion of the Smarter Balanced assessment. Students will enter their responses to PT items on the same system used during the CAT portion of the Smarter Balanced assessment.

Smarter Balanced recommends that the ELA/literacy PT be completed over two sessions and the mathematics PT be completed during a single session.

Each student will complete a PT in each content area in which they are participating in the Smarter Balanced assessment. A student's content area test will not be complete and valid without participation in both the PT and CAT portions of the assessment.

To minimize the effect of student fatigue, Smarter Balanced recommends that students take the PT and the CAT portions of the assessment on separate days. Smarter Balanced also recommends that students begin with the CAT portion of the assessment, followed by the PT. However, the PT and CAT portions of the assessment may be administered in a different order, if necessary.

### Duration of the Performance task

Once a student has started a PT, it will be available for up to 10 calendar days (not school days). It is recommended that students complete each part of a PT within a single day. Members may elect to authorize districts to allow a PT to remain open for an additional 10 calendar days. Test delivery systems could be configured to automatically expire PTs after 20 calendar days unless the member has requested from their vendor for the PTs to expire after 10 days.

### Scheduling the Performance Task

Information on scheduling the PT for English language arts/literacy is provided in Table 4.

Table 4: Scheduling Information for the PT, English Language Arts/Literacy

| English Language Arts | Performance Task (PT) |
|---|---|
| Number and Duration of Sessions | The PT is presented in two parts.<br>Recommendations:<br>• Administer in two sessions corresponding to Parts 1 and 2 of the PT.<br><br>Session durations range from 60-120 minutes. Part 2 should be scheduled to allow students enough time to complete the full write in one test session. |
| Breaks within Sessions | The PT is presented in two parts. Students can take breaks within Parts 1 and 2; however, once a student moves to Part 2, he or she will not be able to review or revise items in Part 1. |

| English Language Arts | Performance Task (PT) |
|---|---|
| | Recommendation: Students complete Part 1 in one test session and Part 2 the next school day in a single test session. |
| Total Duration | The PT will remain active for 10 calendar days after the student has begun the PT (unless member delegates authority to extend time). <br><br> Recommendation: Student completes each part of the PT within one day. |
| Additional Required Resources | Headphones are required for some PTs. |

Information on scheduling the PT for Mathematics is provided in Table 5

Table 5: Scheduling Information for the PT, Mathematics

| Mathematics | Performance Task (PT) |
|---|---|
| Number and Duration of Sessions | Recommendations: <br> • Administered in one session. <br> • Session duration ranges from 40-120 minutes. |
| Breaks within Sessions | Students can take breaks during PT test sessions. Mathematics PT items are presented on a single screen. Following a break, the student will have access to the same items. |
| Total Duration | The PT will remain active for 10 calendar days after the student has begun the PT (unless member delegates authority to extend time). <br> • Recommendation: Student completes the PT in one day. |
| Additional Required Resources | Plain graph paper is required beginning in sixth grade and can be used on all mathematics assessments. |

**Scoring Performance Tasks**

The PTs in English language arts/literacy and mathematics will include a variety of item types. Many items on the PTs will require hand scoring. Some items on mathematics PTs require use of a scoring engine capable of scoring graphic and equation items. Refer to *Section 7.7 Item Scoring* for more information on the scoring requirements for items on the PTs.

**Resources**

- *Performance Task Overview Training Module* (Available in the Smarter Balanced Member Portal and the Secure File Transfer Server
- *Performance Tasks Specifications*

**Member Responsibility**

Members are responsible for working with their service providers to implement the design for assignment of the Smarter Balanced-provided PTs via online test delivery and paper/pencil form (if administered). For each grade and content area, members will randomly assign PTs to students. Random assignments ensure that performance tasks are not over-assigned within the state/territory.

## 7.7 ITEM SCORING

Item scoring comprises both machine and external scoring, including hand scoring. States will need to contract with service providers to obtain hand scoring and possibly automated scoring with artificial intelligence (AI) for items that cannot be machine scored. Smarter Balanced does not require the use of automated scoring. The computer adaptive test (CAT) component for the ELA/literacy and mathematics assessments includes a variety of item types that are machine scored by the service provider's test delivery system in real time. In addition, for ELA/literacy in grades 6-8 and high school, the CAT component also includes short text items that may require hand scoring. The Performance Task (PT) component of each content area assessment requires both machine-scoring and hand scoring.

Table 6 shows the item types and typical scoring method for the summative assessment item pool by content area.

Table 6: Item types by Subject Found in Summative Assessment Item Pools

| Item Types | ELA/literacy | Mathematics | Typical Scoring Method |
|---|---|---|---|
| Multiple Choice (MC) | Y | Y | Machine |
| Multiple Select (MS) | Y | Y | Machine |
| Evidence-Based Selected Response (EBSR) | Y | | Machine |
| Match Interaction (MI) | Y | Y | Machine[*] |
| Hot Text (HTQ) | Y | | Machine |
| Short Answer Text Response (SA) | Y | Y | Hand-scored |
| Essay/Writing Extended Response (WER) | Y | | Hand-scored |
| Equation Response (EQ) | | Y | Machine[*] |
| Grid Item Response (GI) | | Y | Machine[*] |
| Table Interaction (TI) | | Y | Machine[*] |

[*] The summative test packages include MI, EQ, GI, and TI items that may require external scoring.

### 7.7.1 Machine-Scored Items

Most item types in the Smarter Balanced online assessment will allow for machine scoring in both ELA/literacy and mathematics.

Smarter Balanced member service providers' test delivery systems can machine score both graphic and equation items for the online assessment. The online test package indicates which items can be machine scored and provides machine scoring rubrics for those items. The text-response items require external scoring, either hand scoring or a combination of hand scoring and scoring with an automated scoring application (artificial intelligence (AI) scoring application). For paper-pencil administration, Smarter Balanced provides files containing item information (including the score keys) for the paper forms. Members will work with their service providers to load these keys into the service providers' scanning/scoring systems to allow for machine scoring of some of the items on the paper forms.

### 7.7.2 External Scoring, Including Hand Scoring

There are ELA/literacy CAT and PT items and mathematics PT items which require external scoring. For example, text response items require external scoring. In addition, some equation

and graphic items may require external scoring. External scoring may be accomplished by hand scoring or by an automated scoring application (AI scoring).

Table 7 shows the number of items in the 2023-24 summative assessment item pool that require scoring by human scorers or an automated scoring application. The entire pool of hand-scored items will be available for each administration. Multiple Performance Tasks will be randomly administered across each state/territory. It is unknown how many students will take each item within a state/territory.

There are no hand-scored items on the CAT portion of the mathematics assessment. Each student will be administered between zero and four items requiring hand scoring on the performance task portion of the mathematics assessment, depending on grade. Each student will be administered between zero and five short text items and one essay item requiring hand scoring on the ELA/literacy assessment, depending upon grade.

Table 7: Total Number of Items in the 2023-24 Summative Assessment Item Pool that Require Hand Scoring

| Content Area | Component | Total Number of Hand-Scored Items Across All Grades |
|---|---|---|
| ELA/literacy | CAT | 373 |
| ELA/literacy | PT | 308 |
| Mathematics | CAT | 0 |
| Mathematics | PT | 381 |

**Hand-Scoring Process**

1. For operational items that require hand scoring, Smarter Balanced provides the following materials for each year's summative test package. The materials are available to member states and service providers in the Scoring Materials folder of the Smarter Balanced Secure File Transfer Server (FX).

   - Training materials

   - Qualification sets

   - Scoring rubrics

   - Training papers

   - Validation papers

   - Specifications for condition codes and alerts

2. The state and its service provider are responsible for training and scoring for operational items that require hand scoring. Smarter Balanced has no consortium-wide requirements for number and location of scoring sites or hiring requirements for scorers. Each state will determine:

   - Any constraints around hand scoring, including preference for on-site, distributed (remote) and/or automated scoring.
   - Whether service providers will use existing hand scoring sites and raters.
   - If automated scoring with an artificial intelligence (AI) scoring application is used, whether or not the state allows the score derived from the automated scoring system to be the score of record.
   - Whether it will allow scoring activities (including hand scoring) to take place outside of the United States.
   - If any scoring will be done in-state/territory by teachers (e.g., professional development opportunity); states should be explicit regarding the distribution of student data to locations, people, or machines inside or outside of the United States.
   - Minimum qualifications for raters; states should be mindful of the importance of including teachers with backgrounds in teaching diverse student populations (e.g., low socio-economic status, English language learners, and students with disabilities) from different geographic locations.
   - How much and how often the state requires monitoring of statistical feedback from the service provider's hand scoring process.

3. The service provider will deliver field test responses for scoring by another party, managed by Smarter Balanced. The transfer of student responses may pass through the state to the Smarter Balanced field test scoring contractor.

4. Smarter Balanced has established minimum qualification standards for qualifying and monitoring external scoring. States are expected to maintain traditional item-level hand-scoring statistics, such as rater agreement rates and inter-rater reliability, and to meet these standards for consistency, but may also establish stricter standards for external scoring.

5. Smarter Balanced does not have any consortium-wide requirements for the rate of double-scoring (percentage of responses for an item that are scored twice). Smarter Balanced recommends that 10% of student responses be the minimum level of double scoring allowed. Each state will determine the percentage of its student responses that will be double scored.

6. Smarter Balanced recommends following the guidelines described in the _Operational Best Practices for Statewide Large-Scale Assessment Programs_.

7. Scoring Spanish Responses for Mathematics—Students who use dual-language supports in the classroom and for whom Dual-Language Translations is a provided support, may respond to items in English, Spanish, or a combination of the two for the mathematics assessments. Allowing this option is a state-level decision. States allowing this option need to be sure their service providers are prepared for flagging and hand scoring mathematics items with Spanish responses.

8. States will work with their service provider to establish procedures to be followed when sensitive responses are received for items that require hand scoring.

## Training

**ELA/literacy.** To score ELA/literacy items, raters should receive training at the level of the task model. Smarter Balanced provides "Baseline" anchor and training sets as well as rubrics by writing purpose (e.g., informational, opinion) for essay items. Qualification and validation sets are provided for each essay. Anchor and training sets are also provided for the task models associated with the ELA/literacy short text items in the CAT and PT sections. For the ELA/literacy short text items in the CAT and the PT sections, raters should receive training by grade span (grades 3-5, 6-8, and high school) instead of by grade level.

Even though training is at the task-model level, qualification occurs on an item-by-item basis for all ELA/literacy hand-scored item types, and Smarter Balanced provides qualification and validation sets for each item in ELA/literacy. For those item types where raters were trained by grade span, raters will qualify on each item within a specific grade.

**Mathematics.** For mathematics, Smarter Balanced provides anchors and training sets for the task models. Smarter Balanced provides item-specific rubrics and item-specific validation sets for all mathematics items. To score mathematics items, raters should receive training and qualify on task models for almost all items.

## Special Cases

**Condition Codes.** Along with other training materials for scoring of operational responses, Smarter Balanced provides specifications for condition codes and alerts. Smarter Balanced has defined special cases of student responses to hand-scored items. The following condition codes are used during hand scoring:

> B: Blank
>
> I: Insufficient
>
> L: Non-scorable language
>
> T: Off topic (essay only)
>
> M: Off purpose (essay only)

In addition to condition codes, service providers will train raters for sensitive papers/alerts, including watching for troubled students and/or cheating. Examples of troubled student alerts include (but are not limited to) suicide, criminal activity, bullying incidents, and/or extreme depression. An indication of cheating may include (but is not limited to) evidence of assistance/intervention by an adult during the testing process. States and their service providers should have a process in place to handle alerts as soon as they arise, including notifying both the state/territory where the alert originated and Smarter Balanced.

**Note:** Starting in school year 2022-23, student responses to the ELA/literacy full write items will receive a condition code of off purpose will still be scored for the dimension, Conventions. See the updated scoring rules, available here.

### Double Scoring

Each state shall determine the percentage of its student responses that will be double scored, and the method for double scoring (e.g., items can be double scored using a combination of automated scoring and hand scoring). Percentages for double scoring will vary by content area and possibly by grade(s). Smarter Balanced recommends that 10% of student responses be the minimum level of double scoring allowed. Smarter Balanced expects states to collect item-level hand-scoring statistics (e.g., inter-rater reliability) for each hand-scored item for all item responses that receive two reads.

Service providers may score Smarter Balanced student responses from multiple states; thus, states may want to request item-level, hand-scoring statistics aggregated across all states scored by the service provider as well as item-level, hand-scoring statistics for an individual state.

Smarter Balanced has established minimum qualification standards for qualifying and monitoring external scoring.

### Quality Monitoring

Before hand-scoring any operational student responses, raters must meet the qualification standards described below:

Table 8: Exact Agreement Standards for Rater Qualification

| Score Point Range | Qualification Standard |
|---|---|
| 0-1 | 90% (no non-adjacent scores) |
| 0-2 | 80% (no non-adjacent scores) |
| 0-3 | 80% (no non-adjacent scores) |
| 0-4 | 70% (no non-adjacent) |

Smarter Balanced provides validity papers for all hand-scorable items. Service providers will use these papers to monitor scoring accuracy. Raters will be expected to maintain these quality standards (Table 8 above) during the qualification and validation process. During hand scoring, raters score a minimum of ten validity responses for each item. Scorers that do not meet the exact agreement standards should be retrained or disqualified from scoring.

Service providers are responsible for providing inter-rater reliability (IRR) and validity reports to states. IRR reports indicate agreement rates among readers and provide a picture of readers' scoring patterns. Validity reports indicate agreement rates between a reader and a set of papers with predetermined scores and provide a picture of readers' adherence to or drift from the correct scoring procedures. Analysis of the inter-rater reliability and validity reports is an excellent source to determine team or reader drift and team leader influence. The supervisor can re-anchor readers or teams with live papers, relevant training materials, and scoring guidelines. For monitoring inter-rater reliability, a minimum 10% second-read rate is required. The inter-rater reliability standards and validity standards are provided in this table.

During hand scoring, raters score a minimum of ten validity responses for each item. Scorers are expected to meet the following exact agreement standards or be retrained or disqualified from scoring.

Table 9: Exact Agreement Standards for Rater Validation

| Score Point Range | Qualification Standard |
|---|---|
| 0-1 | 90% |
| 0-2 | 80% |
| 0-3 | 80% |
| 0-4 | 70% |

## Inter-rater Reliability

Inter-rater reliability reports indicate agreement rates among readers and provide a picture of readers' scoring patterns. Analysis of the inter-rater reliability reports is an excellent source to determine team or reader drift and team leader influence. The supervisor can re-anchor readers or teams with live papers, relevant training materials, and scoring guidelines. For monitoring inter-rater reliability, a minimum 10% second-read rate is required. The inter-rater reliability standards are provided in this table.

Table 10: Inter-Rater Reliability (IRR) Standards

| Score Point Range | IRR Standard (Exact Agreement) |
|---|---|
| 0-1 | 80% |
| 0-2 | 70% |
| 0-3 | 70% |
| 0-4 | 60% |

## Member Responsibility

Members are expected to meet these standards but may also establish stricter standards for external scoring.

Each member shall determine the percentage of its student responses that will be double scored. Some items will be double scored using a combination of AI scoring and hand scoring. Percentages for double scoring will vary by content area and possibly by grade(s). Smarter Balanced recommends that 10% of student responses be the minimum level of double scoring allowed.

Smarter Balanced recommends following the guidelines described in the *Operational Best Practices for Statewide Large-Scale Assessment Programs*.

## Resources

Smarter Balanced provides available training materials, qualification sets, scoring rubrics, training papers, and validation papers for these items during each testing window; these are found in the Scoring Materials/Summative Scoring Materials folder of the Secure File Transfer Server.

*Scoring Spanish Responses for Mathematics*—Students who use dual language supports in the classroom and for whom Dual Language Translations is a provided support, may respond to items in English, Spanish, or a combination of the two for the mathematics assessments. Allowing this option will be a member-level decision. Members allowing this option need to be sure their service providers are prepared for hand scoring mathematics items with Spanish responses.

Additional scoring resources:

- *Smarter Balanced Hand Scoring Rules* located in the Scoring Materials Folder on the Secure File Transfer Server.
- *Appendix F: Large-Scale Scoring: Involving Teachers and Producing Results.*
- Technical Report on Automated Scoring with Artificial Intelligence.

**Recommendation Regarding Documentation of Evidence**

In their evidence documentation, members should plan to include their external scoring policies and procedures. This evidence includes demonstrating that:

1. The training materials, qualification sets, scoring rubrics, training papers, and validation papers provided by the Consortium were used as expected in all hand scoring activities following test administration.

2. External scoring met the minimum standards described above.

3. There is an adjudication process for resolving read-behind discrepancies that is consistent with the process used for the member's other assessment programs.

### 7.7.3 Scoring of Sensitive Papers

Members need to be sure that the procedures put in place by their scoring service providers allow for a process to be deployed when sensitive responses are received on constructed response items. Examples of what Smarter Balanced has identified to be potentially sensitive in nature include, but are not limited to:

- Suicide
- Criminal activity
- Alcohol or drug use
- Extreme depression
- Extreme violence
- Sexual assault or physical abuse
- Self-harm or intent to harm others
- Neglect

- Bullying
- Threats of violence

Additional information about item and test scoring is available in the *Implementation Guide for States and Service Providers* available in the *Manuals and User Guides folder of the Member Portal Document Library*.

## 7.8 TEST INTEGRATION AND TEST SCORING

Test Integration and Test Scoring are two separate components of the assessment delivery system. Both are discussed here.

### 7.8.1 Test Integration

Members will need to ensure that their procured assessment delivery system has the capability to integrate test data from the CAT and PT portions of the assessment. In addition, the assessment delivery system will need to send some item responses to external scoring systems while holding other item results that are machine-scored in real time. The system will then need to re-merge the item-level scored data once externally scored results are available. Once the item-level scored data are available, the system will need to assign test scores for each student.

### 7.8.2 Test Scoring

In addition to scoring individual items, members are responsible for producing test scores for individual students. Test scores are produced in accordance with specifications in the Smarter Balanced Scoring Specifications. Student test scores are based on a combination of item scores from the CAT and PT portions of the assessment.

### 7.8.3 Types of Scores

Smarter Balanced scores in English language arts/literacy and mathematics will be reported as a content area-level overall score and an achievement-level score. In each content area claim level scores will be produced and used to assign designations of "Below Standard," "At/Near Standard," or "Above Standard".

#### Overall Score

An individual student's overall performance in each content area will be reported in terms of a scale score and error band. Members will compute the overall score and standard error of measurement in accordance with specifications provided by Smarter Balanced.

**Achievement Level**

A student's overall performance in each content area will be classified into one of four achievement levels. Smarter Balanced provides members with achievement-level cut scores and specifications for classifying student performance.

**Claim-level Score**

Mathematics and English language arts/literacy assessments are reported by claims. Within each claim, individual student performance will be classified as "Below Standard," "Near Standard," or "Above Standard" in accordance with specifications provided by Smarter Balanced.

**Other Member-specific Scores**

With the exception of those scores specified by Smarter Balanced, members **will not be permitted** to produce and report individual student test scores on Smarter Balanced tests without prior written approval of Smarter Balanced.

## 7.8.4 Combining Performance on the PT and CAT

An individual student's overall score will be based on his or her performance across the PT and CAT portions of the assessment as described in the *Smarter Balanced Scoring Specifications*. Student performance on the individual PT and CAT portions of the assessment is not calculated and **will not** be reported separately.

**Computing an Overall Score**

Smarter Balanced provides the algorithm, or scoring model, needed to compute Smarter Balanced scaled scores from an individual student's raw scores on the items that were administered to the student. Smarter Balanced provides item parameters and all information needed for computing the overall score and conditional standard error of measurement. Smarter Balanced utilizes the two-parameter logistic (2PL) model and the generalized partial credit model (GPCM) to calibrate selected-response and polytomous items, respectively.

**Missing Data**

Smarter Balanced will provide specifications on computing an overall score when a student has not participated in one of the two portions of the assessment. Student participation in each portion of the assessment will be determined based on participation rules established by the member in a manner consistent with their approved federal workbooks and waivers.

### 7.8.5 Lowest Obtainable Scale Score (LOSS)/Highest Obtainable Scale Score (HOSS)

Options for Lowest Obtainable Scale Score (LOSS) and Highest Obtainable Scale Score (HOSS) values have been set in policy. Implementation of the option desired by each member needs to be negotiated with the test-scoring contractor. Beginning in 2015-16 Smarter Balanced members have the following options:

**Option 1:** Members may choose to retain the LOSS/HOSS values defined during the first test administration (SY 2014-15) which are shown in Table 11.

Table 11: 2014-15 Lowest and Highest Obtainable Scores

| Subject | Grade | Theta Metric | | Scale Score Metric | |
|---|---|---|---|---|---|
| | | LOT | HOT | LOSS | HOSS |
| ELA/literacy | 3 | -4.5941 | 1.3374 | 2114 | 2623 |
| ELA/literacy | 4 | -4.3962 | 1.8014 | 2131 | 2663 |
| ELA/literacy | 5 | -3.5763 | 2.2498 | 2201 | 2701 |
| ELA/literacy | 6 | -3.4785 | 2.5140 | 2210 | 2724 |
| ELA/literacy | 7 | -2.9114 | 2.7547 | 2258 | 2745 |
| ELA/literacy | 8 | -2.5677 | 3.0430 | 2288 | 2769 |
| ELA/literacy | HS | -2.4375 | 3.3392 | 2299 | 2795 |
| Math | 3 | -4.1132 | 1.3335 | 2189 | 2621 |
| Math | 4 | -3.9204 | 1.8191 | 2204 | 2659 |
| Math | 5 | -3.7276 | 2.3290 | 2219 | 2700 |
| Math | 6 | -3.5348 | 2.9455 | 2235 | 2748 |
| Math | 7 | -3.3420 | 3.3238 | 2250 | 2778 |
| Math | 8 | -3.1492 | 3.6254 | 2265 | 2802 |
| Math | HS | -2.9564 | 4.3804 | 2280 | 2862 |

**Option 2**: Members may choose to use other LOSS/HOSS values beginning in SY 2015-16 as long as the revised LOSS values do not result in more than 2% of students scoring below the LOSS

level and the revised HOSS values do not result in more than 2% of students scoring above the HOSS level.

**Option 3:** Members may choose to eliminate LOSS/HOSS altogether.

Additional Considerations

- All-wrong/All-right tests:
  - For all incorrect tests, score by adding 0.5 to an item score with smallest a-parameter among the administered operational items (CAT and PT) for a test.
  - For all correct cases, score by subtracting 0.5 from an item score with smallest a-parameter among the administered operational items (CAT and PT) for a student.
- Smarter Balanced will need to retain both the calculated theta score and the reported scale score for students whose scores fall into HOSS/LOSS ranges.
- If using Option #1 or #2 above:
  - When the scale score corresponding to the estimated theta is lower than LOSS or higher than HOSS, the scale score will be assigned associated LOSS and HOSS values. The theta score will be retained as originally computed.
  - LOSS and HOSS scale score rules will be applied to all tests (Summative, ICA, and IAB) and all scores (total and subscores).
- The standard error for LOSS and HOSS will be computed using theta ability estimates given the administered items. For example, in the formula in Section 5.1, $\hat{\theta}$=theta for the LOSS or HOSS, a and b are for the administered items.
- If using Option #3, the scale score is calculated directly from estimated theta.

### 7.8.6 Assigning an Achievement Level

An individual student's overall performance is classified into one of four achievement levels. Smarter Balanced provides members with achievement-level cut scores and specifications for classifying student performance.

### 7.8.7 Computing Claim-Level Scores

For each reported claim, student performance is classified into one of three claim-level performance categories. Members will compute a claim-level score based on items administered to the student and assign the appropriate reporting category ("Below Standard," "Near Standard," or 'Above Standard") in accordance with specifications provided by Smarter Balanced.

If composite claim scores are available for a member state, please refer to the composite score information in the Smarter Balanced Scoring Specifications.

### 7.8.8 Missing Data

Smarter Balanced provides specifications for computing and reporting claim-level scores when students have not responded to items within a claim area in the Smarter Balanced Scoring Specifications available in the Technical Documentation folder of the Document Library in the Member Portal.

### 7.8.9 Confirming the Accuracy of Test Scoring

Smarter Balanced released specifications and requirements needed to properly deliver the Smarter Balanced Assessment. As part of this process, Smarter Balanced developed an Implementation Readiness Package of specifications, sample data, tests, and other materials necessary to show compliance with Smarter Balanced specifications and requirements. Members should require their contractor(s) to show evidence that they are able to accurately apply scoring specifications provided by Smarter Balanced to produce Smarter Balanced test scores from raw scores across individual items, including demonstrating that the software and procedures used to implement Smarter Balanced scoring algorithms produce comparable Smarter Balanced test scores.

## 7.9 SENDING DATA TO SMARTER BALANCED

Smarter Balanced requests the submission of de-identified data from the summative assessment (at a minimum) from its members for the purposes of, but not limited to:

- Producing an annual technical report to:
  o Ensure that the assessment provides fair and accurate information regarding what students know and can do,
  o Provide information to stakeholders in a manner described by the 2014 Standards for Educational and Psychological Testing (AERA, APA, NCME), and
  o Provide annual counts of passages and other media exposure to the Copyright Clearance Center.

### 7.9.1 Authorizing Data Sharing

Member states may use the existing MOU between the state and the Regents of the University of California as authorization for sharing de-identified student level data. If a state requires an additional data sharing agreement, the state needs to inform Smarter Balanced of the need for a separate data sharing agreement.

### 7.9.2 Smarter Balanced Student Privacy Principles

In September 2013, Governing States adopted the policy that "Each member state retains control of its student-level data." Smarter Balanced affirms that:

- Members control the data that Smarter Balanced collects.

- Smarter Balanced and its members will collect a minimum amount of data.

- Members may elect not to submit official identifiers, student names, and dates of birth.

- Smarter Balanced, its members, and their authorized contractors will use industry best practices to ensure that student data are secure when the data are transmitted and stored.

### 7.9.3 Secure File Transfer

All files will be transferred via the Secure File Transfer Server. Smarter Balanced staff will provide the host name/IP address to member leads and member designated service providers where Consortium access is limited to Smarter Balanced and contractor staff respectively who have a legitimate need to access the information.

### 7.9.4 Data for Submission

Each member determines what data is submitted. At a minimum, members should provide de-identified data, which includes:

- A unique student ID (that may not be linked to the educational record, but is consistent over time).

- Gender and subgroups (as defined by the Elementary and Secondary Education Act).

- Eligibility for designated supports and accommodations.

- Test event information, including test instrument ID and test event ID.

- Test scores, claim scores, and standard errors of measurement.

- Response data, including item IDs, raw responses, and scored responses.

**Accepted File Types**

Data will be delivered in *Test Results Transmission (TRT) format* or an alternative format. The *Smarter Balanced Test Results Data Dictionary and Logical Data Model* on SmarterApp includes a data model and data dictionaries for test (student/examinee) and item-response level results. If a member submits data in a manner different from the TRT format, they should provide documentation regarding the differences.

# 8.0 ONLINE SUMMATIVE ASSESSMENTS

## 8.1 ONLINE SUMMATIVE TEST PACKAGES

All item content and test administration specifications for Smarter Balanced assessments are delivered in test packages posted in the Test Packages folder of the [Secure File Transfer Server](#). Separate Test Package files are provided for mathematics and ELA/Literacy.

For the 2023-24 school year, members may continue to administer the full form blueprint, or may administer an adjusted form blueprint that has been shortened to make it easier to administer.

The adjusted form blueprints were derived based on the previously adopted Smarter Balanced summative assessment blueprints (full form):

- The computer adaptive test (CAT) portion of the adjusted form blueprint was reduced by approximately 50% in each claim.

- Given that performance tasks are designed to be integrated tasks, the blueprints associated with the performance tasks were not adjusted.

The adjusted form blueprint is another option for members to leverage based on local needs.

**Resources**

- [Assessment Package Types](#)

## 8.2 ONLINE SUMMATIVE TEST ADMINISTRATION MANUAL

The *Online Summative Test Administration Manual (TAM)* is designed to support the administration of the Smarter Balanced online summative assessments. Additional information about the Online Summative TAM is available in *Section 5.1.1*.

**Member Responsibility**

It is expected that all members will need to customize the *Online Summative TAM* to meet their needs. This should include the contact information for the member organization and help desk, as well as changes due to state/local policy, configuration of the test administration system, and/or preferential edits that do not impact the validity of the assessments.

Throughout the Online Summative TAM, members will see highlighted text. This indicates that members should customize the section/topic according to member policies and procedures and according to the platforms used to administer the test. Members may edit the *Online Summative TAM* in sections that are not highlighted but should send any such changes to Smarter Balanced in advance to mitigate risks to validity, reliability, or comparability.

Some examples of needed areas of customization are:

- Member testing dates
- Member contact information
- Help desk information

- Student and Teacher Administration and Registration System user roles

- Test security policy

- Links for posted materials and modules

- Test security/administration training policy

## 8.3 ONLINE APPEALS—ONLINE SUMMATIVE ASSESSMENT

For test security incidents that result in a need to invalidate, reset, reopen, or restore, as defined in *Appendix B: Online System—Appeal Types and Conditions for Use*, for individual student assessments, **Smarter Balanced requires that requests be approved/denied by the member.** In most instances, an appeal will be submitted to address a test security breach or irregularity. *In some cases, an appeal may be submitted to address incidents that are not security related such as reopening an assessment for a student who becomes ill and is unable to resume the test until after the test has expired.*

Because the appeals system is for *action items only*— when a certain action needs to be taken on a test—all appeals that are also test security incidents must be recorded in a log or other mechanism devised by the member for tracking test security incidents.

The online appeals process and conditions for use are described in *Appendix B*. *Appendix B* also provides some further instruction on when a condition requires a notation as a test security incident versus only an appeal.

Only tests administered within the online summative system can be entered in the online appeals system. Paper-and-pencil appeals are addressed in the *Paper and Pencil TAM*.

### Member Responsibility

Smarter Balanced requires that all appeals be approved/denied at the member level, and the member will be responsible for monitoring and checking on appeals requests on a daily (or more frequent) basis during test administration.

Members can choose to approve or deny the request. Once a member takes action on a requested appeal, the requestor will be notified via email that an action has been taken and will be prompted to log in to view the details. Smarter Balanced recommends that members use the comments field to indicate the reason why an appeal was denied in order to allow school personnel a better understanding of the appropriate conditions for appeals.

### Member Monitoring

Members should monitor the requests made through the appeals system and ensure that the requests made and approved are valid, appropriate, and within the guidelines of appeals. As members identify recurring themes (e.g., students submitting the ELA/literacy PT before completing the full write), members may need to provide retraining opportunities with

schools/districts to mitigate these issues. Most appeals can be avoided with proper administration of the assessments.

**Resources**

*Appendix B* shows the Online System Appeal Types, with a description and condition for use for reset, reopen, invalidation, and restore.

## 8.4 ONLINE TEST SCHEDULING—ONLINE SUMMATIVE ASSESSMENT

The *Online TAM* contains general guidelines for scheduling/timing of the assessments—including approximate amount of time to administer, general guidelines for number of days to administer, and order of administration.

**Member Responsibility**

*Timing:* Members will need to work with their service providers to make decisions regarding timing and expiration of tests. Members may choose to be more restrictive (but not less) than Smarter Balanced requirements.

*Expiration:* Smarter Balanced requires that tests stay open no longer than 10 calendar days for the performance task (PT) and 45 calendar days for the computer adaptive test (CAT) after the student has begun the test; however, members may configure the system to have the tests open for shorter durations. Members may approve or delegate approval to districts, on a case-by-case basis, an extension of up to 10 additional calendar days for performance tasks within the test administration window. Test delivery systems could be configured to automatically expire performance tasks after 20 calendar days, unless the member has requested from their vendor for the performance tasks to expire after 10 days.

*Reopen:* Smarter Balanced recommends that CAT and PT tests reopened as a result of a "reopen appeal" remain open for ten calendar days from approval.

**Recommendation Regarding Documentation of Evidence**

In their evidence documentation, members should plan to include their test expiration policy for both the CAT and PT.

Smarter Balanced documents members' test expiration policies in the Member Test Administration Policy document available on the K-12 Leads page of the Member Portal.

## 8.5 REMOTE ADMINISTRATION OF SUMMATIVE ASSESSMENTS FOR 2023-24

Members may authorize LEAs to administer summative assessments remotely during the 2023-24 school year according to the following policy.

LEAs will administer tests consistent with the established policies for in-person test administrations, as described in the *Member Procedures Manual* and *Online Summative Test Administration Manual (TAM)*. Two important requirements to highlight, that apply to in-person and remote test administration are:

- Tests must be proctored in real time with a trained Test Administrator, who is an employee of the Local Education Agency (LEA). In lieu of LEA proctors, the state may allow for proctoring by a remote proctoring contractor; and,

- The test must be administered in a secure environment, where students have access to allowable resources.

While a primary obligation of Test Administrators is to support students during in-person testing, extra effort is necessary to support students during remote test administration; therefore, tests may be administered remotely only when there are **additional provisions to assist the student while the student is testing, as determined by the state.**

If a member allows LEAs to administer summative assessments remotely, the LEA must attempt to make available opportunities for all eligible students to test. This may include the following options:

- **In-person testing:** The Test Administrator and students are in the same room

  – traditional in-person testing centers in schools or other secure facilities;

  – testing resources brought to students at the students' homes or other location (e.g., library) by LEA staff for in-person testing; or,

- **Remote Testing**: The Test Administrator and students are in different locations and the test is administered consistent with this policy.

More detailed guidance for remote test administration is provided in the Remote Summative Testing webpage, *Remote Summative Testing Implementation Guidance* document and the *Supplement: Online Summative Test Administration Manual – For the Remote Administration of the Test of English Language Arts/Literacy and Mathematics*.

# 9.0 PAPER-PENCIL SUMMATIVE ASSESSMENT

## 9.1 ADMINISTRATION OF PAPER-PENCIL TESTS

The Smarter Balanced summative assessment is intended to be administered as an online assessment; however, Smarter Balanced makes available a paper-pencil test for members that elect to use it.

For the 2023-24 school year, members who elect to allow for the remote administration of paper-pencil tests may do so provided that there is a process in place to ensure test security

(E.g., a process to securely provide materials to the students and securely return the materials back to the school, a process for monitoring the administration of remotely administered paper-pencil tests).

### Paper-pencil Test Administration Window

Schools have a maximum of a three-week window in which to administer a paper-pencil assessment for a grade and content area. The window begins on the day a booklet for the content area/grade is unsealed at the school site or made available to the school unsealed (e.g., by the district). Members may establish a shorter window. In addition, members may approve exceptions on a school-by-school basis, based on the Member's determination that the exception does not constitute a material increase in risk to test security.

### Member Responsibility

Members will need to determine whether they plan to offer paper-pencil tests, and if so, should have a process in place for identifying the schools or districts or individual students requiring paper materials. Members will need to work closely with their service providers on all details of the paper administration including ordering, shipping, receiving, scanning, scoring, and reporting. States are responsible for establishing a process for monitoring the administration of paper-pencil tests to ensure they are administered in a way to ensure comparability between online and paper-pencil Smarter Balanced assessments, in areas such as timing, scheduling, accommodations, scoring, and reporting.

### Member Monitoring

Members are responsible for establishing a process for monitoring the administration of paper-pencil tests to ensure they are administered in a way to ensure comparability between online and paper Smarter Balanced assessments, in areas such as timing, scheduling, accommodations, scoring, and reporting.

### Resources

Smarter Balanced provides members with a variety of resources, as described in the next section, to support the paper-pencil assessment. The paper-pencil materials are available in the Test Packages/PaperPencil folder on the Secure File Transfer Server.

### Recommendation Regarding Documentation of Evidence

In their evidence documentation, members should plan to include their processes for all aspects of paper-and-pencil testing (e.g., test security, administration protocols).

## 9.2 TEST FORMS—PAPER-PENCIL SUMMATIVE ASSESSMENT

Smarter Balanced provides one operational paper-pencil form per grade and content area. Smarter Balanced provides the test books and answer booklets in a print-ready PDF format. The Smarter Balanced logo appears on the front cover of the test book. Members may also insert their own insignia on the front cover of the test booklet. The form is available in braille: Unified English Braille (UEB), UEB with UEB Math, and UEB with Nemeth Math codes from the American Printing House for the Blind (APH). The mathematics form is available translated into Spanish with a Dual Language Translation presentation that includes both English and Spanish.

## Items Included on the Paper-pencil Assessment

For each grade and content area, the paper-pencil test will include only operational items from both the adaptive and performance task sections of the online test. The paper-pencil forms include both machine and hand scored items.

The following item types are included on the paper-pencil form:

- Multiple Choice
- Multiple Select Response
- Table Response
- Two-Part Multiple Choice, with Evidence Responses
- Equation Response
- Short Text Response
- Essay/Writing Extended Response

In some instances, items were modified for delivery in a paper format. The modified items may have revised item parameters, and items that were originally machine-scored may require hand scoring. Updated item information for items appearing on each paper-pencil form is available in the Test Packages/Paper-Pencil folder on the Secure File Transfer Server.

## Member Responsibility

Members will work with their service providers to customize, print, package, distribute, retrieve, scan, score, and integrate the scores for any needed paper materials. This includes working with service providers to determine counts for printed materials, printed ancillary documents, as well as production and distribution of the paper form and accommodated materials such as Spanish forms of the mathematics test. Members are responsible for working with their service providers to order braille forms from APH or for ordering directly. Members can consider other methods besides scanning to enter student responses from paper tests into the system.

Service providers will need to merge student data from paper-pencil and online administrations for reporting and prior to submission of data to Smarter Balanced.

Smarter Balanced recommends that members engage in conversations with service providers *early* in a school year (late summer or early fall) to set up materials ordering systems and mechanisms for collecting school/district paper counts, agreeing upon print overages, packing specifications, mechanisms for identifying quantities, etc.

The assessments will be delivered in separate test booklets for each content area and grade, and students will respond in scannable response booklets by content area and grade. These will be provided in blueline copy in a print-ready format (except for logos and scan marks).

### 9.2.1 Braille Versions

Smarter Balanced will develop one operational braille test form for each grade and content area annually.

Smarter Balanced will provide student ancillary materials in braille, if necessary and appropriate. English glossaries in braille are included with the braille forms.

Members are responsible, either independently or in coordination with their service providers for ordering braille forms from the American Printing House for the Blind (APH) as well as for printing the accompanying teacher notes/scripts for the administration of the braille version of the test.

### 9.2.2 Spanish Translated Versions for Mathematics

Spanish translated mathematics paper-pencil operational forms for each grade level are available in the Test Packages/PaperPencil folder on the Secure File Transfer Server. The accompanying answer booklets and the test administrator script portions of the Spanish Paper-Pencil Test Administration Manual are also translated into Spanish. The Spanish-translated test booklets, answer booklets and Test Administration Manual are presented in Dual Language Translation format in both English and Spanish.

### 9.2.3 Large Print

Smarter Balanced will not provide a large-print Smarter Balanced Assessment version; however, members may require their vendor to create a large-print version of the Smarter Balanced paper-pencil assessment using the files provided for the regular form.

### 9.2.4 Summary of Provided Test Materials

Table 12 provides a summary of all test materials that Smarter Balanced provides for paper-pencil assessment.

Table 12: Summary of Paper-Pencil Assessment Materials for 2023-24

| Subject | Forms | Components of Each Form |
|---------|-------|-------------------------|
| ELA/Literacy | • Form 6 | **Print:**<br><br>• ELA/Literacy Paper-Pencil TAM<br>• Test booklets for each grade<br>• Scannable answer booklets for each grade<br><br>**Audio CD:**<br><br>• Listening passages<br><br>**Glossaries:**<br><br>• English<br><br>**Braille:**<br><br>• Braille test booklets for each grade<br>• English glossaries in braille for each grade |
| Mathematics | • Form 6 | **Print:**<br><br>• Math Paper-Pencil TAM<br>• Test booklets for each grade<br>• Scannable answer booklets for each grade<br><br>**Glossaries:**<br><br>• English<br>• Translations<br><br>**Spanish Print:**<br><br>• Spanish Translated Math Paper-Pencil TAM with Spanish Scripts |

| Subject | Forms | Components of Each Form |
|---|---|---|
| | | • Spanish Translated test booklets for each grade<br>• Scannable Spanish Translated answer booklets<br><br>Braille:<br><br>• Braille test booklets for each grade<br>• English glossaries in braille for each grade |

### 9.2.5 Test Booklets and Answer Booklets

The paper-pencil version of the Smarter Balanced Assessment includes test books and scannable answer booklets for each grade and content area. Members will need to procure services related to ordering, printing, packing, shrink wrapping secure test books, distributing, managing, retrieving, scanning, storing, and securely disposing of the paper-pencil test books. Table 13 shows the item and page counts for the operational and breach forms.

Table 13: 2023-24 Test Booklets Item and Page Counts

| Content Area | Grade Level | Form 6 (Operational) | | |
|---|---|---|---|---|
| | | # of items, including PT items | # of pages in test booklet | # of pages in answer booklet |
| Math | 3 | 36 | 36 | 12 |
| | 4 | 38 | 36 | 12 |
| | 5 | 39 | 36 | 12 |
| | 6 | 39 | 36 | 16 |
| | 7 | 40 | 40 | 16 |
| | 8 | 39 | 44 | 12 |
| | 10 | 41 | 40 | 12 |
| | 11 | 42 | 40 | 12 |
| ELA/ literacy | 3 | 38 | 56 | 16 |
| | 4 | 38 | 64 | 16 |

| Content Area | Grade Level | Form 6 (Operational) | | |
|---|---|---|---|---|
| | | # of items, including PT items | # of pages in test booklet | # of pages in answer booklet |
| | 5 | 38 | 60 | 16 |
| | 6 | 38 | 60 | 16 |
| | 7 | 38 | 64 | 16 |
| | 8 | 41 | 68 | 16 |
| | HS | 39 | 76 | 16 |

## 9.3 ANCILLARY MATERIALS—PAPER-PENCIL SUMMATIVE ASSESSMENT

### 9.3.1 Paper-and-Pencil Test Administration Manual

Smarter Balanced provides Paper-Pencil Test Administration Manuals for each form and content area for members choosing to offer schools a paper administration. The paper-pencil manuals work in conjunction with the Online Summative TAM to provide all necessary directions to administer the paper-pencil assessments. The Braille Test Administration Manual Insert provides additional information about administering a braille version of the test.

### 9.3.2 Listening Passages Resources

Smarter Balanced provides files required to produce audio CDs of the ELA/literacy listening passages. Members are responsible for mastering, replicating, and distributing the audio CDs. The audio files are available in the Test Packages/Paper Pencil folder on the Secure File Transfer Server. Alternatively, test administrators may read aloud scripts of the listening passages. The scripts are included in the secure ELA/literacy Paper-Pencil Test Administration Manual.

### 9.3.3 Manipulatives

The paper-pencil assessments do not include or allow for the use of manipulatives.

### 9.3.4 Calculators

Starting in Grade 6, the mathematics test is segmented into calculator and no calculator sessions. Members must ensure that calculators are available to students for testing purposes. Refer to the *Mathematics Paper-Pencil Test Administration Manual* for information on allowable calculators by grade.

### 9.3.5 Translated Glossaries

Smarter Balanced provides glossaries for the mathematics paper-pencil forms in each of 13 languages. The glossaries are one to multiple pages in length. Members are responsible for making the glossaries available to districts who plan to administer the paper-pencil test in advance of test administration so that they may print these glossaries locally.

## 9.4 TEST SECURITY—PAPER-PENCIL SUMMATIVE ASSESSMENT

Maintaining the integrity and security of the Smarter Balanced assessments is critical, and members will be required to develop a mechanism for tracking and logging all Smarter Balanced secure test materials. Members will be expected to work with their service providers regarding test security of paper-pencil assessments. This may include barcoding of all secure documents with a unique tracking number, coding documents with assigned student information, or manually logging test booklets out/in and then reconciling any differences within your state. Members should follow best practices for distributing and collecting secure test materials. Test security is outlined in the Operational Best Practices for statewide Large-Scale Assessment Programs (2013).

### 9.4.1 Retention of Secure Test Materials in Support of the FOIA and FERPA

Smarter Balanced does not prescribe the length of time that secure test materials must be retained. Individual member policies guiding the retention of secure materials in support of the Freedom of Information Act (FOIA), Family Educational Rights and Privacy Act (FERPA), and any member-specific policy should be followed for Smarter Balanced materials.

## 9.5 SCANNING AND SCORING TESTS

The paper-pencil tests are comprised of items that require both machine and hand scoring. Once items are scored, vendors will need to assign test scores for students taking the paper-pencil test forms.

### 9.5.1 Scanning and Machine Scoring of Responses

Members will need to procure services to securely scan answer documents or enter responses into a data system and machine score applicable item types. Members should ask vendors to guarantee and demonstrate accuracy of the scanning machines for the paper-pencil forms. Members should ask vendors to describe methods for scoring the paper-pencil student response booklets, beginning with receipt control methods and continuing through the step at which all data have been transferred to a master file.

Smarter Balanced provides test maps for machine-scored items on the paper-pencil forms. These documents are available in the Test Packages/PaperPencil folder on the Secure File Transfer Server.

### 9.5.2 Hand Scoring

Smarter Balanced provides scoring rubrics, training papers, and validation papers to support scoring of paper-pencil test responses. These documents are available in the Scoring Materials folder on the Secure File Transfer Server.

In addition, some paper pencil items that are machine scored when delivered online will need to be hand scored. Smarter Balanced provides the Supplemental Scoring Materials for hand scoring these items available in the Test Packages/PaperPencil folder on the Secure File Transfer Server.

### 9.5.3 Double Scoring

Smarter Balanced recommends that 10% of student responses be the minimum level of double scoring allowed, however, each member shall determine the percentage of its student responses that it wants double scored. When there is not exact agreement between two raters, Smarter Balanced currently uses the resolution read performed by a senior rater as the score of record. Smarter Balanced expects members to collect item-level, hand-scoring statistics (e.g., inter-rater reliability) for each hand-scored item for all item responses that receive two reads.

Vendors may score students responses from multiple member clients; thus, members may want to request item-level, hand-scoring statistics aggregated across all members scored by the vendor as well as item-level, hand-scoring statistics for an individual member.

### 9.5.4 Scoring and Scaling Paper-Pencil Tests

Smarter Balanced includes the item-parameter files needed to calculate paper-pencil test scores in the test maps. The test maps are available in the Test Packages/PaperPencil folder on the Secure File Transfer Server. Refer to *Section 7.7 Item Scoring* and *Section 7.8 Test Integration and Test Scoring* for additional information about item and test scoring.

# 10.0 INTERIM ASSESSMENTS

The Smarter Balanced interim assessments allow teachers to check student progress throughout the year, so educators can use the results to inform instruction and help students meet the challenge of college- and career-ready standards. The interim assessments may be administered at multiple points throughout the school year. For the 2023-24 school year, the start date for interim assessments may be August 1, 2023, or later.

The interim assessments are available in ELA/literacy and mathematics in grades 3–8 and high school; however, teachers may administer individual assessments at any grade level. Out of grade-level interim administration is a local decision (e.g., administration of a grade 4 interim to students in grade 3, or a grade 3 interim to students in grade 4). The optional and flexible interim assessments can be administered in either a standardized or non-standardized manner.

In a standardized administration the test is administered using the same protocols as a summative assessment administration. Non-standardized administration refers to any administration that is not consistent with the administration requirements of the summative assessment. Some examples of non-standardized administration might include (but are not limited to):

- Administering tests while students answer cooperatively in pairs, in small groups, or as a class. Teachers may elect to include some discussion time between test items.

- Providing access to classroom resources that may support the students' understanding of the assessed content.

Results from a non-standardized administration of the interim assessments are more appropriately used as assessment for learning (formative assessment process) rather than the assessment of learning. Because non-standardized administration does not necessarily describe the performance of individual students in a comparable manner, caution must be used when making instructional decisions based on data from a non-standardized administration.

There are three types of Smarter Balanced interim assessments: The Interim Comprehensive Assessments (ICAs), the Interim Assessment Blocks (IABs) and the Focused IABs. All ICAs and IABs draw from a shared bank of items, including performance tasks, meaning there are overlapping items in the IABs and ICAs in each content area and the performance task is the same whether given as an IAB as part of the ICA.

The interim assessments are designed to be administered as computer-based assessments and are not available in paper-pencil format. All interim assessments offer the same breadth and depth of universal tools, designated supports and accommodations available on the summative assessment. Unlike the Smarter Balanced summative assessments (which are adaptive), the ICAs and IABs are fixed-form tests, which means the tests are static and do not adapt according to student responses. Because the interim assessments are used to support instruction, providing each student with the same test items allows teachers to more easily interpret their students' performance on the same items. It also allows teachers to better manage hand scoring since all students respond to the same constructed-response questions.

Consortium members have the flexibility to allow teachers to administer, and students to take interim assessments remotely under the following conditions:

- An authorized employee (e.g., teacher, test administrator) in a school administers the test consistent with the district or school policies for in-person interim assessment administration.

- The test administrator monitors the test activity such that tests are open only for the minimum amount of time necessary for students to complete and submit their responses.

- The test administrator uses established test administration practices to support students accessing the interim assessments; this may be a phone call or chat with a parent/guardian in advance of starting the test.

- The test administrator maintains student data privacy with student State-wide Student Identifiers (SSID) and other Personally Identifiable Information (PII), which are required to take an interim assessment. (Do not send PII over email, chat or text or other non-secure transmission methods. Please refer to local policies regarding communicating PII.)

- The test administrator follows state and local policies regarding test security and immediately escalates to the test coordinator any suspected item security issue.

  - A state or territory may establish a policy that allows for interim assessments to be administered without the use of a secure browser provided that the applicable test security protocols from *Section 10.5* are followed.

  - A state's policy must include procedures to address item security as described in *Section 10.5 Interim Test Security* (e.g., posting on social media).

Additional information on remote administration of interim assessments can be found on the Administering Interim Assessments webpage.

## 10.1 TYPES OF INTERIM ASSESSMENTS

### 10.1.1 The Interim Comprehensive Assessments (ICA)

The Interim Comprehensive Assessments (ICAs) measure similar content to the summative assessment. Therefore, the ICAs may be helpful for purposes such as determining the knowledge and skills of students who are new to the district or the state and providing interim information after a significant period of instruction.

There is one ICA per grade and content area, which includes a performance task. The ICAs take about the same amount of time to administer as the Smarter Balanced summative assessments when administered under standardized conditions.

### 10.1.2 The Interim Assessment Blocks (IAB) and Focused IABs (FIABs)

The Interim Assessment Blocks (IABs) are short item-sets that include 6–18 items. Throughout the school year, teachers can use these sets of related concepts in ELA/literacy and mathematics to check students' understanding. Since the IABs are more granular than the ICAs, teachers can administer these assessments throughout the school according to the sequence of their curricula. Some of the IABs assess as many as 4 to 8 targets and other more focused IABs (FIABs) assess only 1 to 3 targets. Each IAB has a corresponding Connections Playlist available in Tools for Teachers, which provides aligned resources teachers can use for instructional next steps to support student learning.

Teachers should know that there are overlapping items in the IABs and ICAs in each grade and content area.

FIABs that assess no more than 3 assessment targets and their associated instructional resources provide educators with a deeper understanding of student knowledge and skills and next steps for instruction. For current information about available interim assessments, see the Interim Assessments Overview available on the Test Development and Design page of the Smarter Content Explorer and in the Document Library of the Smarter Balanced Member Portal.

**Interim Assessments at a Glance:**



## 10.2 INTERIM ASSESSMENT ITEM PORTAL (IAIP)

The Interim Assessment Item Portal (IAIP) provides educators with the ability to use interim assessment items in more flexible ways that support student learning using the formative assessment process. The IAIP includes all live items on the available Interim Assessment Blocks (IAB), Focused IABs (FIABs) and the Interim Comprehensive Assessments (ICA). The IAIP may be used by educators to view and select or deselect individual assessment items to tailor content covered by an interim assessment, so it better aligns with the sequence of their instruction. Educators may use the IAIP in addition to the fixed-form IABs, FIABs, and ICAs for instruction, as determined by the member state. The IAIP is available for integration into Tools for Teachers by member service providers.

## 10.3 INTERIM ASSESSMENT BLUEPRINTS

The interim assessment blueprints provide additional information about the content measured by each assessment, including which claim(s), assessment target(s), and depth of knowledge level(s) are addressed by the items, as well as the number of items by target or target group.

The ICA blueprint describes the assessed content for the comprehensive assessments, which is the same content assessed on the summative assessment. The IAB blueprints can be used by educators to plan how to integrate the IABs effectively within classroom instruction or to better understand results that are reported. The IAB blueprints can be used to determine which assessment targets are addressed in a specific IAB and the emphasis of each target relative to the other targets in the IAB. A fifth-grade teacher, for example, may wish to collect more information regarding her students' knowledge about geometry. The teacher could use the Grade 5 Geometry blueprint to see that this IAB is composed of 13 machined-scored items across the four claims: concepts and procedures, problem solving, communicating reasoning, and modeling and data analysis.

The current interim assessment blueprints are available for download from the Test Development and Design page of the Smarter Content Explorer and in the Document Library of the Smarter Balanced Member Portal.

ELA/Literacy ICA Blueprint

[Mathematics ICA Blueprint](#)

[ELA IAB Blueprint](#)

[Mathematics IAB Blueprint](#)

[ELA Focused IAB Blueprint](#)

[Mathematics Focused IAB Blueprint](#)

## 10.4 INTERIM ITEM POOL

The interim assessment items were developed using the same specifications as the items in the summative assessment item pool. The items in both the summative and interim pools were field tested in 2014 or in subsequent years as embedded field test items within the operational test.

The Consortium provides interim assessments that include all of the universal tools, designated supports, and accommodations listed in the *Usability, Accessibility, and Accommodations Guidelines*.

## 10.5 INTERIM TEST SECURITY

Interim assessments can serve a variety of educator needs. To better support the range of possible uses consistent with member education agency policies, educators may establish the timeframe, administration policies, and local scoring practices for interim assessments. The interim assessments are designated as student and teacher facing. The student- and teacher-facing designation means that educators have the flexibility to access the test questions and their students' responses to the test questions. This allows educators to use the interim assessments as part of the formative assessment process to improve teaching and learning. Because of this flexibility, the interim assessments are not intended to be used for accountability purposes.

The student- and teacher-facing designation also means that interim assessments are not for public use, display, or distribution. For this reason, any use, display, or distribution of the interim assessments that results in access to individuals beyond authorized local education agency staff and students is prohibited. Finally, interim assessment items must **not** be copied into a third-party application that is not approved by the member state education agency and in consultation with Smarter Balanced.

Note: A state or territory may establish a policy that allows for interim assessments to be administered without the use of a secure browser provided that the above test security protocols are followed.

## 10.6 SPECIFICATIONS FOR THE ADMINISTRATION OF INTERIM ASSESSMENTS

The ICAs and IABs are administered online through the member service provider's test administration system. Members should provide educators with the following considerations for using the interim assessments.

- **Multiple configurations:** Districts/schools may elect to administer both the ICAs and IABs during the school year.

- **Multiple administrations:** The ICAs and IABs (including those for the same block) may be administered multiple times within an academic year. Members may determine the schedule for interim assessments or delegate the administration of interim assessments to schools/districts. The Consortium does not limit the number of times that the ICAs and/or IABs are administered. Members will need to work with service providers to determine contractual expectations related to the number of administrations.

Despite this allowance, members should be aware of risks associated with administering these assessments multiple times within an academic year.

- *Item over-exposure:* testing multiple times a year limits the item pool available to students, which will increase the possibility of students encountering the same item several times. Over-exposed items are unlikely to hold their original parameters and may skew performance results. To prevent this, schools and classrooms may want to limit their testing program to either a judicious use of ICAs or to coordinated use of IABs.

- *Hand scoring:* Some of the interim items require hand scoring. Members will have to allocate resources (either vendor resources or local resources) to hand scoring and multiple administrations may add to this effort.

- **Grade levels:** The ICAs and IABs are comprised of items aligned to the CCSS in grades 3–8 and high school. However, the interim assessments are not constrained by grade level. Students may take any grade level interim assessment.

Smarter Balanced does not prescribe the order or timeframe in which the ICAs or IABs are administered; this decision will be made at the local (e.g., member, district, school, or teacher) level.

In addition, a member can create its own version of the interim assessments using a unique member blueprint and the Smarter Balanced bank of interim items. If a member chooses to create their own version of the interim assessments, the member would be responsible for test development and calibrations for these tests.

### 10.6.1 Interim Assessment Guide for Administration

*The Interim Assessment Guide for Administration* provides detailed information about the interim assessments, interim test administration options, the Smarter Balanced Open-source Reporting System, and connections to Tools for Teachers. This customizable document is available in the Manuals and User Guides folder of the Member Portal Document Library and in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer.

**Resource**

- The Interim Assessment Overview and Blueprints are posted on the Test Development and Design page of the Smarter Content Explorer and in the Member Portal Document Library.

## 10.7 INTERIM HAND SCORING

Both the ICA and IAB contain some items that require hand scoring. The member service provider's interim assessment hand scoring application enables educators to hand-score items locally and submit those scores for integration into the overall score of the test. Smarter Balanced provides rubrics, exemplars, and training guides for each hand-scored item. Training Guides are available in the Scoring Materials/Interim Scoring Materials folder of the Secure File Transfer Server. Member service providers may make them available through a password protected link on the member's assessment portal. Members may want to develop training protocols to support educators scoring these items locally.

The interim assessments are fixed-form and all students take the same hand-scored items.

**Member Responsibility**

Members are responsible for determining the approach for the hand scoring of interim items (scored by teachers or contracted through a service provider).

## 10.8 INTERIM REPORTING

### 10.8.1 Options for Reporting Interim Results

The results from the interim assessments may be stored and reported through the Smarter Balanced Data Warehousing and Reporting System or through a member service provider's reporting system. Members are not required to send interim test results to Smarter Balanced. The only reason for members to send interim results to Smarter Balanced is if they are using the Smarter Balanced instance of the Data Warehouse and Reporting system to generate interim test reports.

### 10.8.2 Reporting ICA Results

The reporting for ICA results is the same as the reporting for the summative assessments. The reports will include an overall scale score with margins of error and an achievement level per content area as well as claim-level information. At the claim level, students are assigned to one of three reporting categories (Below Standard, At/Near Standard, Above Standard) related to the overall scale score at the established member-wide Level 2/Level 3 cut point.

### 10.8.3 Reporting IAB Results

Results for IABs will be classified into one of three reporting categories (Below Standard, At/Near Standard, Above Standard) related to the overall scale score at the established member-wide Level 2/Level 3 cut point.

The Smarter Balanced open-source reporting system provides detailed information about student responses to ICA and IAB test questions, including, a scale score and standard error of measure, an item viewer with student responses, the item's rubric and exemplar, and item metadata (e.g., claim, target, standard, item difficulty, Depth of Knowledge).

Note: The results from the interim assessments are not intended to be used for accountability purposes.

# 11.0 HELP DESK

**Member Responsibility**

Members will need to procure or provide the member-level help desk support for schools and districts for the Smarter Balanced assessments, including the Practice and Training Tests, summative assessments, interim assessments, and Tools for Teachers.

Members will address their specific requirements with their service providers around help desk reporting, hours, contact methods, service-level agreements, tiered support, etc. The *Implementation Guide for States and Service Providers* provides specific considerations for members when soliciting customer support/help desk services as well as guidance around what type of support is handled by each tier. *The Implementation Guide* is available in the Manuals and User Guides folder of the Member Portal Document Library and in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer Server.

A customer support and help desk solution will require Tier 1, Tier 2, and Tier 3 services. States need to provide Tier 1 and Tier 2 support. Smarter Balanced provides Tier 3 support for the open-source Smarter Balanced Reporting System, Tools for Teachers, and for any legacy users of the open-source Administration and Registration Tool (ART) or Smarter Balanced Single Sign-On. States need to provide Tier 3 support for all other aspects of the assessment system,

including any proprietary applications. States typically procure help desk and customer support services from their service provider.

---

**Resource**

- The *Implementation Guide for States and Service Providers* is available in the Manuals and User Guides folder of the Member Portal Document Library and in the Assessment Training and Operations/Manuals and Guides folder of the Secure File Transfer Server.

---

# 12.0 RIGHT TO VIEW STUDENT RECORDS

---

**Member Responsibility**

Smarter Balanced requires that members and districts maintain the security of the Smarter Balanced assessments and student data. In addition, members must comply with the Family Educational Rights and Privacy Act (FERPA). Unless otherwise specified by existing member policy or regulation, members may allow parent(s)/guardian(s) to review secure test items and student responses as necessary to meet FERPA requirements. If members establish a Right to View Policy, Smarter Balanced requires that members must have a process to comply with the following requirements:

- Request to View Secure Materials:

    – Require parent/guardian to show reasonable proof of identity and parent/guardian status prior to review.

    – Require that the parent/guardian sign a nondisclosure agreement (NDA) prior to review.

- Parent/Guardian Review of Secure Materials:

    – Review must be supervised at all times by member or district staff.

    – Student(s) or other non-authorized individuals are not present in the meeting room.

    – All electronic devices (including cell phones) are turned off and made inaccessible during the review.

    – Note taking, photography, or recording is not allowed.

    – Upon completion of the meeting, the member or district staff must return the materials to be archived, securely shredded, or deleted from any system, as appropriate.

# 13.0 DATA WAREHOUSE AND REPORTING SYSTEM

The Smarter Balanced Data Warehouse and Reporting System is accessed through a single-sign-on feature and it consists of two main components.

## 13.1 DATA WAREHOUSE

The data warehouse component compiles the data into two data stores: individual item response and reporting. See *Section 4.1* for information about the open-source system. Additional technical information, including specifications and layouts regarding the data warehouse, can be found on [SmarterApp.org](SmarterApp.org).

- **Individual Item Response Data Store** holds the student responses to individual items.
- **Reporting Data Store** holds the scored student test data that the reporting system will use.

## 13.2 SMARTER REPORTING SYSTEM

Smarter Balanced has developed its own reporting system. Members can choose to utilize the Consortium-hosted Smarter Reporting System, deploy their own instance of the open-source system adapted to their needs, or utilize a different system.

The Smarter Reporting System allows users to produce public-facing and secure score reports for both summative and interim assessments. These online reports comprise individual student reports and aggregate reports used by school, district, and state-level users. The Smarter Reporting System is also used to display student results in a variety of different ways, including individual item responses for interim assessments. All reports are available online. The Individual Student Reports (ISR) include an option for printing or downloading a PDF version.

The reporting system allows for the creation of optional student groups (e.g., teacher, class). This feature allows a user with the Group Administrator role to create and manage groups for a district or school. Student groups are assigned to users who may view test results for groups.

The Smarter Reporting System allows users the ability to filter by certain criteria

Filters defined by US Department of Education:

- Gender
- Race/Ethnicity
- English learner status
- Military connectedness
- Migrant status

- Lower socioeconomic status

- Disability designation (504/IEP)

Non-US Department of Education filter:

- Primary Language

The reporting system is an online reporting platform that provides reports at the individual student, classroom, grade, school, district, and member levels (in addition to data extract capabilities). These reports include:

- **Individual Student Report (ISR)[1]** — This secure report provides the student's mathematics and ELA/literacy overall scale scores and achievement level for the selected assessment (summative or ICA). It includes a description of the achievement level, claim information, writing trait scores (ELA/literacy), and the option to include accommodations available to the student during testing.[2] The ISR for the Interim Assessment Blocks (IAB) reports a scale score, standard error of measurement, and student performance as one of three reporting categories: Below Standard, Near Standard, or Above Standard.

- **Interim Assessment Block (IAB) Dashboard** — This report provides an overview of student performance on all IABs administered to a group of students.

- **Results by Student** — This report provides a roster of students with summative, IAB or ICA results for a particular grade level in a school or district. The report displays individual student achievement levels (summative/ICA) and reporting categories (IAB) and each student's scale score with an error band.

- **Claim Report (Summative and ICA)** — This report provides the distribution of claim scores for a group of students. The Claim Report available to teachers provides both aggregate and individual student data. The Claim Report, available to district/school administrators, is an aggregate report.

- **Target Report (Summative only)** — This report provides target-level data for the ELA/literacy and mathematics summative assessments. The Target Report available to teachers is an aggregate report at the student group (classroom) level. The Target Report available to district/school administrators is an aggregate report.

---

[1] ISR for 8th and 11th grade reports—members may develop *a higher education conditions policy* and work with their service providers to have a link to that policy inserted into this report (placeholder will be there to identify exact location).

[2] The ISR is intended for educator use to be provided to students and parents/guardians to review student performance. Members may also develop a customized student score report for distribution to parents/guardians.

- **Results by Item (Interim only)** — This report provides item-level information for IABs and ICAs. The report displays individual student and group performance on each interim test item. It includes the claim and target, item difficulty, content standard assessed, and depth of knowledge (DOK) for each item. The report also allows teachers to view student responses to interim test items and provides the rubric and exemplar for each item.

- **Key and Distractor Analysis (Interim only)** — This report displays the percentage of students who selected each answer option for multiple choice and multiple select items.

- **Writing Trait Scores** — This report displays student performance on the three writing traits: Evidence/Elaboration, Organization/Purpose, and Conventions. IAB and ICA reports include individual and group performance, while summative reports include only group performance.

- **Custom Aggregate Reports —** School and district administrations can create customized reports for overall student performance, claim and target performance, and longitudinal reports, including the ability to disaggregate results by subgroup performance.

**For more information, see *Section 14.3 Types of Reports Available* in the Implementation Guide for States and Service Providers.**

### Member Responsibility

Members will need to determine how they want to utilize the Smarter Balanced reporting system and then work within their state, territory or agency or with their service providers to adapt the system if needed (or develop a new/different system).

### Member Monitoring

Members should monitor whether reports are received and distributed appropriately (e.g., if the member sends parent/guardian reports to districts/schools for distribution, they should ensure parents/guardians receive those reports in a timely manner) or that reports are available to parents/guardians, schools and the public electronically in a timely manner.

### Recommendation Regarding Documentation of Evidence

In their evidence documentation, members should plan to include information about report production, availability, and an appropriate communications plan to inform students, educators, parents/guardians and other stakeholders about student performance on the summative assessment.

### Resources

- The Smarter Reporting System User Guide provides information about the functionality and reports available in the reporting system. This document is available in the Smarter Balanced Member Portal.

## 13.3 DATA SHARING WITH HIGHER EDUCATION INSTITUTIONS

Higher education institutions in member states and territories are taking part in the Consortium with the goal of recognizing an agreed-upon achievement level on the high school assessments to determine readiness for placement into college-level, credit-bearing courses. According to the Consortium's College Content-Readiness Policy, students who score at Achievement Levels 3 or 4 and abide by any Grade 12 conditions jointly established by K–12 and higher education can be assured they will be exempted from developmental courses at participating colleges.

**Member Responsibility**

So that students can take advantage of this policy, members who administer the Smarter Balanced high school test need to establish a process for securely transferring high school summative assessment results to higher education institutions in compliance with member and federal privacy requirements. Members should consult with their state or territory's higher education agency or college/university system offices to determine the best method for making the needed assessment results available to colleges and universities. In the meantime, Consortium members are exploring the joint development of a mechanism that would allow members to share scores with out-of-state higher education institutions as well as other in-state school districts.

# 14.0 FIELD TESTING NEW ITEMS

To maintain the item bank, it will be necessary to continue online field testing throughout the life of the assessment program. Smarter Balanced is responsible for future item development and may contract with a variety of entities, including members, to develop items. Members are encouraged to recruit educators to participate in the item writing, content review, fairness review, and item-data review processes. This supports the Smarter Balanced Theory of Action that includes educator involvement in the assessment development process. Smarter Balanced will field test online, newly developed items during each test administration cycle. All members participating in Smarter Balanced will participate in the field testing of new items.

## 14.1 FIELD TESTING COMPUTER ADAPTIVE TEST ITEMS

Field test computer adaptive test (CAT) items will be embedded within the operational CAT for each student and will appear seamless to students. These items will be included and identified in the test package delivered to members for each online administration and will be controlled through the computer adaptive testing algorithm. The CAT item selection algorithm determines which field test items are administered, when they are administered, and which students receive field-test items. States and their service providers will not be responsible for determining which students are administered field-test items.

The number of students administered field-test items within a particular state or territory will be based on the number of students in the state or territory and the number of items field tested within a particular administration. The assignment of field test items to students will follow the same procedure in all member states.

The CAT field test items are randomly administered to all students. No embedded math field test items in the CAT require hand scoring. For ELA/literacy, Smarter Balanced field tests items in alignment with the blueprint, which includes 'short text' items in the CAT for grades 6-8 and high school which the field-test scoring vendor may need to hand-score. Smarter Balanced uses approximately 1,500 responses per item for computing item statistics and calibration.

Smarter Balanced will assume responsibility for the scoring of all field test items. The amount of field testing per student is minimal—no more than 6 items per student—and is not expected to impact the overall testing time.

There will be no field-test items on the paper-and-pencil form provided by Smarter Balanced.

## 14.2 FIELD TESTING PERFORMANCE TASKS

States will administer performance tasks (PTs) from the operational pool, as well as performance tasks that need to be field tested. States, in collaboration with their service providers, will randomly assign the field-test performance tasks to a small percentage of students throughout the testing window. To support the reporting of claim results without requiring the rapid calibration cycle of field test performance tasks, students who are administered a field-test PT will be presented with a small number of additional items on the computer adaptive portion of the test. Students who are assigned to a field-test PT in English language arts/literacy (ELA/literacy) will not receive writing trait scores for the ELA/literacy full write from their operational test. Smarter Balanced provides states with these writing trait scores for ELA/literacy field-tested PTs after they are subsequently scored and calibrated.

The Enhanced Computer Adaptive Test (CAT) Blueprint provides information about additional CAT items administered to students who participate in the embedded field test of PTs. This document is available in the Summative Assessment Blueprints and Estimated Testing Times folder in the Document Library of the Smarter Balanced Member Portal and on the Test Development and Design page of the Smarter Content Explorer.

### 14.2.1 Roles and Responsibilities Regarding Embedded Field-Test Responses

As part of the embedded field test, at least four PTs will be included in each grade and content area assessment. Each task requires approximately 2000 student responses across the Consortium.

Members are responsible for:

- Administering the tasks in a manner consistent with the *Smarter Balanced Online Summative Test Administration Manual*,

- Sensitive responses until they are handed off to the Smarter Balanced field test-scoring contractor (See *section 7.7.3 Scoring of Sensitive Papers*),

- Providing the student responses associated with the uncalibrated tasks as well as the student responses to all other items to the Smarter Balanced scoring contractor.

Smarter Balanced is responsible for:

- Developing the PTs for each test administration year,

- Range finding, assigning scores to student responses, conducting data reviews and calibrating the embedded field test performance task items,

- Producing anchor papers to support operational scoring, and

- Screening responses after receiving them from members and alerting members of sensitive responses (See *section 7.7.3 Scoring of Sensitive Papers*) per the guidance about topics that may require action in the *Smarter Balanced Online Summative Assessment Test Administration Manual*.

### 14.2.2 Plan for the 2023-24 Embedded Field Test of Performance Tasks

For the 2023-24 embedded field test, there will be four (4) ELA/literacy field-test performance tasks (PTs) per grade in grades 3–8 and high school. There will be four (4) mathematics field-test PTs in grades 3- 8, and high school.

Students who receive an embedded field-test PT will also receive an enhanced CAT that includes additional items.

For English Language Arts/literacy (ELA/literacy), students in grades 3-8 and 11 will receive four additional items. Three of these items will contribute to students' scores for the ELA/literacy writing (claim 2) reporting category. One of the four items will contribute to students' scores for the ELA/literacy research (claim 4) reporting category. For mathematics, students in grades 3-8 and 11 will receive two additional items. Both items will contribute to the reporting of students' scores for the Problem Solving & Modeling and Data Analysis (claims 2 and 4) reporting category. On average, it will take approximately 10–15 additional minutes for students to answer the additional CAT items.

# 15.0 TOOLS FOR TEACHERS

The Tools for Teachers website is an online collection of instructional and professional learning resources. These resources are aligned with the Common Core State Standards and help educators implement the formative assessment process to improve teaching and learning. Tools for Teachers was developed by educators for educators and provides a variety of resources, including:

- Instructional Resources developed by educators through a collaborative development and quality review process utilizing a set of consistent criteria,

- Formative Assessment Strategy and Accessibility Strategy resources embedded within instructional resources and available independently for educator use,

- Interim Connections Playlists, which are collections of resources that provide educators with a means of understanding the student performance progressions of a concept, and

- Professional Learning Resources focused on educator professional development, mentoring, and coaching.

Tools for Teachers includes several innovative design features that make it user-friendly. State-of-the-art search features help educators locate materials which can be bookmarked and annotated with the "notes" feature. In addition, Tools for Teachers is web accessible (i.e., WCAG 2.1 AA compliant). Resources are built with the realities of classroom instruction in mind: they include options for differentiating instruction, student-focused accessibility strategies, and are embedded with cross-cutting formative assessment strategies that can be transferred to other lessons and activities across content areas. Also embedded across the site are hyperlinks to navigate to other Smarter Balanced applications to allow educators easy access to address questions and curiosities with ease.

Tools for Teachers is available to eligible educators in member states or local educational agencies (LEAs). Members may also provide higher education faculty, especially faculty to teach in educator preparation programs, with access to Tools for Teachers. Members, or their service providers, are responsible for working directly with districts and institutions of higher education to set up and maintain user accounts and to provide help-desk support.

Tools for Teachers is embedded with "help features" throughout the site. Additional support may be needed beyond what the help function provides. Technical support for Tools for Teachers will be provided by the member. For more information regarding technical support, see *Section 16.3 Tier 3 Support* of the *Implementation Guide for States and Service Providers*.

## 15.1 MEMBER-OWNED TOOLS FOR TEACHERS RESOURCES

Tools for Teachers is built using a multi-tenant approach. Multi-tenancy affords members with options for customization in terms of the user interface as well as the underlying database. One such customization is the development and rendering of member-owned resources (i.e., resources only available to educators within a specific consortium member state/territory).

Members may fund and organize the development and posting of additional Tools for Teachers resources outside those provided by Smarter Balanced (e.g., resources related to additional subjects like science). In terms of access to these resources, the federated Single Sign-On (SSO) model allows members and service providers to provision access to Tools for Teachers to additional user groups (e.g., science teachers) as needed.

There are four components of the development and rendering of member-owned resources which are further described in the paragraphs below:

- Resource Authoring: resource development, quality assurance, copyright permission, and resource attachment accessibility conformance, and entry into the Smarter Balanced Test and Item Management System (TIMS) Workflow approvals in TIMS are completed by the appointed member-owned resource administrator for copyediting, copyright permission validation, and the resource attachment accessibility audit.

- Resource Storage: which takes place and is managed within TIMS.

- Resource Posting: involves a sync from TIMS to the Tools for Teachers data layer, which requires a final quality review by the member-owned resource administrator.

- Viewing & Accessing (i.e., availability of resources in Tools for Teachers and access to the site).

The authoring of member-owned resources will be wholly funded and organized by the member including the costs of any in-person or virtual resource development workshops or events, resource author contracts or other compensation, copyediting, document accessibility remediation, and other associated costs. Members will ensure the quality of all resources and are responsible for obtaining and storing copyright permission associated with resource development. Members will use existing resource authoring guidance specifying the sections and meta-data fields for each resource and will author the resources in TIMS. Smarter Balanced will share knowledge, materials, and practices for the established resource creation and approval process for member consideration.

Member-owned resources will be authored, reviewed, and stored within the Smarter Balanced TIMS. A limit to the number of member-owned resources will be established as an up-to limit. Limits to the number of stored resources and TIMS authoring accounts will be established once costs can be accurately estimated and sustainability as well as fairness ensured.

Posting of member-owned resources to the Tools for Teachers application will be approved by a designee for each member. The designee will ensure adherence to quality, accessibility,

copyright, and bias and sensitivity guidelines. The designee will only be allowed to approve the posting of resources for their own specific tenant. Smarter Balanced will share knowledge, materials, and practices for the established process for quality assurance, accessibility, and copyright for member consideration. Smarter Balanced will provide training to the member designee on how to use the tenant-specific administrative tab in TIMS and how to sync approved tenant-specific resources in Tools for Teachers for posting.

The final step in the process is the viewing of member-owned resources and access to them by educators. Members will manage the associated quality assurance of the rendering of member-owned resources. Members will work with their service providers to manage SSO user accounts and access for educators. Smarter Balanced will maintain the Tools for Teachers website and ensure that the functionality of the application allows for the rendering of member-owned resources. Smarter Balanced will also maintain the existing access model.

Should a member wish to share resources with another member, or members, they may enter into an agreement to do so without the involvement or approval of Smarter Balanced. Members who agree to share resources with another member will indicate the state/territory with which the resources should be shared in TIMS. A sync will be required to ensure the resource is live for the tenant with which it's being shared.

## RESOURCES

- Tools for Teachers Overview
- *Tools for Teachers Help Desk Guide*
- *The Formative Assessment Process*
- *Instructional Resource Quality Criteria*
- *TIMS Resources Authoring Technical Guide*

# APPENDIX A: SAMPLE—MEMBER ADMINISTRATION CHECKLIST—ONLINE SUMMATIVE ASSESSMENTS

| Member User (Member) Example Activities | Notes/Resources |
|---|---|
| Determine what, if any, paper needs your organization (state, bureau, territory) has and work with Smarter Balanced to retrieve the materials and then work with your service provider to prepare for paper administration. | Appendix B, Section 7.0 |
| Address all policy issues in your organization around technology including developing a system in which to provide technical support for users. | Device Requirements and Approved Browsers on www.smarterbalanced.org |
| Address all policy issues in your organization around accessibility and accommodations. | Accessibility and Accommodations page on www.smarterbalanced.org |
| Designate user roles and permissions for your organization and configure the test registration tool for these roles. | Test Registration Tool User Guide Template |
| Identify roles and establish communication lines with service providers, District Test Coordinators, and others around breach procedures. | Section 4.0 of the Online TAM (Secure File Transfer Server) |
| Customize the Online TAM to reflect the decisions and needs of your organization, such as contact information, help desk details, certification requirements, user roles, timelines, etc. | Online TAM (Secure File Transfer Server) |
| Establish quality control procedures for implementing Smarter Balanced open-source software or certification protocols. | SmarterApp.org |

| Member User (Member) Example Activities | Notes/Resources |
|---|---|
| Review training modules to identify areas where information needs to be adapted to your organization's policy(s). | ([Secure File Transfer Server](#)) |
| On an ongoing basis, share and communicate important information to districts/schools that they need to prepare for administration, such as when the training modules are available, when the administration materials are available, where they are posted, and timelines and calendars. | |
| Communicate an organization-level policy with your districts around required training expectations. | |
| Investigate and make a decision on all test security incidents reported in the appeals system. | |
| Monitor and review all test security incidents submitted within your organization and make an appropriate appeal in the appeals system as necessary. | |
| Review test security breaches immediately. | |
| Monitor and review appeals as they come in during test administration. | |

Note: This list is not intended to be a comprehensive list or to cover all the steps members need to take to administer the summative assessments. It will be amended and updated over time.

# APPENDIX B: ONLINE SYSTEM – APPEAL TYPES AND CONDITIONS FOR USE

| Type of Appeal | Description | Conditions for Use |
|---|---|---|
| RESET | Resetting a student's test removes that test from the system and enables the student to start a new test. | The member may reset any test - Computer Adaptive Test (CAT) or performance task (PT) - if any of the following settings need to be changed because they were incorrectly set: <br><br> • American Sign Language (for mathematics and ELA/literacy listening) <br> • Braille (braille is a presentation setting) <br> • Closed captioning (for ELA/literacy listening stimuli) <br> • Streamlined interface (the streamlined interface is a test shell setting) <br> • Dual Language Presentation (for mathematics tests only) <br><br> Note: Dual language translations are automatically provided when the selected language or presentation is Spanish. <br><br> • Translations—glossaries (for mathematics tests only) <br> • Text-to-speech as an accommodation (e.g., for ELA/literacy reading passages in grades 3–8 or high school) <br><br> Note: Text-to-speech for items only is a designated support and NOT a valid reset request. <br><br> • Any non-embedded accommodation(s) <br><br> Log as Testing Irregularity: YES. <br><br> Additional Testing Irregularity considerations: <br><br> CAT: |

| Type of Appeal | Description | Conditions for Use |
|---|---|---|
| | | 1. Student has been presented with five or fewer items.<br><br>Log as Testing Irregularity: NO.<br><br>2. Student has been presented with more than five items.<br><br>Log as Testing Irregularity: YES.<br><br>PT:<br><br>1. Student has been presented with any items in a performance task.<br><br>Log as Testing Irregularity: YES. |
| INVALIDATION | Invalidating a student's test deletes the test. The test will not be scored. | The member may invalidate any test (CAT or PT) if:<br><br>1. There is a test security breach.<br><br>Log as Test Breach: YES.<br><br>2. The test is administered in a manner inconsistent with the Online Summative Test Administration Manual (TAM).<br><br>Log as Testing Irregularity: YES.<br><br>3. There is a testing session in which a student deliberately does not attempt to respond appropriately to items.<br><br>Log as Testing Irregularity: YES.<br><br>Note 1: Invalidated tests will not be scored.<br><br>Note 2: After reviewing the circumstances of an incident, a member may elect only to document the above conditions as testing irregularities, rather than to invalidate the test. |

| Type of Appeal | Description | Conditions for Use |
|---|---|---|
| REOPEN | Reopening a test allows a student to access a test that has already been submitted or has expired.<br><br>If an expired test is reopened, the test will reopen at the location at which the student stopped the assessment.[3] The student will be able to review items within the current segment of the assessment but cannot return to previous segments.<br><br>If a submitted test is reopened, the test will reopen at the last page of the test. The student can review items in the current segment but cannot return to previous segments. | The member may reopen any test (CAT or PT) if:<br><br>1. A student is unable to complete a test due to a technological difficulty that results in the expiration of the test.<br><br>Log as Testing Irregularity: YES.<br><br>The member may reopen a CAT if:<br><br>1. A student is unable to complete the test before it expires (45 days) due to an unanticipated excused absence or unanticipated school closure or technology issues.<br><br>Log as Testing Irregularity: YES.<br><br>The member may reopen a PT if:<br><br>1. A student is unable to complete the test before it expires (10 days) due to an unanticipated excused absence or unanticipated school closure or technology issues.<br><br>Log as Testing Irregularity: YES.<br><br>2. A student starts a PT unintentionally—for example, selects a PT instead of a CAT, or selects a mathematics PT instead of an ELA/literacy PT—and the student is unable to complete the test before it expires (10 days).<br><br>Log as Testing Irregularity: YES. |

[3] A test that is reopened following an expiration will remain open for ten calendar days from the date it was reopened.

| Type of Appeal | Description | Conditions for Use |
|---|---|---|
| | | 3. A student unintentionally submits a test before he or she has completed it—for example, a student submits the ELA/literacy PT before completing Part 2.<br><br>Log as Testing Irregularity: NO. |
| RESTORE | Restoring a test returns a test from the Reset status to its prior status. This action can only be performed on tests that have been reset. | The member may only restore a test if a test was inadvertently or inappropriately reset. |

Note: Members may allow for additional appeals (e.g., Grace Period Extension)

# APPENDIX C: TEST SECURITY CHART

The test security chart shows the test security incident levels and examples of types of issues.

| Level of Severity & Potential Effect on Test Security | Types of Issues |
|---|---|
| **LOW** <br><br> Impropriety | Student(s) making distracting gestures/sounds or talking during the test session that creates a disruption in the test session for other students. |
| | Student(s) leave the test room without authorization. |
| | Disruptions to a test session such as a fire drill, school-wide power outage, earthquake, or other acts. |
| **MEDIUM** <br><br> Irregularity | Student(s) cheating or providing answers to each other, including passing notes, giving help to other students during testing, or using hand-held electronic devices to exchange information. Student(s) accessing the Internet or any unauthorized software or applications during testing. |
| | Student(s) accessing or using unauthorized electronic equipment (e.g., cell phones, smart watches, PDAs, iPods, or electronic translators) during testing. |
| | Administrator or Coordinator leaving related instructional materials on the walls in the testing room. |
| | Administrator or Coordinator failing to ensure administration and supervision of the Smarter Balanced assessments by qualified, trained personnel. |
| | Administrator giving incorrect instructions that are not corrected prior to testing. |
| | Administrator or Coordinator giving out his or her username/password (via email or otherwise), including to other authorized users. |

| Level of Severity & Potential Effect on Test Security | Types of Issues |
|---|---|
| | Administrator allowing students to continue testing beyond the close of the testing window. |
| | Administrator or teacher coaching or providing any other type of assistance to students that may affect their responses. This includes both verbal cues (e.g., interpreting, explaining, or paraphrasing the test items or prompts) and nonverbal cues (e.g., voice inflection, pointing, or nodding head) to the correct answer. This also includes leading students through instructional strategies such as think-aloud, asking students to point to the correct answer or otherwise identify the source of their answer, or requiring students to show their work. |
| | Administrator providing students with non-allowable materials or devices (e.g., calculators during non-calculator sections) during test administration or allowing inappropriate designated supports and/or accommodations during test administration. |
| | Administrator allowing anyone other than a student to log in to the test unless prescribed as an allowable accommodation in the student's Individualized Education Program (IEP). This includes Test Administrators (TAs) or other staff using student information to log in or allowing a student to log in using another student's information. |
| | Administrator providing a student access to another student's work/responses. |
| HIGH<br><br>Breach | Administrator or Coordinator modifying student responses or records at any time. |
| | The live Student Interface or TA Interface being used for practice instead of the Interim, Training or Practice Tests. |
| | Adult or student posting items or test materials on social media (Twitter, Facebook, etc.). |

| Level of Severity & Potential Effect on Test Security | Types of Issues |
|---|---|
| | Administrator allowing students to take home printed test items, reading passages, writing prompts, or scratch paper that was used during the test or failing to otherwise securely store test materials. |
| | Adult or student copying, discussing, or otherwise retaining test items, reading passages, writing prompts, or answers for any reason. This includes the use of photocopiers or digital, electronic, or manual devices to record or communicate a test item. This also includes using secure test items, modified secure test items, reading passages, writing prompts, or answer keys for instructional purposes. |
| | Secure test materials being shared with the media (such as the writing prompts, test items, or reading passages), or allowing the media to observe a secure test administration. |
| | Adult or student improperly removing secure testing materials such as test items, stimuli, reading passages, writing prompts, or scratch paper from the testing environment. |

Other Documentation

- *Online Summative Test Administration Manual* (Customizable version available in the Assessment Training and Operations folder of the Secure File Transfer Server)
- *Test Administrator User Guide Template*
- *Online Training Modules* (Assessment Training and Operations/Training Modules folder in the Secure File Transfer Server)

# APPENDIX D: GUIDANCE FOR SOCIAL MEDIA SUMMATIVE ASSESSMENT MONITORING

## SMARTER BALANCED TEST SECURITY

Maintaining test security during administration of the summative assessment is critical to preserving the integrity of test items and validity of the test itself. These guidelines provide recommendations for monitoring social media.

## TEST ADMINISTRATION PROCEDURES

It is important to be vigilant before, during, and after testing for any situations that could lead to or be an impropriety, irregularity, or breach. Please remember that only individuals who have been appropriately trained and whose presence is required may be present during the administration of the summative assessment.

To get ahead of the problem and reduce the number of security breaches on social media, Smarter Balanced encourages members to refer to Appendix D: Test Security Chart for detailed information on the impact and definition of incidents as well as the timeline for reporting these activities.

## SOCIAL MEDIA MONITORING

Smarter Balanced and members have a vested interest in ensuring that assessments are supported by security protocols that establish both fairness for student engagement and validity in the interpretation of results. Maintaining test security during administration of the summative assessment, especially during remote administration, is critical to preserving the integrity of the test.

Educators and policymakers need valid information about student performance to improve schools, and students deserve to have real information about what they know and can do related to the assessed content. When some students see questions in advance, the test results don't have the same meaning for all students. Test security has always been a concern, and states have addressed these issues since they first began administering standardized tests. The biggest differences are the new landscape of social media—instead of copying tests, students post questions online - and remote test administration. During the 2022-23 test administration, there were regular social media posts regarding the summative assessment from students in a remote testing environment.

Smarter Balanced works with a major test security company to monitor the Web and social media, including Twitter, Instagram, Tik Tok, and Facebook. Members are encouraged to perform their own monitoring. If Smarter Balanced staff confirm a breach, they will:

- Post the image from social media to the Operational Test Security folder of the [Secure File Transfer Server](#);

- Provide as much information as possible, including the name and location of the individual who posted the breach; and

- Notify the appropriate K-12 Lead via email.

**Member State Responsibilities**

- Contact the district/school where the breach occurred.

- Request verification that the image was posted by a student from the district/school.

- Once verified, request that the image of the item be taken down.

- Alert Smarter Balanced when the item has been removed.

When a member finds a potential security breach, SEA staff will:

- Contact the district/school where the breach occurred;

- Request verification that the image was posted by a student from the district/school;

- Post the image from social media to the Operational Test Security folder of the [Secure File Transfer Server](#);

- Request that the image of the item be taken down; and

- Notify Smarter Balanced ([sb@smarterbalanced.org](mailto:sb@smarterbalanced.org)) of the resolution, providing as much information as possible, including whether the item is from the interim or summative assessment.

## TEST ADMINISTRATION PROCEDURES

States and districts may choose to monitor social media for potential security breaches. Smarter Balanced recommends the following:

### Twitter, TikTok, Instagram, and Facebook

- If schools in your state or territory do not have social media accounts, Smarter Balanced recommends they create them to help monitor possible test security issues.

- Following [@SmarterBalanced](#) will also help you to monitor the Smarter Balanced news feed.

- To search for conversations and posts about the summative assessment, consider the following search queries:

    - #sbac #smarterbalanced #commoncore #ccss
    - #[insert name of school] or @[insert school Twitter handle]

– "smarter balance(d)" "sbac" and any acronym used for your state test

## What to Look for

- Images of the computer screen that show ELA/literacy or mathematics test items.

- Any photographs that appear to have been taken in the test administration room; these can be images  students have taken of themselves or their classmates as well as pictures taken by test administrators of the testing session.

- Tweets that indicate test security policies are not being upheld.

## New Technologies

Evolving technology can complicate test security. Members should amend their test security guidance to ensure new devices such as smartwatches do not compromise the security of test items.

# APPENDIX E: ITEM RISK RUBRIC

## SMARTER BALANCED ITEM EXPOSURE RISK ANALYSIS

Smarter Balanced members have an individual and collective interest in maintaining the security of the summative assessment items. Members may use the following rubric to evaluate the risk that item exposure creates.

A greater degree of exposure and/or a longer duration of exposure increase the threat to the validity of the test.

| Exposure Duration | Levels of Risk |
|---|---|
| 1. Degree of cumulative exposure | a. Newspaper/mainstream media – High Risk<br>b. Social Media – Medium Risk<br>c. Local Exposure - Medium Risk |
| 2. Duration of exposure | a. Permanent (e.g., paper exposed) – High Risk<br>b. Moderate (online, but removed within 24 hours) – Medium Risk |

The nature of content that is exposed will help to determine the threat to the validity of the test. Content that isn't accessible to a large portion of the student population and/or otherwise isn't likely to be present on their tests constitutes a lower risk to the tests' validity.

| Exposure Scope | Levels of Risk |
|---|---|
| 1. What type/portion of the item is exposed? | a. All information is self-contained – High Risk<br>b. Items that have an External Reference – High Risk<br>c. Stimuli – Medium Risk |
| 2. What is the likelihood students will see the exposed content on their tests? | a. Likely – High Risk<br>b. Unlikely – Medium Risk |
| 3. Presentation of item | a. English – High Risk<br>b. Language other than English – Medium Risk<br>c. Braille and ASL –Medium Risk |

Smarter Balanced will collaborate with members and service providers to determine the action members need to take. The tests that the Consortium will apply are described by the following questions:

- Will the validity of the test be improved based on removing the item from the pool?
  - If the item is highly exposed but unlikely to be seen by a student, the risk of interacting with the item pool may not be offset by the reduction of risk to validity.
  - If the item measures a unique dimension of the construct (either via content or difficulty) the validity of the test could be reduced by eliminating the item.
- Will public perception of the test be improved by the action?
  - The Consortium needs to take a strong stance on test security.
  - Removing exposed items may be necessary to retain public confidence.

If the Consortium determines that an action is required, the Consortium will notify the members' designated service providers, the member project managers identified on the MOU, and the K-12 Lead of the item ID that must be removed from the pool. Service providers will need to remove the item within 1 business day.

Unless otherwise determined by the Consortium on a case-by-case basis, members do not need to re-score tests for students who respond to an item exposed between the time the item was exposed and when the item was removed from the service providers' systems.

# APPENDIX F: LARGE-SCALE SCORING: INVOLVING TEACHERS AND PRODUCING RESULTS

**Marianne Perie, University of Kansas**

*A report prepared March 27, 2014 for the Smarter Balanced Assessment Consortium*

The transition of assessments toward college and career readiness and adoption of the Common Core Member Standards requires students to analyze text, show their thinking processes, and write out responses. Performance tasks, short constructed-responses, and written essays are once again becoming an important component of member testing. This transition has caused members to increase the use of hand scoring far beyond recent practice; however, those with a longer history in testing may remember a time in the 1990s when open-ended responses requiring hand scoring was commonplace.

The Smarter Balanced Assessment Consortium provides assessments aligned to the Common Core Member Standards but requires members to administer and score the assessments and return the results to the consortium for analysis. Thus, the onus for scoring assessments rests squarely on the members. The current assessment design of Smarter Balanced includes one performance task (including an extended written response) and three short constructed responses for English language arts (ELA), and one performance task and one short constructed response for mathematics. Each student, therefore, will have six written responses that need to be hand scored.

In early plans, both Smarter Balanced and the Partnership for Assessment of Readiness for College and Careers (PARCC) had hoped to use automated scoring technology. Automatic scoring is done entirely by computers aggregating quantifiable text features to evaluate the quality of an essay. These systems work exclusively with variables that can be extracted and combined mathematically, and must be programmed to do so (Zhang, 2013). This programming involves human judgment, but the implementation is automated. A sample of human-scored responses are used to determine which text features best predict the scores. Scores for new responses are then based on the scoring algorithm from the original sample. Thus, automated scoring focuses on summing individual pieces, while human scorers can make holistic decisions based on many interacting factors. Research reveals "the primary strength of automated scoring compared to human scoring lies in its efficiency, absolute consistency in applying the same evaluation criteria across essay submissions and over time, as well as its ability to provide fine-grained, instantaneous feedback," (Zhang, 2013, p. 3). Once the prediction equation for a particular essay prompt has been determined, automated scoring can achieve greater objectivity than human scoring (Williamson, Bejar, & Hone, 1999) and generate nearly real-time performance feedback on various dimensions of writing.

In spite of all its advantages, automated scoring has rarely been used in a large-scale, K–12 assessment of ELA.[4] There are still issues that concern researchers and stakeholders. For example, automated scoring cannot distinguish nuances or intent, or detect relatively uncommon errors. It is also not yet ready to assess creativity, logical progression, quality of ideas, or any other unquantifiable feature. Additionally, if schools receive information about how a scoring algorithm functions, it opens the possibility for coaching students to achieve higher scores than their writing deserves.[5] At this time, both Smarter Balanced and PARCC have decided to start with human scoring as automated scoring continues to be researched and refined.

But human scoring is not simply a fallback position. Bringing in educators to score student work can benefit both the validity of the assessment scores and the professional development of the educators. The challenge is to balance the need for reliable and accurate scores with the desire to provide a positive experience for teacher scorers without significantly increasing cost. The goal of this paper is to summarize the literature on best practices for scoring open-ended responses. Multiple options for human scoring will be described, and the paper will conclude with recommendations for Smarter Balanced members to consider moving forward.

## INCORPORATING A SCORING MODEL THAT INVOLVES EDUCATORS

Developing a scoring model for a large-scale member assessment that involves educators can have many benefits, if done well. Ultimately, training on and application of scoring methods can provide educators with a chance to deepen their understanding of content standards, as well as improve their assessment literacy. During a well-designed process that includes training on content standards, tasks, rubrics, and sample student work, educators learn more about the testing process, and about interpreting and applying content standards. Comparing exemplar and typical responses provides educators with information about teaching methods that can help students avoid common errors and produce responses that best demonstrate their knowledge and understanding.

> In the course of scoring, teachers learn to apply common criteria and standards to the work of all their students, rather than just comparing one student's work to another student's work. Learning to use evidence as a result of participating in standards-based scoring often transforms the way teachers evaluate student work. (Darling-Hammond & Falk, 2013, p. 7)

---

[4] Both West Virginia and Louisiana have used automated essay scoring in their member programs.

[5] For example, in a recent competition, an algorithm that merely counted the number of commas in a response was almost as accurate as human scoring. Therefore, students who randomly inserted extra commas would have gotten higher scores than their writing deserved.

Additionally, involving current classroom teachers in the scoring process helps member policymakers create buy-in for new assessments and ownership of the resulting scores. Being involved in the evaluation process can also help dissipate educators' fears of outside testing companies judging their students' work.

It is important to keep in mind that all the benefits of involving educators in assessment scoring can only be achieved when the process is done well. A well-implemented process requires thorough training materials, a certification process, constant monitoring during scoring, and recalibrating raters throughout the scoring window. A poorly implemented process can cause a lack of comparability across scorers and scoring sites. For a project such as Smarter Balanced, a poorly implemented process in one member will decrease the comparability of results with other members in the Consortium.

The benefits of using human scoring are accompanied by financial costs. As part of the process for determining a scoring model the first year and subsequent years, these financial costs will need to be balanced against the benefits of providing professional development to educators through their involvement in scoring.

## RESEARCH AND PAST PRACTICES

Over the years, much research has been done on how to improve scoring. More recently, some studies have focused on the benefits teachers receive from scoring in terms of their classroom instruction. In addition, while some member testing programs use vendor-lead scoring, many others rely on teachers. Thus, members have learned best practices from those teacher scoring sessions.

### Research Findings

Effective training techniques for scoring procedures can also provide teachers with professional development. A good training session clarifies content standards, articulates scoring rubrics, provides concrete examples of student work at each score point, and examines different approaches to the same task. Training benefits are maximized in group settings where teachers can interact with one another and discuss each of the points of training. Darling-Hammond and Ancess (1994) found that reviewing and discussing student work with colleagues could help teachers improve their teaching practices and enhance their evaluation of student work in the classroom.

Studies of teachers involved in assessment scoring have found that teachers reported benefits related to increased knowledge of the scoring process. Goldberg (2012) reported that teachers tended to cite benefits such as clarification of standards, understanding multiple approaches that students take to a task, increased assessment literacy, and a better understanding of

desirable teaching practices. Other studies done on a Washington state writing assessment focused on how the gains in knowledge regarding the scoring process would translate to teaching practices. After participating in scoring, teachers indicated that, going forward, they would familiarize students with how to align their responses to a rubric, specifically by encouraging voice and expression in writing, and suggesting that students move away from the five-paragraph essay, emphasize purpose, and provide explicit details (Eliott-Schuman, 2005, 2006). Another study in Washington found a direct link between teacher participation in scoring and improvement of student scores. Specifically, a model where teachers were trained by a professional scorer and then trained other teachers in their own schools resulted in students in those schools showing greater score gains (Heuschel, 2004).

However, the benefits are not guaranteed. Research has also identified relevant factors for a successful scoring session. First, it is important to train teachers on the difference between grading and scoring (Jones & Vickers, 2011). Training on the rubrics with multiple sample papers helps with this goal, and it can be easier to do with an analytic scoring rubric than with a holistic one. In addition, when having teachers score sample papers (often called validity papers), the type of feedback given to the new scorers is critical. Teachers need annotated feedback on any inaccuracies, not just the "correct" score. Providing teachers with time to talk about any misperceptions will improve both scoring accuracy and teachers' professional development experiences.

Regardless of the benefits, there are still some concerns with using participation in assessment scoring as professional development. The first goal of scoring is to produce accurate, reliable scores of student work. Test results often have high stakes attached to them, and the validity of the scoring is of paramount importance. Policymakers are also concerned with reducing costs, when possible. Some of the less costly designs allow for less teacher interaction, which reduces the learning benefits for teachers from the scoring experience.

Goldberg (2012) found that the benefits teachers received from scoring participation might be overstated. Several of his studies showed that teachers struggled to apply lessons learned from scoring to classroom instruction (Goldberg & Roswell, 2000). Conversations with scoring coordinators at several testing companies involved in large-scale state assessments revealed that they see diminishing returns over time. There is a lot of discussion among teachers the first day and during the scoring of about the first 50 papers. However, discussion decreases as scoring continues. Scoring becomes more cost-effective as teachers become more experienced, but the more experienced teachers become, the fewer benefits they seem to receive from the process.

One of the testing companies, Data Recognition Corporation in Minnesota, indicated that it could cost more to involve teachers in the scoring process compared with using professional

scorers because of the time it takes to score each paper. In contrast, an article by the Assessment Solutions Group found a scoring cost of $125/day for a teacher, which is lower than the cost for a professional scorer (Topol, Olsen, & Roeber, 2013). Both the hourly rate and the number of papers scored in an hour need to be compared for each group. The estimated time for scoring decreases as scorers become more skilled, but efficiency will be realized more quickly with experienced scorers. Kingston (1997) found that fast scorers who took less than half the time of slow scorers had comparable accuracy rates.

## State Examples

In the 1990s, prior to the enactment of No Child Left Behind, many states were conducting performance assessments, including extended performance tasks, portfolios, and extended writing prompts. Connecticut, Kentucky, Maine, Maryland, New York, and Vermont involved students in classroom performance tasks of longer duration—from one class period to several—designed at the state level and administered and scored locally, with a moderated scoring process to ensure consistency. At this time, assessment results were not used punitively, either at the school or teacher level, so local scoring was more common. Some states conducted audits of randomly-selected districts during the summer to check the reliability and accuracy of scoring. The results of the audits were then used to improve the process.

Both Vermont and Kentucky had portfolio assessments in the 1990s that were scored locally by teachers. As described by Darling-Hammond and Rustique-Forrester (2005), teachers engaged in dialogue about what good work looked like, resulting in a shared definition of quality work across schools. Teachers working within schools were trained to score the portfolios using analytical rubrics. Two readers, using double-blind scoring, scored each piece. If scores did not match in any domain, a third reader would also score. By 2008, the agreement rate for independent readers who audited school-level scores was more than 90% (Conley & Darling-Hammond, 2013). By participating in scoring, teachers gained multiple perspectives on student learning and quality responses. They took advantage of the time together to discuss types of teaching approaches that helped elicit high quality responses.

Maryland implemented the Maryland State Performance Assessment Program in the 1990s. Students were engaged in performance tasks, usually in teams, that lasted an entire class period. The program employed centralized scoring with teacher volunteers. These teachers had the opportunity for professional interaction, specifically through conversation focused on learning targets. A study conducted by Goldberg (2012) found that participation in scoring led to an increased understanding of the standards, objectives, and learning targets and how to use that understanding to improve classroom instruction. He also found that these teachers were better able to implement their own classroom assessment more naturally into instruction. One

teacher claimed that participating in scoring was an enduring type of professional development, because it involved "seeing through the students' eyes" (Goldberg, 2012, p. 44).

Currently, teachers in New York score the open-ended items of the Regents examination, including the writing portion. The state allows two of the four built-in professional development days to be used for scoring. The state is in the process of changing the scoring model. In the past, scoring has been done at the building or regional level, depending on the density of the location. However, because the results will now be used for teacher evaluation, in 2014 all schools are required to move to a regional scoring model. A pilot of the regional scoring model was conducted in 2012–2013. A trained site supervisor led each regional scoring site, providing deeper training and improved norming practices to ensure high-quality scoring. They found that working together across schools allowed for greater collaboration and problem-solving. Some challenges they encountered included finding sufficient numbers of trained site supervisors to cover the state, standardizing the training across all of the sites, and providing sufficient specialized expertise at specific sites. For example, some regions needed scorers fluent in other languages or skilled in more complex science subjects. In the future, New York hopes to move to a distributed scoring model as technology becomes more prevalent across the state.

Oregon includes local performance assessment as part of its assessment program, and also has a statewide writing test. The local performance assessments are scored locally using state-provided scoring guides. For the writing assessment, Oregon uses volunteer teachers as scorers and either compensates them financially for their time or provides them with continuing education credit. The state uses a hybrid model of distributed scoring and centralized scoring. One of the challenges the state faces is how to balance using new teachers each year for centralized scoring to share the learning opportunity, while retaining trained teachers to increase efficiency.

Pennsylvania currently administers a performance assessment as the alternate assessment based on alternate achievement standards for students with the most significant cognitive disabilities. Each test consists of 25 performance tasks that are scored based on a rubric. A teacher videotapes a student engaging in the tasks and sends the video to the state for centralized scoring. Scoring centers across the state are staffed by educators and employees of the University of Pittsburgh. Each video is scored by a two-person team of educators who discuss each task and reach consensus on the score. If they cannot reach consensus, they call for a third-party expert. Because the rubrics include features of student-teacher interaction, teachers learn about the types of mistakes test administrators make that can lower a student's score. Thus, the greatest professional development for teachers is learning how to best administer tasks to their own students to increase their probability of success.

New Hampshire will be integrating performance assessment with the Smarter Balanced assessment to create high school portfolios that will count toward graduation. The state indicated it would create regional support networks led by assessment experts to help build capacity in schools and to support regional task validation and calibration scoring sessions, with a goal of 80% or greater inter-rater reliability[6] on locally scored tasks (Conley & Darling-Hammond, 2013).

In addition to these examples, multiple states administer writing tests—typically in three grades per year—that are scored by hand. In some of these states, the testing vendor conducts the scoring using trained scorers, and in others teachers are brought to a centralized scoring facility and trained by the vendor. Additionally, West Virginia currently uses an automated essay scoring model. Human scorers provide the training papers and are used to validate the scores assigned to a portion of the writing prompts by the automated engine (Rich, Harrington, Kim, & West, 2008).

## Current Best Practices

To investigate current best practices for scoring, interviews were conducted with several testing vendors, and research articles produced by those vendors were analyzed. Most of the larger assessment vendors have developed a system of distributed scoring where a computer assigns papers across multiple scorers, scorers review papers and enter their scores online, and statistics are calculated in real time. Training and validation can also be conducted online. Each vendor has their own set of practices and recommendations, three of which are outlined in this section.

Measured Progress uses both distributed scoring and scoring centers.[7] For their distributed scoring models, they created training modules with voiceover to standardize the training that all teacher scorers receive. In addition, they find that using the same anchor papers, practice papers, qualification papers, and validation papers across all scores also helps with standardization and ensures reliability of scores across all scorers working remotely. Teachers must be certified through Measured Progress's program before they can score operationally. During scoring, Measured Progress uses text chats, real-time monitoring, and live interventions to individualize training and ensure accuracy in scoring. They prefer to use double-scoring, also known as 100% read behind, as this technique increases the reliability of scoring and provides greater opportunity for a lead scorer to intervene if

---

[6] The goal was defined by an agreement rate calculated as the percentage of exact plus adjacent scores.

[7] Information on Measured Progress was obtained in an interview with Scott Hinders and Karin Evans, November 5, 2013.

reliability checks show that a particular teacher tends to score either more leniently or strictly than the norm.

One example of using distributed scoring that resulted in strong professional development for teachers was in a pilot program where teachers developed performance tasks aligned with the Common Core State Standards. The teachers built the rubrics, were trained by Measured Progress, and then scored the responses. Multiple classrooms within a school worked together to increase collaboration and reduce the burden on any one teacher. In one case, there were 75 students and 8 teachers participating. Even with each paper being scored twice, each teacher only had to score about 20 papers. The school enabled group training and discussion of results, which allowed for professional dialogue and resulted in a rich learning experience for the teachers. Measured Progress also facilitated webinars across schools to allow teachers to connect with each other and share lessons learned from the experience.

The Data Recognition Corporation (DRC) has also conducted distributed scoring, although they indicated they find scoring centers to be more efficient.[8] One conflict they noted with trying to conduct an efficient, accurate, and reliable scoring process while providing rich professional development for teachers centered on the number of different tasks scored. To increase efficiency, scorers tend to be trained on one task and continually score that same task across multiple students. However, teachers will be better served by scoring every task for a fewer number of students. Thus, professional raters tend to focus on one task over multiple papers; teachers on one paper with multiple tasks. It is more efficient if each rater scores fewer items, but it doesn't provide the rich professional development that scoring all the items gives.

DRC provided an example that used 900 teachers for each subject to score open-ended items. These items included short constructed responses, extended written essays, and "show-your-work" math tasks. The teachers were trained at a centralized scoring facility and then spent two weeks scoring all the open-ended items. The training lasted approximately a day and a half. The first day of scoring was spent answering technical questions and ensuring teachers understood the process. DRC noted that the second day of scoring brought multiple "aha!" moments for teachers as they saw the types of errors students made and discovered patterns in responses. However, during the next 10 days of scoring, there seemed to be diminishing returns on the amount learned from scoring the items. The patterns seen in the first 50 papers did not change significantly while scoring additional papers.

---

[8] Information on DRC was obtained in an interview with Sherri Woolf, November 6, 2013.

Although grading additional items may provide new insights for teachers, at a certain point the emphasis moves to scoring accurately without drift. Greater professional development occurs by allowing teachers to confer, so pairing teachers is particularly effective. Also, having teachers with specific expertise, such as special education, pair with a content expert or be available to teams of content experts enhances the experience by drawing on multiple backgrounds.

A couple of concerns noted by DRC include the awkwardness of having to dismiss a teacher who does not qualify, even after repeated training, or one who consistently scores validation papers incorrectly. Additionally, DRC has experienced resistance from teachers who found the scoring process tedious after the first 50 papers. They recommend scoring in pairs to increase collaboration and reduce the drudgery. They also note that holistic scoring goes faster than analytic scoring, but analytic scoring may be clearer, which reduces the difficulty of training and scoring. Another recommendation is to consider working within schools, allowing collaboration across teachers who continually work together. An example might be asking an elementary school with 200 students to score 200 papers from other schools. Teachers can work as teams within the school to increase the benefits of the experience.

Although Pearson is a testing vendor that touts the benefits of automated scoring, they also use centralized and distributed scoring to meet the needs of their clients. Pearson published a paper by Jones and Vickers (2011) that lists one benefit of distributed scoring as promoting inclusion and diversity in the scoring pool, resulting in a large scoring pool that can increase turnaround time and reduce the number of papers any individual teacher has to score. Some of Pearson's recommendations parallel those of other vendors:

1. Score at the item level, not the booklet level, to increase efficiency and reduce the chance of a halo effect across the items within a booklet.
2. Supplement the scoring pool of current teachers with retired teachers, substitute teachers, and qualified scorers.
3. To increase professional development benefits, use regional scoring when there are only a few papers to score (e.g., alternate assessment portfolios), and local scoring (within a school) for formative or classroom assessments.

Pearson also raises the issue of whether teacher scoring should be voluntary or mandated, and whether teachers should be paid for their time or count it as continuing education or professional development. They note that unions may take issue with mandating an additional task without adequate compensation.

# MODELS FOR HAND SCORING

Several models for hand scoring have been described in the review of best practices. This section outlines the descriptions, benefits, and challenges of each model to help members decide which one is best for their assessment program.

## Centralized Scoring

Centralized scoring requires that a selected group of educators meet in a central location, receive training, and score all papers across an agency, state or territory. This model has typically been used during the past 20 years. Although it is expensive, the model produces results in a known timeframe, allows for daily review of statistics, and supports immediate retraining. All experts can be assembled in one place to work directly with the scorers. Fewer scorers are involved, but those participating can learn a lot about the scoring process and have in-depth discussions about the construct and scoring rubrics. Thus, this method is highly interactive.

In some cases, real papers are used for scoring, but in others, image scoring is used. This occurs when student responses are scanned and viewed on a computer. Image scoring has logistical advantages, including easy randomization, prevention of lost or damaged papers, and initial scores remain hidden during rescoring.

## Distributed Scoring

Distributed scoring is an Internet-based scoring strategy that uses an electronic platform to assign computer-entered or imaged papers to scorers, enabling them to score papers anywhere they have an Internet connection. As mentioned earlier, many programs have the capability of training scorers online, certifying them and continually checking for scorer drift through the insertion of validation papers. Measured Progress described a system in which they intervene individually and retrain those scorers who are incorrectly scoring validation papers.

Benefits of distributed scoring include that it can be done much less expensively than any scoring that requires a central location; it can involve more educators, as the only requirement is an Internet connection; and it is flexible in that scorers can score any time day or night during the scoring window. Because more teachers can be involved in distributed scoring, each teacher should ultimately have to score fewer papers. Training and validation can be easily seeded throughout the process with ongoing monitoring and correction for score drift.

On the other hand, distributed scoring is the least interactive of the scoring models. Teachers lose the opportunity to interact with one another when they score alone. Although teachers will benefit from the training, increase their assessment literacy, and learn more about implementing rubrics, they lose the benefits of discussing these features with others. Scoring can become an isolating activity rather than one that forms connections and enhances understanding.

## Teacher Scoring with Distributed Read Behind

Another model allows teachers to score their own students' work in addition to having responses uploaded to a central site for scoring by others through distributed scoring. In addition to scoring their own students' work, teachers agree to score an additional number of anonymous papers from other schools. Members can decide what percentage of read behind they want to implement. One option is to vary the percentage as time goes on, using a higher percentage initially until there is sufficient data to validate each scorer's accuracy. Then the read behind rate for each scorer can be lowered. The read behind percentage determines how many papers teachers would need to score through the distributed scoring system.

One benefit of this approach is that teachers gain instant feedback from seeing how their own students responded to the test questions. Being able to reflect on mistakes their students made and reviewing the approaches students took should provide some professional development, particularly if teachers have a chance to discuss their reflections with other teachers in their school. Teachers will gain additional understanding by viewing the work of other students outside of their schools. The distributed scoring should help teachers distinguish common approaches or mistakes across the agency, state or territory from ones specific to their classes. From a statistical standpoint, the distributed scoring read behind should detect any bias effect for any individual teacher, allowing for adjustments. This approach should have relatively low costs for members.

Some concerns with this approach include the fear that the students' teachers will be biased. This bias could be based on the teacher's knowledge of the students or it could be an overall bias to score high if the scores are used in teacher evaluation. If a bias is found, additional scoring will be needed and thus increase the burden on all teachers.

## School Scoring with Distributed Read Behind

In this model, staff members within a school work together as content teams scoring collectively. Papers are uploaded and a 20% read behind is done to check for any systematic bias. In schools where students enter their work online, the papers can be made anonymous. Staff can be trained through an online training program, by a central staff person, or using a train-the-trainer model resulting in one scoring expert in each school. The same type of distributed scoring read behind is employed as in the teacher scoring with distributed read behind model.

Having teachers work collaboratively in content teams within a school provides great professional development, and reviewing how students from their schools did on the items provides important information to teachers. Additionally, comparing approaches and errors made by their students to those made by students from other schools provides teachers with a

valuable perspective. Even the distributed scoring element can be done as a team, with teachers working together around a computer. However, this approach is only successful in larger schools. For example, if a school has only one science teacher, then the better approach is to have several schools team together.

If a professional development day is used for scoring, members could save money, as they would not have to pay the teachers to score. However, this approach can be time-consuming. Depending on the size of the school and the number of items requiring hand scoring, the process can take several days. It may not be feasible to do the distributed scoring as a team, in which case teachers will learn less from that portion of the task.

### Regional Scoring

In regional scoring, staff members from multiple schools gather at a scoring site central to their geographic locations and collectively score exams. Training is typically done by an expert scorer who certifies teachers before they score operationally. This can either be an outside expert or someone within the region who has attended additional training sessions. Validation papers are seeded within the operational papers as in a standard scoring process, and mid-course corrections are made as needed.

As with school scoring, having teachers from the same region working collaboratively to understand the standards, interpret the rubrics, and learn about the scoring process provides good professional development. Once papers have been entered or uploaded to an online environment, they can be made anonymous to prevent any competition among schools within the region. Even if all the papers are randomized so that teachers do not know which papers are from which district, working together to identify common themes and trends can help teachers improve their teaching practices.

As with school scoring, regional scoring can be time-consuming. If professional development days are assigned for this scoring task, the cost will be reduced, but it may take several days to score all papers. More time is typically spent on standardized training in this model than in the previous two. Although standardization is important in all contexts, it is difficult to add validity papers at the class or school level without it being obvious that they are validity papers, which could negatively affect their utility.

### Moderated Scoring within Schools

Moderated scoring within schools is a more formal protocol for doing distributed scoring in teams. Committees are formed within schools, either by content expertise or by purposefully combining multiple teacher backgrounds, such as math and special education teachers. These committees must first be trained and qualify together, and then work together to score each

paper. Training is typically conducted through an online training module, although a train-the-trainer model can be used if district-wide training is held with each school sending a representative. The committees enter scores online as they score each paper so reliability can be monitored in real time. Likewise, validity papers can be seeded throughout the process so that the scoring accuracy of committees can be monitored, and interventions can be applied if drift is observed. The papers scored by each committee could be within the school or from anywhere in the member, but all identifying information is removed.

As with similar models, working in committees provides teachers with the interaction needed to turn the scoring process into a rich professional development activity. The moderation activities ensure standardization across scores and promote accurate and reliable scoring.

Again, as with school and regional scoring, the moderated scoring can take time. Providing professional development days for this activity will help cut costs, but it will be important to look at the number of educators per committee, the number of committees, and the number of student responses that must be scored.

## Hybrid Models

Many of these approaches can be combined, particularly if members want all student responses to be double-scored. For instance, if teachers are each assigned 75 essays to score, they could receive training as a group and score the first 30 essays as teams, and then complete the last 45 on their own through a distributed scoring approach.

Likewise, teachers could be brought to a regional or central location for training and then score a day's worth of papers face-to-face with other colleagues. Afterward, teachers can continue to score essays online, at which point they could be paid by the hour or by the essay.

Another approach is to have all teachers score a small number of papers (50 or fewer) as a group, either within a school or larger group, to receive professional development experience. From that initial pool, scoring coordinators would select some teachers to be certified as expert scorers who would then score more papers and be paid for their effort. In subsequent years, those expert scorers could also conduct trainings themselves in addition to scoring on their own in a distributed scoring model.

Table 16: Summary of Available Scoring Models

| Model Name | Quick Description | Benefits | Considerations | Best for |
|---|---|---|---|---|
| Centralized scoring | Selected committee of educators or professional scorers meeting in a central location to score all papers | • Highly standardized<br>• Chance for interaction among committee members | • Costly to rent space<br>• Few educators receive professional development benefits | Members with low connectivity, a pre-existing pool of trained scorers, and less interest in providing professional development for a large number of teachers |
| Distributed scoring | Training, scoring, read behind, validity, and reliability all done on computer, usually from remote locations | • Inexpensive<br>• Open to any educator who has a computer with an Internet connection<br>• All teachers can receive training | • Little to no opportunity for interaction among teachers<br>• Requires a large number of teachers | Members with high connectivity but few resources to pay for face-to-face meetings |
| Teacher scoring with distributed read behind | Teachers score their own students' work in addition to imaging it; read behind is done | • Good professional development for teachers | • Time-consuming<br>• Concern of teacher bias | Members with professional development days built into the school schedule and |

| Model Name | Quick Description | Benefits | Considerations | Best for |
|---|---|---|---|---|
| | through distributed scoring randomly assigned to teachers | to see their own students' work as well as the work of students from other schools<br>• Low cost | • Requires every teacher to participate | that want to improve the assessment literacy of their teachers; good for lower-stakes tests, or tests without consequences to teachers |
| School scoring with distributed read behind | Educators within a school gather in content teams and score collectively; papers are imaged for distributed read behind | • Good professional development by allowing teachers to work in teams<br>• Low cost | • Time-consuming<br>• There may not be sufficient teachers per content area in smaller schools<br>• Requires every teacher to participate | Members with professional development days built into the school schedule; works best in larger schools |
| Regional scoring | Educators from multiple schools within a region gather at a | • Can be more standardized | • Logistical issues for rural areas or | Urban areas with small schools close together |

| Model Name | Quick Description | Benefits | Considerations | Best for |
|---|---|---|---|---|
| | scoring site central to their geographic locations and collectively score exams | • Works well for smaller schools with close geographic neighbors<br>• Allows for collaboration among educators | remote schools<br>• Time-consuming | |
| Moderated scoring within schools | Formal protocol for doing distributed scoring in teams within schools | • Allows for collaboration of educators within a school<br>• Standardization increases reliability and validity | • Time-consuming<br>• Requires many teachers to participate | Members trying to maximize professional development and standardization |
| Hybrid models | Several approaches that combine face-to-face meetings with distributed scoring | • Provides enough face-to-face scoring for every teacher to receive professio | • Could be costly, depending on location of scoring and requirement to | Members trying to provide professional development for everyone without overburdening all teachers |

| Model Name | Quick Description | Benefits | Considerations | Best for |
|---|---|---|---|---|
| | | nal development<br><br>• Best scorers continue scoring all papers to ensure accuracy and reliability<br><br>• Distributed portion maximizes efficiency and standardization | pay distributed scorers | |

## RECOMMENDATIONS

The best scoring model is one that is highly efficient, provides a rich learning environment for teachers, and results in accurate and reliable scores. Yet these three goals are often inherently contradictory. Members may need to prioritize these goals or determine which two are most important. In the interest of maximizing the achievement of all three goals, first consider these guidelines:

1. **Standardize scoring within the consortium.** It is important that papers are scored the same across the consortium. In the case of Smarter Balanced, if comparability of assessment results across members is an important goal, then the papers must be scored the same across all members within the consortium. This can be accomplished relatively easily if there is one set of training modules, training papers, certification papers, and validation papers that are used by every member. The certification standards and validation criteria should be standardized and practiced in every scoring session across the consortium.

2. **Maximize professional development time.** Consider first which tasks will benefit teachers the most. For example, use teachers to score the essays and performance tasks, but leave the short constructed-responses to the vendors for professional scoring. Teachers will learn much more by examining students' extended work than by checking the correctness of the short constructed-responses.

3. **Find cost efficiencies.** Automate as much as possible. Use online delivery systems for training, delivering student papers, collecting scores, completing read-behind scoring, and calculating reliability statistics. Distributed scoring saves costs compared with a centralized scoring system.

The best model appears to be one that allows teachers to work in teams to score, and also contains a read-behind element to improve the reliability of the scores. The research and anecdotal evidence all indicate that teachers receive the most benefit working in teams where they can share their insights and hear others' viewpoints. To avoid the cost of bringing teachers to a centralized location, consider using a moderated scoring model either within a school or within a region, depending on the size of the school and its proximity to other schools. Another cost-saving idea could be using the school or regional scoring model to score each extended response once, and using distributed scoring for the read behind. After the initial scoring, expert teacher scorers could be identified for the read behind, and professional scorers could fill in as needed to ensure a complete second scoring is done quickly. Any discrepancies in scoring would need to be mediated by a lead scorer or expert, who could be provided by the vendor coordinating the scoring effort. In addition, the vendor could set up a hotline to contact lead scorers so that any confusion occurring within school or regional scoring sessions could be addressed immediately. Although this step would add to the cost, it would also add to the professional development of the teachers and to the accuracy of the scoring.

Members will also need to determine if scoring is mandatory or voluntary. Should all schools participate every year or on a rotating basis? These considerations are both financial and political. The more teachers who participate, the quicker the papers will be scored and the fewer papers any one teacher will have to score. However, mandating this activity without reimbursing teachers may cause problems with unions. Consider using professional development days for scoring and/or providing continuing education credit to participating teachers to ensure that they are not taking on extra work without proper compensation.

Finally, the scoring process should be monitored the first few years to determine how teachers react to the process. Are they receiving the intended benefits of the professional development? How does that translate to changes in their instructional approaches? Cost efficiencies need to be monitored to find the balance between providing professional development, including professional scorers where needed, and not overburdening teachers.

# REFERENCES

Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Darling-Hammond, L., & Ancess, J. (1994). *Authentic assessment and school development*. New York: National Center for Restructuring Education, Schools, and Teaching. Teachers College, Columbia University.

Darling-Hammond, L., & Falk, B. (2013). *How student-performance assessments can support teacher learning.* Washington, DC: Center for American Progress.

Darling-Hammond, L., & Rustique-Forrester, E. (2005). The consequences of student testing for teaching and teacher quality. *Yearbook of the National Society for the Study of Education, 104*, 289–319.

Elliott-Schuman, N. (2005). *Grades 4, 7, and 10 Washington Assessment of Student Learning in writing: 2005 teacher scoring*. Seattle, WA: Office of Public Instruction.

Elliott-Schuman, N. (2006). *Grades 4, 7, and 10 Washington Assessment of Student Learning in writing: 2006 teacher scoring.* Seattle, WA: Office of Public Instruction.

Goldberg, G. (2012). Judgment-based scoring by teachers as professional development: Distinguishing promises from proof. *Educational Measurement: Issues and Practices, 31*(3), 38–47.

Goldberg, G., & Roswell, B. (2000). From perception to practice: The impact of teachers' scoring experience on performance-based instruction and classroom assessment. *Educational Assessment, 6*, 257–290.

Heuschel, M. (2004). *A study of training 7th grade teachers in scoring the Washington Assessment of Student Learning (WASL) in writing and mathematics* (Doctoral dissertation). Seattle Pacific University, Seattle, WA.

Jones, M., & Vickers, D. (2011). *Considerations for performance scoring when designing and developing next generation assessments*. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/Performance_Scoring_for_Next_Gen_Assessments.pdf

Kingston, N. M. (1997). *Scoring student responses through imaging: Evidence that new technology will solve a major concern of performance assessment.* Paper presented at the 27th Annual Large-Scale Assessment Conference, Colorado Springs, CO.

Rich, C. S., Harringon, H., Kim, J., & West, B. (2008). *Automated essay scoring in state formative and summative writing assessment.*

Topol, B., Olsen, J., & Roeber, E. (2013). *Getting to higher-quality assessments: Evaluating costs, benefits, and investment strategies.* Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Williamson, D. M., Bejar, I. I., & Hone, A. S. (1999). Mental model comparison of automated and human scoring. *Journal of Educational Measurement, 36*, 158–184.

Wolfe, E., Matthews, S., & Vickers, D. (2010). The effectiveness and efficiency of distributed online, regional online, and regional face-to-face training for writing assessment raters. *The Journal of Technology, Learning, and Assessment, 10*(1). Retrieved from http://jtla.org

Zhang, M. (2013). *Contrasting automated and human scoring of essays*. Retrieved from the Educational Testing Service website: http://www.ets.org/Media/Research/pdf/RD_Connections_21.pdf

# APPENDIX G: PROTECTING STUDENT PRIVACY WITHIN EXEMPLARS OF STUDENT WORK

Exemplars of student work serve a critical role in educational assessment. Exemplars of student work are used to:

- Train educators and other staff to reliably evaluate student work.
- Create consensus among educators regarding expectations of student achievement.

Maintaining the privacy of students whose work is used as exemplars is of paramount importance. Prior to including any element of student work as an exemplar, the student work will be reviewed to determine if the student included any information that either discloses or makes practicable, the discovery of a student's identity. The review of the student's work will include, but not be limited to any of the following information:

- Student or parent/guardian last name,
- Student or parent/guardian employment and financial information,
- Student or parent/guardian residence (e.g. state, county, city, and address),
- Student or parent/guardian phone number,
- Student or parent/guardian email address or online names (Twitter, etc.),
- Teacher name,
- District or School name, or
- A unique combination of any information that explicitly identifies a student or otherwise makes practicable, the discovery of a student's identity.

If any of the above information is present in the student work, either the student work will not be used as an exemplar, or the identifying information will be redacted from the student work before it is used as an exemplar.

# REVISION LOG

Updates to the *Member Procedures Manual* after the August 21, 2018 draft are noted below*.

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| 1.0 | 7 | Changed role of members from gathering and submitting evidence of valid test administration via an annual survey to members gathering and documenting evidence for multiple purposes.<br><br>Added a note about the transition from the current sFTP site to the new Secure File Transfer Server. Updated links are highlighted gray throughout the document. | 8/22/18 |
| 1.2 | 8 | Added new section about the Smarter Balanced Technical Advisory Committee | 8/22/18 |
| 2.0 | 8 | Changed section title from "Evidence Submission" to "Post-Administration Evidence Collection and Documentation." Changed reference to members submitting documentation in an annual survey to members documenting and maintain evidence for various purposes. Added a reference to the Member Test Administration Information document posted on the Member Portal where some member implementation and test administration policies are documented. | 8/22/18 |
| 3.0 | 8 | Section revised to address member service provider test delivery systems since members are not currently using the Smarter Balanced open source TDS. | 8/22/18 |
| 3.1 | 9 | Section revised to address member service provider test delivery systems since members are not currently using the Smarter Balanced open source TDS. | 8/22/18 |
| 3.2 | 9 | Section revised to address member service provider test delivery systems since members are not currently using the Smarter Balanced open source TDS.<br>Changed "Evidence Submission" to "Recommendation | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
|  |  | Regarding Documentation of Evidence" and reference to member submitting evidence to documenting evidence. |  |
| 3.3 | 10 | Renamed this section from "Open-Source System Deployment" to "Test Delivery System Deployment" and included information to address member service provider test delivery systems. | 8/22/18 |
| 3.4 | 10 | Deleted section 3.4 "Documentation and Ancillary Materials for the Open-Source System" and replaced with "Bring Your Own Device (BYOD) Local Option" (formerly section 3.5)<br><br>Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. Added "if applicable" to the recommendation regarding documentation of evidence. | 8/22/18 |
| 4.1.2 | 12 | Removed references to the Practice and Training tests being hosted by Smarter Balanced and replaced with language about member-hosted practice and training tests.<br><br>Clarified information about the use of the Practice Test.<br><br>Added language about the Smarter Balanced Sample Tests hosted on the Smarter Balanced website and clarified the purpose and recommended uses of each instance. | 8/22/18 |
| 5.1 | 14 | Added reference to Test Administrators communicating test security information to students.<br><br>Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | 8/22/18 |
| 5.2 | 15 | Updated the description of interim assessments to align with the Interim Assessments Overview document. | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| 6.1.2 | 16 | Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | 8/22/18 |
| 6.1.3 | 16 | Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. Added language about member's private school and home-instructed student participation policies to be included in the Member Test Administration Information document posted on the Member Portal. | 8/22/18 |
| 6.2.1 | 17 | Added reference to section 6.2.2.

Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. Added language about universal tools, designated supports and accommodations not made available to students documented in the Member Test Administration Information document posted on the Member Portal. | 8/22/18 |
| 6.2.2 | 18 | Changed "adults" to "educators." | 8/22/18 |
| 6.3 | 21 | Under Additional High School Flexibility, updated language regarding Assessing Grades Other than Grade 11.

Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. Added language about members' annual summative test administration windows documented in the Member Test Administration Information document posted on the Member Portal. | 8/22/18 |
| 6.4 | 21 | Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | documenting evidence. | |
| 6.5 | 24 | Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | 8/22/18 |
| 6.8.1 | 25 | Changed reference to the Smarter Balanced open-source test delivery system to member service providers' test delivery systems. | 8/22/18 |
| 6.8.2 | 26 | Removed reference to Math CAT hand-scored items to align with blueprint changes for 2018-19. Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | 8/22/18 |
| 6.8.3 | 28 | Added "Treats of violence" to list of potentially sensitive responses | 8/22/18 |
| 7.3 | 29 | Removed "Members may use the Administration and Registration Tools (ART) open source application provided by Smarter Balanced or use an alternate test administration and registration system as described in section 3.0 Open-Source System Information," from the first paragraph. Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | 8/22/18 |
| 7.5 | 32 | Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. Added language about members' test expiration policies documented in the Member Test Administration Information document posted on the Member Portal. | 8/22/18 |
| 8.1 | 32 | Replaced language regarding use of paper and pencil | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | tests "for those schools and districts whose technology does not meet minimum requirements" with a more general statement about a paper and pencil test being available for members that elect to use it.<br><br>Under Member Responsibility, added "summative results in the Smarter Balanced open-source Reporting System" to the information about transfer of paper/pencil scored data for reporting.<br><br>Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | |
| 8.2 | 33 | Clarified language about availability of and ordering braille and Spanish paper and pencil test forms. | 8/22/18 |
| 8.2.1 | 33 | Updated language regarding the number of available braille forms per grade and content area. Added information about English glossaries in braille available upon request and clarified the ordering process for braille forms. | 8/22/18 |
| 8.2.2 | 34 | Removed reference to test forms numbers and added test administrator script portions of the Spanish test available on the secure sFTP site. | 8/22/18 |
| 8.2.3, Table 4 | 35 | Updated form numbers for 2018-19.<br>Added English Glossaries for ELA/Literacy<br>Added English and Translated Glossaries for Mathematics<br>Removed Audio CD for listening passages for Mathematics | 8/22/18 |
| 9.0 and 9.1 | 36 | Updated language about assessed content for ICAs to match the Interim Assessments Overview document. | 8/22/18 |
| 9.3 | 37 | Changed "open source" to "member service provider's" test administration system.<br>Clarified information about the ICA blueprints.<br>Removed reference to adaptive interims available as | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | the item pool supports them and replaced with information about Focused Interim Assessment Blocks for future development. | |
| 9.4 | 38 | Replaced reference to Smarter Balanced open-source application for interim hand-scoring to "member service provider's interim assessment hand-scoring application." | 8/22/18 |
| 9.5.3 | 39 | Added information about item-level reports available for interim assessments in the Smarter Balanced open-source reporting system. | 8/22/18 |
| 10.0 | 39 | Under Member Responsibility, removed Table 5 that described help desk tiers for Smarter Balanced open-source software and replaced it with, "Members need to provide Tier 1 and Tier 2 support. Smarter Balanced will provide Tier 3 support for the Data Warehouse and Reporting System if members choose to stand up their own instance of the open source software. Members otherwise need to provide Tier 1, 2, and 3 supports for proprietary applications." Renumbered subsequent tables accordingly. | 8/22/18 |
| 12.2 | 41 | Updated list of filters available in the Smarter Balanced open-source reporting system. Updated information about the reports available in the reporting system. Changed "Evidence Submission" to "Recommendation Regarding Documentation of Evidence," and reference to member submitting evidence to documenting evidence. | 8/22/18 |
| 12.3 | 43 | Removed reference to an outline of available mechanisms to facilitate data transfer from the Association of Collegiate Registrars and Admissions Officers and replaced it with information about future work to develop a data transfer mechanism. | 8/22/18 |
| 13.1.1 | 44 | Updated information about the design of the embedded field test of performance tasks. Removed references to "uncalibrated" performance tasks, and | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | "without requiring the rapid calibration cycle of field test performance tasks." | |
| 13.1.2 | 45 | Added two references to section 6.8.3 Scoring of Sensitive Papers and updated dates. | 8/22/18 |
| 13.1.3 | 45 | Corrected section number from 13.1.4 to 13.1.3. Updated the date and information about the number of field test performance tasks and the number of additional items in the enhanced CAT. | 8/22/18 |
| 14.0 | 45 | Changed reference to non-secure/non-public interim items to "student- and teacher-facing" to match the language in the Interim Assessments Overview document.<br><br>Under Member Responsibility, changed references to "secure and non-secure items" to "summative and interim items," and changed reference to Smarter Balanced Practice Test to Smarter Balanced Sample Test. | 8/22/18 |
| Appendix B, Section 1 | 50 | Removed Table 7 that described help desk tiers for Smarter Balanced open-source software and replaced it with, "Members need to provide Tier 1 and Tier 2 support. Smarter Balanced will provide Tier 3 support for the Data Warehouse and Reporting System if members choose to stand up their own instance of the open source software. Members otherwise need to provide Tier 1, 2, and 3 supports for proprietary applications."<br><br>Renumbered subsequent tables accordingly.<br><br>Under Emergent Issues, added Smarter Balanced "provided services" | 8/22/18 |
| Appendix B, Section 3 | 57 | Under Functionality of the Test Delivery System, replaced, "If a member does not use the open-source test delivery system, then the member will need to ensure that their test delivery system supports all functionality needed to successfully administer Smarter Balanced assessments," with new language that doesn't reference the open source system. | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---------|------|------------------------|----------------|
| | | Updated information about calculators to use of the Desmos calculators. | |
| | | Under Other Universal Tools, Designated Supports and Accommodations, removed reference to updates for the open-source test delivery system. | |
| | | Under Adaptive Engine, removed reference to the open-source adaptive engine. | |
| | | Under Student Testing Interface, changed "Using online scratch paper" to "Using online note pad." | |
| | | Under Relationship of Other Components to Test Delivery System, removed a reference to the Smarter Balanced open-source test delivery system. | |
| | | Under Contractor-Specific Test Delivery System, removed references to the option of vendor use of the Smarter Balanced open-source test delivery system. | |
| Appendix B, Section 4 | 61 | Under Machine-Scored Items, clarified the information about the Smarter Balanced open-source scoring engine and machine scoring keys and the option to use the open-source solution or a proprietary engine. | 8/22/18 |
| | | Under Item Scoring for the Summative Assessment, changed the date and removed a reference to the Smarter Balanced open-source test delivery system. | |
| | | Under Machine Scored Items, removed references to the Smarter Balanced open-source test delivery system. | |
| | | Under Hand-Scored Items, updated the number of hand-scored items for 2018-19. | |
| Appendix B, Section 5 | 68 | Under Test Integration, removed a reference to the Smarter Balanced open-source test delivery system. | 8/22/18 |
| | | Under Test Scoring, Option 1, removed a reference to the Smarter Balanced open-source test delivery system. | |
| Appendix B, Section | 74 | Under Scoring Performance Tasks, removed a reference to the Smarter Balanced open-source test | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| 6 | | delivery system.<br>Updated links under Other Documentation. | |
| Appendix B, Section 7 | 75 | Updated Paper and Pencil form numbers for 2018-19 and clarified the information about braille forms | 8/22/18 |
| Appendix B, Section 7, Table 13 | 76 | Removed reference to an Audio CD as a component for Math forms | 8/22/18 |
| Appendix B, Section 7 Table 14 | 77 | Updated number of items and test/answer book pages for 2018-19 Paper and Pencil tests. | 8/22/18 |
| Appendix B, Section 7 | 78 | Under Special Forms, updated the information about ordering braille forms.<br>Under Scoring and Scaling P&P Tests, removed "The item-parameter files are the same for 2015-16 as for 2014-15." | 8/22/18 |
| Appendix B, Section 8 | 82 | Test Administration/Test Delivery System, removed reference to administering interims through the open-source test delivery system.<br>Under Interim Assessment Item Pool, changed the reference to non-secure/non-public to "student and teacher facing," and updated the language to match the Interim Assessments Overview document.<br>Under Test Design/Interim Comprehensive Assessment, replaced reference to the summative blueprints with information about the ICA blueprints. | 8/22/18 |
| Appendix B, Section 9 | 86 | Under Customizable Manuals and User Guides:<br>• Added information about the Test Administrator User Guide Template<br>• Updated the information about the availability of an Interim Assessment TAM.<br>Under Non-Customizable Manuals and User Guides: | 8/22/18 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | • Replaced a section about "Technology Specifications Manual for Online Testing" with a section about "Student Device and Secure Browser Requirements." Under Summative Assessment Training Modules, removed reference to narrated version of the training modules. | |
| Appendix B, Section 10 | 90 | Under Data Warehouse and Reporting System removed reference to "public-facing score reports." Under Accommodations, updated the information about individual student reports to include non-embedded accommodation supports. Under Individual Student Reports, Batch Printing, updated the information about teachers printing reports. Under Aggregate Reports, updated the list of reports available on the Smarter Balanced open-source reporting system. Under Longitudinal Data, updated the information about the availability of longitudinal data reports. | 8/22/18 |
| Appendix C | 98 | Deleted previous Appendix C: Open Source System and Supporting Materials since this information is no longer relevant at the level of the MPM. Renamed Appendices D through I accordingly as Appendices C through H. | 8/22/18 |
| Appendix D | 100 | Added smart watches to the list of unauthorized electronic equipment. | 8/22/18 |
| Appendix E | 102 | Added language to support the need for social media monitoring. Removed "whether the item is from the interim or summative, and" from the list of information to provide about an image of an item on social media. | 8/22/18 |
| 1.0 | 6 | Added language regarding removal of Appendix B: Procurement Information for Administering the | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | Smarter Balanced Assessment System for use as the source document for the Implementation Guide for States and Service Providers. Applicable sections of Appendix B were also added to the main body of the Member Procedures Manual. | |
| 1.0 | 6 | Added information about the new Implementation Guide for States and Service Providers | 8/21/19 |
| 1.0 | 6 | Removed language about transitioning from the Secure FTP site to the new Secure File Transfer Server | 8/21/19 |
| 2.0 and other sections | Various | Changed the name of the *Member Test Administration* Information document to *Member Test Administration Policies* document in this section and throughout the document. | 8/21/19 |
| 3.0 | 8 | Added a new section 3.0 – Student and Teacher Administration and Registration System and Single Sign-on (SSO) with information that was previously included in Section 7.3 – Student and Teacher Administration Registration System and Appendix B, Section 2 – Administration and Registration Tool (ART) from the previous version | 8/21/19 |
| 4.0 – 14.0 | Various | Renumbered sections due to inclusion of new Section 3.0 | 8/21/19 |
| 4.3 | 12 | Changed "verification" to "an evaluation" | 8/21/19 |
| 4.4 | 13 | Added language about referring to the Implementation Guide for States and Service Providers for additional information regarding the components, functionality and configuration of a test delivery system. | 8/21/19 |
| 5.0 | 13 | Removed "General" from the title of this section and moved the information about accessibility resources under 4.1 in the previous version to the Policy and Implementation resources section. Created two new sections that describe test administration resources and policy and implementation resources. Added the *Interim Assessment Guide for* | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | *Administration* and the *Implementation Guide for States and Service Providers* to the Member Responsibility section | |
| 5.1 | 13 | Added a new section, "Test Administration Resources," that describes the various test administration resources available to support members, including the new *Interim Assessment Guide for Administration* | 8/21/19 |
| 5.2 | 14 | Added a new section, "Policy and Implementation Resources," that describes the *Member Procedures Manual*, *UAAG*, and the new *Implementation Guide for States and Service Providers* | 8/21/19 |
| 6.1 | 17 | Changed the title from "Summative Assessments" to Summative Assessments Test Security" | 8/21/19 |
| 6.2 | 19 | Changed the title from "Interim Tests" to "Interim Assessments Test Security" | 8/21/19 |
| 7.1.2 | 20 | Changed the title from "Participation Definition" to "Participation and Performance Calculation." Replaced the reference to "see the Participation and Performance Calculation document" with the language from the document to eliminate the need to go to another document for the information | 8/21/19 |
| 7.2 | 24 | Changed the title from "Accessibility and Accommodations" to "Universal Tools, Designated Supports and Accommodations." Added Table 2 from Section 4.1 of the previous version of the MPM that provides definitions for the three accessibility resources. Renumbered tables thereafter accordingly. | 8/21/19 |
| 7.2.2 | 25 | Changed "April" to "spring" in the second paragraph and removed "(accessibility tool)" after "designated supports" in the fourth paragraph | 8/21/19 |
| 7.5 | 32 | Replaced language about an example of a Test Security Incident Log from the 2014 Field Test with information about a customizable Test Security Incident Log template available on the Secure File | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
|  |  | Transfer Server |  |
| 7.5.1 | 32 | Replaced references to "the Consortium" with "Smarter Balanced" | 8/21/19 |
| 7.5.2 | 32 | Changed "Appendix C" to "Appendix B – Online System – Appeal Types and Conditions for Use" | 8/21/19 |
| 7.6 | 33 | Added language about performance tasks from Appendix B, Section 6 – Performance Tasks of the previous version of the MPM. The only text in the Performance Task section was the Member Responsibility. | 8/21/19 |
| 7.7 | 36 | Changed title from "Scoring" to "Item Scoring. Added language about item scoring from Appendix B, Section 4 – Item Scoring for the Online Summative Assessment of the previous version of the MPM. | 8/21/19 |
| 7.7.1 | 36 | Added "automated scoring application" before "(artificial intelligence (AI) scoring application)" | 8/21/19 |
| 7.7.2 | 37 | Added language from Appendix B, Section 4 – Item Scoring for the Online Summative Assessment of the previous version of the MPM and the *Implementation Guide for States and Service Providers* about external scoring. Added missing table numbers and titles. | 8/21/19 |
| 7.8 | 43 | Added a new section, Test Integration and Test Scoring from Appendix B, Section 5 – Test Integration and Scoring of the previous version of the MPM | 8/21/19 |
| 7.9 | 47 | Added a new section, Sending Data to Smarter Balanced. This information was previously available via email to K-12 Leads but was not included in the previous version of the MPM. This language is also included in the *Implementation Guide for States and Services Providers.* | 8/21/19 |
| 8.1 | 48 | Changed the title from "Online Test Packages" to "Online Summative Test Packages" | 8/21/19 |
| 8.2 | 48 | Changed the title from "Ancillary Test Administration Materials – Online Summative Assessments" to "Online, Summative, Test Administration Manual." | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | Added "Summative" before all "TAM" references in this section and under Member Responsibilities, and new language about additional information about the Online Summative TAM in Section 5.1.1 | |
| 8.4 | 50 | Replaced "total length" with "approximate amount" of time in the first paragraph | 8/21/19 |
| 9.1 | 51 | Under Member Responsibility, removed text about members being responsible for implementing a process for transferring the scored data from the paper assessment into the Smarter Balanced secure FTP site for reporting results in the Smarter Balanced open-source Reporting System. | 8/21/19 |
| 9.2 | 51 | Changed the title from "Provided Materials – Paper and Pencil Summative Assessment" to "Test Forms – Paper and Pencil Summative Assessment. Added language from Appendix B, Section 7 – Paper and Pencil Assessments of the previous version of the MPM. Removed "Gridded Response" items from the list of "Items Included on the Paper and Pencil Assessment." Under Member Responsibility, added language about service providers needing to merge student data from paper pencil and online administrations for reporting and prior to submission of data to Smarter Balanced and deleted language about CDs for the Listening portion of the ELA/literacy assessment. | 8/21/19 |
| 9.2.1 | 52 | Changed the title from "Braille" to "Braille Versions" and added "the American Printing House for the Blind" before "(APH}" | 8/21/19 |
| 9.2.2 | 53 | Changed the title from "Spanish" to "Spanish Translation for Mathematics" and changed the location of the files from the "secure FTP site" to the "Secure File Transfer Server." | 8/21/19 |
| 9.2.3 | 53 | Added a new section 9.2.3 – Large Print with language from Appendix B, Section 7 Paper and Pencil Assessments of the previous version of thee MPM | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| 9.2.4 | 53 | Updated the paper and pencil test forms information for 2019-20. Removed language about seeing Appendix B for additional information about page counts and item counts. | 8/21/19 |
| 9.2.5 | 54 | Added a new section 9.2.5 – Test Booklets and Answer Booklets with language from Appendix B, Section 7 Paper and Pencil Assessments of the previous version of the MPM. Updated Table 13 – Test Booklets Item and Page Counts for 2019-20. | 8/21/19 |
| 9.3 | 54 | Replaced language in 9.3.1 – Paper and Pencil Test Administration Manual, and added section 9.3.2 – Listening Passages Resources, section 9.3.3 – Manipulatives, section 9.3.4 – Calculators, and section 9.3.5 – Translated Glossaries with language from Appendix B, Section 7 Paper and Pencil Assessments from the previous version of the MPM. | 8/21/19 |
| 9.5 | 55 | Added a new section 9.5 – Scanning and Scoring Tests with language from Appendix B – Section 7 Paper and Pencil Assessments from the previous version of the MPM. Added new section 9.5.1 – Scanning and Machine Scoring of Responses, section 9.5.2 - Hand Scoring, section 9.5.3 – Double Scoring, and section 8.5.4 – Scoring and Scaling Paper and Pencil Assessments. | 8/21/19 |
| 10.0 | 56 | Updated section 10.0 with language from the Interim Assessments Overview doc and the *Interim Assessments Guide for Administration* | 8/21/19 |
| 10.1 | 57 | Created new section 10.1 – Interim Comprehensive Assessments and updated the language to match the *Interim Assessments Guide for Administration* | 8/21/19 |
| 10.2 | 58 | Created new section 10.2 – Interim Assessment Blocks and updated the language to match the *Interim Assessments Guide for Administration* including information about focused IABs | 8/21/19 |
| 10.3 | 58 | Changed title from "Interim Test Design" to "Interim Assessment Blueprints" and replaced the language to | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | match the *Interim Assessments Guide for Administration* | |
| 10.4 | 59 | Replaced "and in thereafter included" with "or in subsequent years" after "2014" | 8/21/19 |
| 10.5 | 59 | Added a new section 10.5 – Interim Test Security using language from the *Interim Assessments Guide for Administration* | 8/21/19 |
| 10.6 | 59 | Deleted language about ICA and IAB blueprints that is included in the new section 10.3 – Interim Assessments Blueprints and replaced "an off-grade level" with "any grade level" in the Grade Levels section. | 8/21/19 |
| 10.6.1 | 60 | Added a new section about the *Interim Assessments Guide for Administration* | 8/21/19 |
| 10.7 | 61 | Removed the section about Member Monitoring since we don't have any guidance about monitoring teacher hand scoring | 8/21/19 |
| 10.8 | 61 | Changed the title from "Interim Data Warehousing and Reporting" to "Interim Reporting" | 8/21/19 |
| 10.8.1 | 61 | Removed reference to a "member operated instance of the Smarter Balanced open-source Data Warehouse and Reporting system" | 8/21/19 |
| 10.8.2 and 10.8.3 | 61 | Replaced "levels of classification" with "reporting categories" | 8/21/19 |
| 11.0 | 62 | Replaced reference to the "Procurement Guidance in Appendix B" with the "*Implementation Guide for States and Service Providers*" and replaced language about help desk support to match the language in the Implementation Guide | 8/21/19 |
| 13.0 | 63 | Changed the title from "Date Warehouse and Reporting" to Data Warehouse and Reporting System." Removed references to the Smarter Balanced test delivery system and Appendix B | 8/21/19 |
| 13.1 | 63 | Removed section about Required Data Elements since | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | this is covered in new section 7.9 – Sending Data to Smarter Balanced | |
| 13.2 | 64 | Replaced "PDF download" with "downloading a PDF version" in the second paragraph. Updated the description of the Reporting System User Guide under "RESOURCES" | 8/21/19 |
| 14.0 | 66 | Changed the title from "Future Field Testing" to "Field Testing New Items" | 8/21/19 |
| 14.1 | 66 | Added new section 14.1 – Field Testing CAT Items and updated the information about CAT field testing for 2019-20 | 8/21/19 |
| 14.2 | 67 | Added new section 14.2 – Field Testing Performance Tasks to replace section 13.1 – Embedded Field Test of Performance Tasks and section 13.1.1 – Design of the Embedded Field Test of Performance Tasks in the previous version of the MPM. Updated the information about PT field testing for 2019-20 | 8/21/19 |
| 14.2.1 | 67 | Replaced reference to "2018-19" with "each test administration year" under "Smarter Balanced is responsible for" | 8/21/19 |
| 14.2.2 | 67 | Updated the information about the plan for the embedded field test of PTs for 2019-20 | 8/21/19 |
| 15.0 | 68 | Added information about Professional Learning Resources available in the Digital Library | 8/21/19 |
| Appendix A | 69 | Added information about addressing policy issues and updated location of resources available in the Secure File Transfer Server | 8/21/19 |
| Appendix B | 71 | Removed Appendix B: Procurement Information for Administering the Smarter Balanced Assessment System and replaced with former Appendix C: Online System – Appeal Types and Conditions for Use. All subsequent Appendices re-lettered accordingly For the Invalidation description, replaced "eliminates" with "deletes" | 8/21/19 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| Appendix C | 73 | Moved "Administrator or Coordinator leaving related instructional materials on the walls in the testing room" from Low Impropriety to Medium Irregularity.<br><br>Moved "Disruptions to a test session such as a fire drill, school-wide power outage, earthquake, or other acts" from Medium Irregularity to Low Impropriety.<br><br>Added "(e.g., calculators during non-calculator sections)" as an example of a non-allowable device. | 8/21/19 |
| Appendix D | 75 | Updated the information about posting images of test items to the Secure File Transfer Server | 8/21/19 |
| Global edits | various | Changed references to the Digital Library to Tools for Teachers<br><br>Changed references to 2019-20 to 2020-21<br><br>Embedded all hyperlinks in the associated text to address accessibility compliance issues<br><br>Deleted all text boxes and moved information to the body of the document to address accessibility compliance issues<br><br>Changed all references to the Administration and Registration Tools (ART) to "test registration tool"<br><br>Changed all references to the ART User Guide to Test Registration Tool User Guide Template<br><br>Changed all references to the Development and Design page of the Smarter Balanced website to the Test Development and Design page of the Smarter Content Explorer | 8/27/2020 |
| 1.0 | 6 | Removed language about first version without Appendix B and creation of the Implementation Guide for States and Service Providers | 8/27/2020 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | Removed "smooth" before "implementation" and added "assessments" after "interim"

Added the Smarter Content Explorer to the locations in which resources may be found | |
| 3.0 | 8 | Changed the section title from Student and Teacher Administration and Registration System to Test Registration Tool and Single Sign-on | 8/27/2020 |
| 3.1 | 8 | Added "and responsibilities" after "roles" | 8/27/2020 |
| 3.1.3 | 9 | Corrected section reference for additional information about accessibility resources for students

Removed "digital" before "test administration delivery system" | 8/27/2020 |
| 3.1.3 | 10 | Added language about the test registration tool assigning accessibility supports for students and clarified the paragraph under Monitoring | 8/27/2020 |
| 3.2 | 10 | Updated the information about the transition to federated SSO | 8/27/2020 |
| 4.0 | 11 | Clarified the information under Member Responsibility and added a link to the Item Types and Accessibility Features document | 8/27/2020 |
| 5.1.4 | 14 | Added information about the ART training model being used as a template for the creation of a training module for the service provider's test registration tool. | 8/27/2020 |
| 5.3 | 16 | Changed the section title to Practice Tests, Training Tests and Released Items
Updated the information about the content for the 2020-21 Practice and Training tests
Replaced reference to the Smarter Balanced Sample Test with the Sample Items Website | 8/27/2020 |
| 5.3.1 | 17 | Added new section about member use of released | 8/27/2020 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | items including a new Member Responsibility section | |
| 6.1 | 19 | Deleted text box to address accessibility issue. The same information is in the body of the page | 8/27/2020 |
| 7.5 | 33 | Deleted text box to address accessibility issue. Added the information from the text box under Recommendation Regarding Documentation of Evidence on page 34 | 8/27/2020 |
| 7.6.2 | 37 | Table 5: Clarified the language about Total Duration for math PTs to match the Online Summative TAM | 8/27/2020 |
| 7.9.3 | 51 | Corrected the name of the Secure File Transfer Server | 8/27/2020 |
| 9.0 | 55-61 | Replaced "paper and pencil" with "paper-pencil for consistency with other documents | 8/27/2020 |
| 9.2 | 56 | Clarified that the Spanish mathematics forms are available in Stacked Spanish presentation | 8/27/2020 |
| 9.2.2 | 57 | Changed the title Spanish Translation for Mathematics to Spanish Translated Version for Mathematics to clarify this is a version of the test. Changed references to Spanish translations of the mathematics form to Spanish translated mathematics forms Changed response booklets to answer booklets Clarified that the Spanish translated forms are presented in Stacked-Spanish. | 8/27/2020 |
| 9.2.4 | 58 | Table 12: Changed name of Spanish form to match previous sections | 8/27/2020 |
| 9.2.5 | 59 | Table 13: Added Grade 10 Math booklets item and page counts and updated item and page counts for Forms 4 and 5. | 8/27/2020 |
| 9.3.3 | 60 | Clarified the language regarding use of manipulatives. Added "or allow for" after "include." | 8/27/2020 |
| 9.3.5 | 60 | Clarified the information about the number of pages for glossaries | 8/27/2020 |
| 10.0 | 62 | Added policy for remote administration of interims as | 8/27/2020 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | approved by members on 6/22/2020 | |
| 10.2 | 63 | Added Focused IABs to the section title and updated the information about Focused IABs | 8/27/2020 |
| 10.3 | 64 | Updated location of the blueprints to include the Document Library of the Member Portal. Added Focused IAB blueprints. Deleted redundant information about interim assessments from the end of this section. | 8/27/2020 |
| 10.6 | 66 | Added clarifying information regarding member responsibility for member-created interim assessments. | 8/27/2020 |
| 10.6.1 | 65 | Updated the location of the Interim Assessment Overview and Blueprints to include the Member Portal | 8/27/2020 |
| 10.7 | 66 | Removed reference to interim hand scoring materials being available in the open-source system. | 8/27/2020 |
| 10.8.3 | 67 | Added "scale score and standard error of measure" to information provided in IAB reports | 8/27/2020 |
| 13.1 | 69 | Updated information about the Data Warehouse | 8/27/2020 |
| 13.2 | 69 | Deleted reference to "including registration statistic, assessment completions" from available data extracts in the reporting system | 8/27/2020 |
| 13.2 | 70 | Added "Members may also develop a customized student score report for distribution to parents/guardians" to footnote #2 | 8/27/2020 |
| 14.2.2 | 74 | Added a note about the plan for the 2020-21 field test possibly changing due to no field testing in 2019-20 and possible implications due to school closures and/or the need for remote testing.. | 8/27/2020 |
| 15.0 | 74 | Replaced this section about the Digital Library with Tools for Teachers | 8/27/2020 |
| 15.1 | 75 | Added new section about member-Owned Tools for Teachers Resources | 8/27/2020 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| Appendix A | 77 | Updated Note/Resources | 8/27/2020 |
| Appendix A | 78 | Replaced ART with "the appeals system"<br>Removed text box around "Note" language to address accessibility issue | 8/27/2020 |
| Appendix B | 81 | Added a note: Members may allow for additional appeals (e.g., Grade Period Extension) | 8/27/2020 |
| Appendix D | 84 | Edited the rational for social media monitoring to highlight the primary use by educators and policymakers | 8/27/2020 |
| Appendix D | 85 | Updated language to match the Social Media Monitoring Resource | 8/27/2020 |
| 8.1 | 54 | Added new information about the adjusted form summative blueprint | 12/15/2020 |
| 8.5 | 57 | Added new section 8.5 – Remote Administration of Summative Assessments for 2020-21 | 12/15/2020 |
| 9.1 | 58 | Added information about remote administration of paper-pencil tests | 12/15/2020 |
| 14.2.2 | 77 | Updated the plan for 2020-21 Embedded Field Test of PTs | 12/15/2020 |
| 1.0 | 8 | Replaced "ensure that comparable scores are generated across the consortium" with "afford members the opportunity to derive scores that are comparable" since we do not generate and report Consortium-wide test scores. | 8/5/2021 |
| 1.2 | 9 | Updated link to the Governance Structure page on the Smarter Balanced website. | 8/5/2021 |
| 3.1 | 12 | Added universal tools to the other accessibility supports under Member Responsibility | 8/5/2021 |
| 4.0 | 13 | Updated the requirement for a physical keyboard in the Student Device and Secure Browser Requirements. Added. "An exemption may be made | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | for students requiring assistive technology" per recommendation from the Student Supports Team. | |
| 5.1 | 15 | Global edit: Spelled out the full name before the acronym for the first mention in a section and used the acronym only in later references in the same section (e.g., computer adaptive test (CAT) and performance task (PT)<br><br>Added information about the Braille Test Administration Manual Insert | 8/5/2021 |
| 5.1.2 | 16 | Added new Section 5.1.2 about the Supplement: Online Summative Test Administration Manual for Remote Test Administration and renumbered subsequent sections accordingly | 8/5/2021 |
| 5.2.2 | 17 | Added "embedded" and "non-embedded" to clarify "digitally-delivered components" from separate components for the test administration system. | 8/5/2021 |
| 5.2.3 | 19 | Added assistive technology manuals to the list of additional materials under Member Responsibility. | 8/5/2021 |
| 5.3 | 19 | Updated the information about Practice and Training Test packages for 2021-22<br><br>Updated the information about use of a secure browser for Practice and Training Tests<br><br>Added information about the Smarter Annotated Response Tool (SmART) | 8/5/2021 |
| 5.3.1 | 20 | Changed "continues to be" to "is" a member in good standing | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| 6.1 | 21 | Moved "The action or inaction of one Consortium member may have consequences for the Consortium as a whole," from the middle of the second paragraph to the beginning.<br><br>Added "proctors, and staff providing non-embedded accessibility resources." and removed duplicate "test administrators" in the second dot point. | 8/5/2021 |
| 6.2 | 22 | Updated the language about the restriction on posting interim items to a third-party application that is "not approved by the member state education agency and in consultation with Smarter Balanced."<br><br>Added clarifying language to the remote administration of interim assessments policy about allowing for the administration of interim assessments without the use of a secure browser.<br><br>Changed "than" to "from" in the last sentence. | 8/5/2021 |
| 7.1.1 | 23 | Global edit: Replace accessibility "tools" with "resources" for consistency with other documents and manuals | 8/5/2021 |
| 7.1.2 | 25 | Changed "Honor Student's Effort" to "Honor Students' Efforts"<br>• Replaced "the" with "an individual" student's control<br>• Replaced "she/he is" with "they are" to use gender-neutral pronouns<br><br>Under Procedures: | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | • Global edit: Added the acronyms for computer adaptive test and performance task and used the acronyms for subsequent references in this section<br><br>Under Valid Tests:<br>• Added a dot point to call out adherence to standardized test administration policies and procedures. (e.g., reading test directions, maintaining a suitable test environment) | |
| 7.2.1 | 27 | Global edit: Correct the names of the accessibility resources. Changed "Tools and Supports to "Universal Tools" and "Designated Supports" | 8/5/2021 |
| 7.2.2 | 28 | Changed the time for member requests for changes to the UAAG from "every spring" to "each year" due to a change in the UAAG timeline effective this year. | 8/5/2021 |
| 7.3 | 30 | Similar to the edit in Section 1.0 - Introduction, changed "across members" to "year-to-year" since we do not aggregate and report Consortium-wide test scores. | 8/5/2021 |
| 7.5<br>Table 3 | 34 | Replaced " An unusual circumstance" and "An event" with "An event that does not align to standardized test administration policy or procedures ..." to clarify that it is anything that does not conform to standardized test administration. | 8/5/2021 |
| 7.5 | 36 | Under Member Responsibility, clarified "Date" to "Date of incident" | 8/5/2021 |
| 7.5.2 | 37 | Global edit, changed all references to Test Administration Manual to Online Summative Test Administration Manual (TAM) | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| 7.6.2 | 40 Table 5 | Clarified that graph paper should be plain (no coordinate planes) to align with the UAAG and TAMs | 8/5/2021 |
| 7.7.2 | 42 Table 7 | Updated the number of hand-scored items for 2021-22

Global edit: Under Hand Scoring Process #7, replaced "Stacked Spanish" with "Dual Language Translation" to align with the updated UAAG | 8/5/2021 |
| 7.9.3 | 54 | Added "access" before "the information" | 8/5/2021 |
| 8.1 | 54 | Updated information about the availability of both a full form and adjusted form blueprint for 2021-22 | 8/5/2021 |
| 8.5 | 57 | Remote administration of summative assessments policy updated per member approval on 5/6/2021 | 8/5/2021 |
| 9.1 | 58 | Remote administration of paper-pencil tests policy updated per member approval on 5/6/2021 | 8/5/2021 |
| 9.2 | 59 | Updated information regarding braille forms | 8/5/2021 |
| 9.2.4 Table 12 | 61 | Updated Paper-Pencil Forms for 2021-22 (Form 4 is operational and Form 5 is breach) | 8/5/2021 |
| 9.2.5 Table 13 | 62 | Updated Table 13 for 2021-22 | 8/5/2021 |
| 9.3.1 | 63 | Added information about the Braille Test Administration Manual Insert | 8/5/2021 |
| 9.5.3 | 64 | Added the Smarter Balanced recommendation for a minimum of 10% of student responses to be double scored as noted in the Member Responsibilities for section 7.7.2 | 8/5/2021 |
| 10.0 | 65 | Added earliest start date for interim assessments may be August 3, 2021, for the 2021-22 school year. | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | Clarified the dot point about TAs following state and local policies regarding test security to include a statement about the option for members to establish a policy to allow for administration of interim assessments without the use of a secure browser and procedures to address item security, as discussed with members on 6/8/2021. | |
| 10.1 | 65 | Renamed the title of this section to "Types of Interim Assessments. Created new subsections:<br><br>10.1.1 The Interim Comprehensive Assessments (ICA)<br><br>10.1.2 The Interim Assessment Blocks (IAB) and Focused IABs | 8/5/2021 |
| 10.1.2 | 67 | Removed "focused" from the first sentence of the description of Interim Assessment Blocks to avoid confusion with Focused IABs. | 8/5/2021 |
| 10.2 | 67 | Added new Section 10.2 - Interim Assessment Item Portal (IAIP) | 8/5/2021 |
| 10.5 | 69 | Updated the language about the restriction on posting interim items to a third-party application that is "not approved by the member state education agency and in consultation with Smarter Balanced."<br><br>Added a note about allowing for the administration of interim assessments without the use of a secure browser in support of a reference to this section in the clarification added to Section 10.0. | 8/5/2021 |
| 13.2 | 73 | Section title renamed Smarter Reporting System. Introductory information updated. | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | Added Military Connectedness to the list of filters | |
| | | Updated information about reporting Writing Trait Scores | |
| | | Updated footnote 2 about the intended use of the ISR | |
| | | Updated the name of the reporting system user guide | |
| 14.1 | 76 | Spelled out the first mention of computer adaptive test (CAT) in this section and used the acronym in subsequent sentences.

Clarified that the number of responses per item used for computing item statistics and calibrations is "approximately" 1,500. | 8/5/2021 |
| 14.2 | 76 | Spelled out the first mention of performance task (PT) and computer adaptive test (CAT) in this section and used the acronyms in subsequent sentences. | 8/5/2021 |
| 14.2.1 | 77 | Changed references to "test administration manual" to the "Smarter Balanced Online Summative Test Administration Manual" | 8/5/2021 |
| 14.2.2 | 77 | Updated the field test plan for 2021-22 | 8/5/2021 |
| 15.0 | 78 | Made clarifying edits to the introductory information about Tools for Teachers | 8/5/2021 |
| 15.1 | 79 | Removed: "Smarter Balanced will support resource author roles in TIMS for those authoring member-owned resources up to an established limit during scheduled authoring windows."

Added: "Smarter Balanced will provide training to the member designee on how to use the tenant specific | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| | | administrative tab in TIMS and how to sync approved tenant specific resources in Tools for Teachers for posting." <br><br> Replaced "UAT" with "the associated quality assurance" <br><br> Added: "Should a member wish to share resources with another member, or members, they may enter into an agreement to do so without the involvement or approval of Smarter Balanced. Members who agree to share resources with another member will indicate the state/territory with which the resources should be shared in TIMS. A sync will be required to ensure the resource is live in the tenant with which it's being shared." <br><br> Under Resources, added: <br> • Instructional Resource Quality Criteria <br> • TIMS Resources Authoring Technical Guide | |
| Appendix B | 83 | Under Reset: <br> Added "or presentation" after "braille is a language" since some test delivery systems refer to braille as a presentation rather than a language. <br><br> Changed "Translation - glossary" to "Translations – glossary" to match the UAAG <br><br> Changed reference to "grade 11" to "high school" | 8/5/2021 |
| Appendix D | 88 | Updated this section to include a reference to remote testing and additional social media platforms | 8/5/2021 |

| Section | Page | Description of Change | Revision Date |
|---------|------|-----------------------|---------------|
| | | Under Test Administration Procedures, replaced references to "Twitter accounts" with more general "social media accounts"<br><br>Added "and any acronym used for your state test" to the search queries | |
| 4.4 | 14 | Created new Section 4.4 Test and Item Management System (TIMS) that includes dates by which only urgent feedback about test content will be reviewed and acted upon for the 2021-22 SY.<br>Renumbered subsequent section accordingly. | 11/29/2021 |
| Global | | Replaced 'tools' with 'resources' throughout when reference accessibility resources. | 8/4/2022 |
| Global | | Replaced 'Grade 11' with 'high school' | 8/4/2022 |
| Global | | Made edits such as removing etc. and other minor grammatical alterations throughout the document. | 8/4/2022 |
| Global | | Changed to tables and text throughout to reflect Smarter Balanced updated style guide. | 8/4/2022 |
| 1.0 | 9 | Spelled out the acronym the first time it was used and for each new section. This is consistent with the practice implemented in 2020-21. | 8/4/2022 |
| 3.1.3 | 12 | Updated the accessibility resources to include the new videos that explain the ISAAP Tool. | 8/4/2022 |
| 3.1.3 | 12 | Revised language. | 8/4/2022 |
| 3.1.3 | 13 | Replaced "all" with "participating." | 8/4/2022 |
| 3.1.3 | 13 | Added the word "profiles" following "students." | 8/4/2022 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
| 3.2 | 14 | Updated language to reflect the ongoing use of the SSO. | 8/4/2022 |
| 3.2 | 14 | Replaced ISAAP Tool Instructions with two new videos: *Understanding the Individual Student Assessment Accessibility Profile*<br><br>*Individual Student Assessment Accessibility Profile: A Tool to Support Students* | 8/4/2022 |
| 4.1 | 15 | Removed *4.1 Open Source System* as it is no longer applicable. | 8/4/2022 |
| 4.1-4.5 | 15-17 | Renumbered 4.2 is 4.1, 4.2 etc. through 4.5. | 8/4/2022 |
| 4.3 | 17 | Revised dates. | 8/4/2022 |
| 5.2.2 | 20-21 | Removed *Resources and Practices Crosswalk* and added a direct link to the Accessibility page. | 8/4/2022 |
| 5.3 | 22 | Referenced 2022-23 test package training materials. | 8/4/2022 |
| 6.1 | 25 | Updated language and removed reference to section 1.3 which is no longer available. | 8/4/2022 |
| 6.2 | 25 | Bolded text for emphasis. | 8/4/2022 |
| 7.1.2 | 28 | Removed unessential wording to increase clarity in the section: Promote Equity. | 8/4/2022 |
| 7.1.2 | 29 | Updated language for tools and accessibility resources. | 8/4/2022 |
| 7.2.3 | 34 | Added 7.2.3 and included information about languages and the process for updating glossaries. New text reflects current practices. | 8/4/2022 |
| 7.2.3 | 35 | Removed *Usability, Accessibility, and Accommodations Guidelines FAQ*, as this information | 8/4/2022 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | has been consolidated and links to the accessibility website. | |
| 7.3 | 35 | Amended text to reflect that members will alter test administration policies as needed. | 8/4/2022 |
| 7.7.2 | 48 | Updated Table 7 to include anticipated hand scored items for 2022-23. | 8/4/2022 |
| 7.7.2 | 51 | Added the note: Starting in school year 2022-23, student responses to the ELA/literacy full write items that receive a condition code of off purpose will still be scored for the dimension, Conventions. | 8/4/2022 |
| 9.2.4 | 68-69 | Added reference to English glossaries in Braille. | 8/4/2022 |
| 10.1.2 | 76 | Added Interims Assessment At-A-Glance graphic. | 8/4/2022 |
| 13.2 | 82-83 | Added language to delineate between Filters categorized by the Department of Education and non-DE filters. Modernized the terms. | 8/4/2022 |
| Change log | 126 | The committee recommended only carrying forward three years of revisions in the change log beginning in 8/2018. This year's log includes pre- and post-pandemic changes. Changes prior to 8/2018 have been removed from the change long. The change log for the MPM prior to 2018 will be archived. | 8/4/2022 |
| Global Edits | Various | • Adjusted formatting of headings for consistency, to align to updated branded template, and improve readability.<br>• Removed all caps from "Recommendation Regarding Documentation of Evidence" sub-headings for consistency.<br>• Copy edits from Smarter Balanced Communications team for consistency and clarity.<br>• Adjusted formatting of all internal document links for consistency. | 10/31/2023 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| | | • Changed "Focused IABs" to "FIABs."<br>• Removed most references to remote teaching and learning throughout document.<br>• Changed "ELA" to "ELA/literacy" for consistency. | |
| Table of Contents | 2 | Adjusted formatting for clarity and consistency. | 10/31/2023 |
| 3.1 | 10 | Inserted paragraph at start of section for consistency with Implementation Guide for States & Service Providers | 10/31/2023 |
| 3.2 | 13 | • Removed "in July 2019" from final paragraph.<br>• Other Documentation: Links updated to reflect current documentation. | 10/31/2023 |
| 4.3 | 16 | • Updated content package release dates for 23-24 administration year.<br>• Added TIMS sentences to end of section to tie in function of TIMS in authoring instructional resources | 10/31/2023 |
| 5.1.3 | 17 | Added link to Test Administrator User Guide template. | 10/31/2023 |
| 5.1.5 | 18 | Changed heading title from "5.1.4 Online Training Modules" to "5.1.5 Online Training Modules." | 10/31/2023 |
| 5.2.3 | 20 | Added direct link to Implementation Guide for States and Service Providers document in Member Portal. | 10/31/2023 |
| 7.1.1 | 25 | Updated first sentence of section to clarify who "participating students" are. | 10/31/2023 |
| 7.1.2 | 27 | Under "Incentivize Desirable Outcomes" heading, adjusted the language in the first sentence for clarity. | 10/31/2023 |
| 7.1.2 | 29 | Under "Calculating Scale Scores": Adjusted capitalization and included acronym for Lowest | 10/31/2023 |

| Section | Page | Description of Change | Revision Date |
|---------|------|----------------------|---------------|
|  |  | Obtainable Scale Score (LOSS) and Highest Obtainable Score (HOSS) for consistency with Implementation Guide for States & Service Providers document. |  |
| 7.2.1, Table 2 | 31 | Updated column headings to "Type" and "Definition" for consistency with the corresponding table in the Implementation Guide for States and Services Providers document. | 10/31/2023 |
| 7.2.3 | 34 | Changed wording of last sentence to reflect past tense of most recent survey administration, and note that results are being processed. | 10/31/2023 |
| 7.7.1 | 46 | Changed section heading to "Machine-Scored Items" to align with the Implementation Guide for States & Service Providers document. | 10/31/2023 |
| 7.7.2, Table 7 | 47 | In the "Total Number of Hand-Scored Items Across All Grades" column, updated the values to reflect 2023-24 item pool that meet this criterion. | 10/31/2023 |
| 7.7.2 | 48 | Number 5 under "Hand-Scoring Process" sub-heading: Added sentence to recommend that 10% of student responses be the minimum level of souble scoring allowed. | 10/31/2023 |
| 7.7.2 | 50 | In final paragraph (beginning with "Note:"), added link to updated scoring rules. | 10/31/2023 |
| 7.7.2 | 53 | Added *Appendix F: Large-Scale Scoring: Involving Teachers and Producing Results* hyperlink to "additional scoring resources" list. | 10/31/2023 |
| 7.8.3 | 55 | • Claim-level Score sub-heading: Changed "At/Near Standard" to "Near Standard." <br> • Other Member-specific Scores sub-heading: reworded paragraph for clarity. | 10/31/2023 |

| Section | Page | Description of Change | Revision Date |
|---------|------|------------------------|---------------|
| 7.8.7 | 57 | • Changed "At/Near Standard" to "Near Standard."<br>• Added sentence to end of sub-section addressing composite scores. | 10/31/2023 |
| 7.8.8 | 57 | Added Smarter Balanced Scoring Specifications hyperlink. | 10/31/2023 |
| 8.5 | 63 | Added link to Remote Summative Testing webpage to last paragraph of section. | 10/31/2023 |
| 9.1 | 64 | Added final sentence to the Member Responsibility sub-heading to bring section into alignment with the Implementation Guide for States & Service Providers document. | 10/31/2023 |
| 9.2 | 64 | Updated the paragraph to reflect that only one operational paper-pencil form is provided per grade and content area. | 10/31/2023 |
| 9.2.1 | 66 | Updated first sentence to reflect that only one operational test form is developed for each grade and content area. | 10/31/2023 |
| 9.2.2 | 66 | Updated first sentence to reflect that only one operational test form is developed for each grade. | 10/31/2023 |
| 9.2.4, Table 12 | 67 | Updated the information in the "forms" column to reflect that only one form (form 6) is developed for each content area. | 10/31/2023 |
| 9.2.5, Table 13 | 68 | Removed all "form 5 (breach)" information from table and updated the numbers of pages in test booklets in the "Form 6 (Operational)" cells. | 10/31/2023 |
| 10.0 | 71 | Formatting: put text into a bulleted list after second paragraph to reflect conceptual relationship. | 10/31/2023 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| 10.0 | 73 | Added sentence to end of section that includes link to Interim Assessment Administration webpage. | 10/31/2023 |
| 10.1.2 | 73 | Added "FIABs" acronym for "focused IABs" for consistency with other documents. | 10/31/2023 |
| 10.6.1 | 77 | Removed second bullet point from "resource" sub-section due to redundancy. | 10/31/2023 |
| 13.2 | 81 | Changed "Economic Disadvantaged" filter to "Lower socioeconomic status." | 10/31/2023 |
| 13.2 | 81 | • ISR bullet point: Added last sentence to align with Implementation Guide for States & Service Providers document.<br>• Claim Report bullet: Added last two sentences to align with Implementation Guide for States & Service Providers document.<br>• Target report bullet: Added last two sentences to align with Implementation Guide for States & Service Providers document. | 10/31/2023 |
| 13.2 | 82 | After "custom aggregate reports" bullet, added sentence separate from bulleted list to align content with Implementation Guide for States & Service Providers document. | 10/31/2023 |
| 15.0 | 86 | Added additional sentence to end of section to reference Implementation Guide for States and Service Providers document. | 10/31/2023 |
| Appendix B | 91 | Fixed bullet point formatting for clarity and readability in "conditions for use" column. | 10/31/2023 |
| Appendix B | 92 | Corrected numbering and formatting in "conditions for use" column. | 10/31/2023 |

| Section | Page | Description of Change | Revision Date |
|---|---|---|---|
| Appendix B | 92-93 | Minor adjustments to formatting of text in "conditions for use" column for readability. | 10/31/2023 |
| Appendix C | 97 | Added "Other documentation" bullet points for consistency with Implementation Guide for States and Service Providers document. | 10/31/2023 |
| Appendix D | 99 | Heading text changed from "Member Responsibilities" to "Member State Responsibilities" | 10/31/2023 |
| Appendix D | 99 | "#common core" changed to "#commoncore" | 10/31/2023 |
| | | | |
| | | | |