December 16, 14

Implementation Readiness Package

Smarter Balanced

Document Version: 2.0

Table of Contents

Implementation Readiness Package (IRP)

Overview

States are responsible for deploying and operating the Smarter Balanced Assessment Delivery System, and in many cases, they will contract for system services. Any platform used to deliver Smarter Balanced assessments should be verified to deliver assessment items with authenticity, score items and tests properly, and deliver assessment results to the Smarter Balanced Data Warehouse according to specifications.

To aid with this effort, an Implementation Readiness Package (IRP) is being developed. IRP is a blend of software test harnesses and written specifications that States and their vendors can use to guide and inform their Assessment Delivery System compliance testing. The purpose of this document is to describe the IRP, and to identify the specific functional requirements of the Assessment Delivery System that it will assess.

Figure 1 below depicts the features of the Assessment Delivery System. Each circled letter in yellow represents a capability for which an IRP module *can* be developed. However, as of this writing, IRP modules are currently underway for items A, B, and G only, as these items address the required *external* capabilities of the Assessment Delivery System to function within the Smarter Balanced framework. An IRP module for item D (Adaptive Engine) is planned as well, but will be released at a later time.

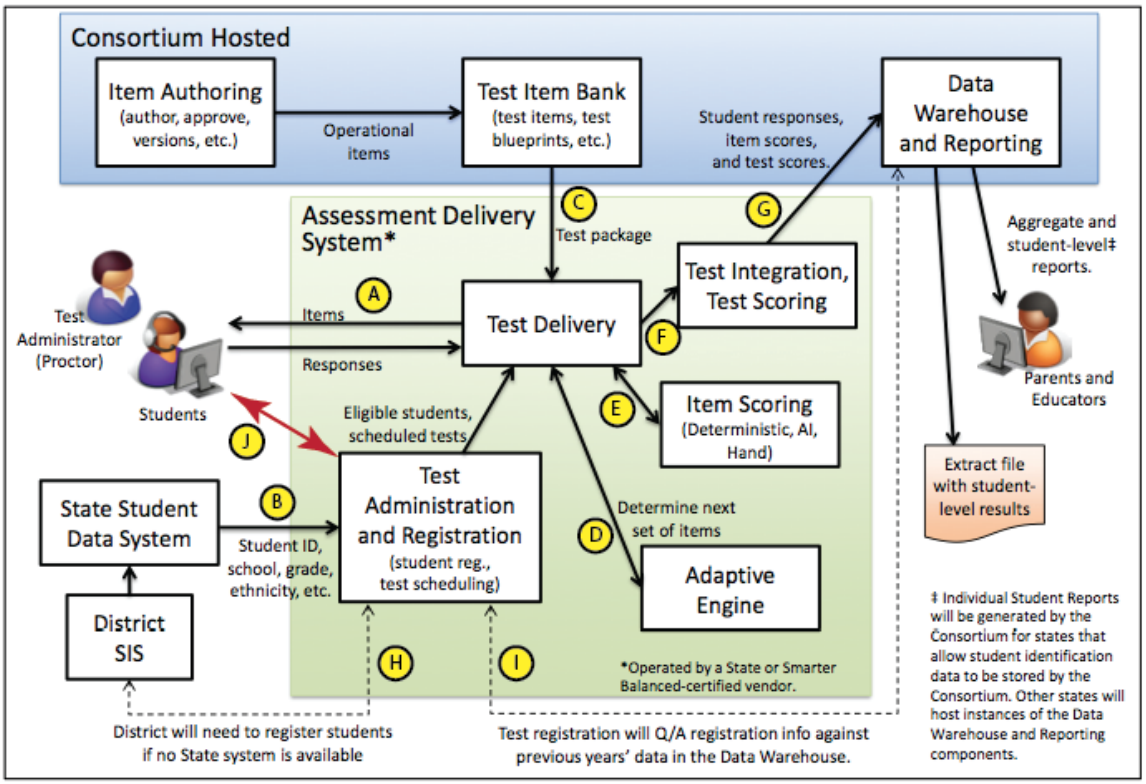

Figure 1: Features of the assessment delivery system

Development – Features & Phases

IRP will be developed in two phases. Phase I (IRP version 1.0) will include the items indicated in Table 1 below, and will be completed by 12/31/2014. Phase II (IRP version 2.X) will include the other more advanced features, as indicated in the table. The versions of IRP Phase II will be rolled out in stages as features become available, with the first version expected in spring of 2015.

Table 1. IRP Version 1.0 Features

IRP Module	Description	Interface	Functional Requirements Tested
A	Compliance of test item delivery to the student.	*IRP Version 1.0* • Manual download of Student Assessment Item Responses from IRP-hosted site	a. Correctness of item rendering (display) and associated interaction on screen (human visual test) b. Capability to manually input and capture student responses to assessment items
G	Compliance of delivery of TDS Output file to the Data Warehouse	*IRP Version 1.0* • Manual upload of TDS Output files to IRP server.	a. Capability to score student responses to individual assessment items b. Compliance with TDS Output file XML structure c. Correctness of data: ○ Student information ○ Item scoring

Table 2. IRP Version 2.x - Planned Features

IRP Module	Description	Interface	Functional Requirements Tested
A	Compliance of test item delivery to the student.	*IRP Version 2.X* • Web Service to send requests (as a sample student) for test items, and to receive test item (response).	a. Correctness of item rendering (display) and associated interaction on screen (human visual test) b. Correctness of test/test items delivered to student, given Student Data (provided by the state), Eligibility Requirements, and test data contained in the Test Package c. Capability to receive student responses to assessment items
B	Compliance of receipt of Student Registration data, and Student Accommodations data.	*IRP Version 2.X* • Web service to deliver Sample Student Data (function will be similar to authentic retrieval of Student Data from State Data System)	a. Receipt verification of School Data, Student Registration data, and Student Accommodations data.
G	Compliance of delivery of fully scale scored XML documents to the Data Warehouse	*IRP Version 2.X* • Web service to receive TDS Output xml file(s) from Vendor (will simulate API to Data Warehouse)	a. Capability to score student responses to individual assessment items b. Compliance with TDS Output file XML structure c. Correctness of data: ○ Student information ○ Item scoring ○ Test Scoring
D	Analysis of Adaptive Engine	*IRP Version 2.X* • Web Service to interface with vendor system, involving student responses to assessment items and resulting adaptations.	a. Efficiency and accuracy of vendor adaptive engine analyzed and compared against "gold standard" adaptive engine using known psychometric properties.

IRP Components

The Implementation Readiness Package currently focuses on two key aspects of the Vendor Assessment Delivery System: Rendering & Interaction, and Scoring.

Rendering & Interaction

Each Vendor Assessment Delivery System must adhere to the requirements outlined below for item rendering on the client machine. These requirements are meant to ensure that the presentation, interaction, and student accommodations associated with an individual item are consistent (within acceptable bounds), regardless of the device type or platform a

student is using. This requirements list is based on the Smarter Balanced Style Guide: Global Style Conventions, found at: http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/StyleGuide/StyleGuide.pdf

As of this writing, the Implementation Readiness Package will *NOT* explicitly (via software interaction) analyze and assess compliance with the Rendering & Interaction requirements listed. Therefore, it is up to states and their developers to incorporate these requirements into their own internal QA testing, and to verify compliance before releasing their system for field use. NOTE: Standards for compliance are part of IRP with reference to Tables 1 and 2 above.

Rendering & Interaction Requirements:

1. Content panes
   - Students should have the option of viewing content in one pane that is the full size of the computer screen, or in sub-panes that divide the screen horizontally or vertically.
   - If a student takes a test on a device that cannot show all the elements of an item (e.g., an iPad device), provision must be made for the student to, at minimum, be able to toggle back and forth between the prompt and answer choices.
2. Scrolling
   - Students should not have to use horizontal or vertical scrolling to view test content in its entirety. The following guidelines are designed to minimize the need for scrolling:
     - Lay out content across the computer screen rather than in columns.
     - Display each item on the screen by itself.
     - Allow students to view a selection on the screen by itself or on the screen with an item. This eliminates the electronic equivalent of turning a page to flip between the selection and the items.
     - Similar options should be provided for viewing a cluster (a group of items associated with the same graphic). Students should be able to view the graphic on the screen by itself or on the screen with an item in the cluster.
3. Items should render in such a way as to clearly provide the needed information related to an item, including:
   - Any graphics or multimedia items are viewed with the correct images or video
     i. Pictures should be drawn to scale where required, or explicit mention of the picture or diagram not being drawn to scale must be captioned
     ii. Audio will require headphones and the volume should be easily adjustable by the student

iii. Text: Smarter Balanced is currently specifying the use of Verdana as the primary font for test content and by default should be black on a white background

Table 2. General font and alignment specifications for text elements in test forms.

Font and Alignment Specifications		
Content	**Font**	**Alignment**
Items	14 pt. Verdana	Left aligned
Part headings in items	14 pt. Verdana Bold Italic	Left aligned
Boxed text	14 pt. Verdana	• Box: left aligned • Text: longest line centered in box; other lines left aligned on longest line
Emphasis terms	14 pt. Verdana Bold	n/a
Directions	14 pt. Verdana Bold	Left aligned
Purpose-setting statements (PSSs)	14 pt. Verdana Italic	Left aligned

Note: At this time, Verdana is specified as the primary font for test content. However, another font may be chosen upon further analysis of the effects fonts have on readability and students' ability to retain information.

- Prompt is to be clearly viewable
- Answer options (choices) render properly (see Table 3 for alignment and order of option in the selected response (SR) items)

Note: Options derived from a stimulus, such as a selection or graphic, are ALWAYS arranged in the same order they appear in the stimulus. This guideline supersedes all other guidelines listed in Table 3.

Table 3. Guidelines for the alignment and order of options in SR items.

Option Alignment and Order			
Option Type	**Alignment**	**Order**	**Example**
Graphic options	• Graphic: left aligned • Option letter: top aligned or vertically centered on graphic (use best judgment)	Arranged for best visual presentation (use best judgment)	n/a
Numeric options *(continued on next page)*	Decimal aligned: • stand-alone numbers • decimal values • numbers that precede or follow symbols: 40°, $20.00 • numbers that precede labels: 6 ties, 12 bananas • numbers that precede units of measure: 15 inches, 30 cm	Arranged in ascending or descending order	What percentage of students prefer strawberry yogurt to blueberry yogurt? A. 25% B. 50% How long, in inches, is each necklace? A. 9 in B. 12 in
	Currency: • dollar signs: left aligned • dollar amounts: decimal aligned	Arranged in ascending or descending order	How much money did Chip spend altogether? A. $ 80 B. $120
	Fractions: • Fraction: left aligned • Option letter: vertically centered on fraction	Arranged in ascending or descending order	How many cups of sugar does Janet need to make two cakes? A. $\frac{1}{2}$ B. $1\frac{1}{4}$

Table 3. Guidelines for the alignment and order of options in SR items.

Option Alignment and Order			
Option Type	**Alignment**	**Order**	**Example**
Numeric options *(continued)*	Times of day: • Decimal aligned on colon	Arranged in ascending or descending order	At what time does Kody eat lunch? A. 11:30 a.m. B. 1:00 p.m.
Text options	Left aligned	Words: arranged in alphabetical order	What does the word pacify mean in the sentence highlighted in the passage? A. greet B. soothe
	Left aligned	Phrases and sentences: arranged by length, longest to shortest or vice versa	Which statement is true about a triangular prism? A. It has six vertices. B. It has four triangular faces.

Note: Options are not arranged in the prescribed order when doing so clues the answer to the item.

- Whenever possible, arrange graphic options in a two-over-two box layout.
- Follow graphics specifications as found in the Graphics Specifications (Part V) of the Smarter Balanced Style Guide.

4. Interactions (Technology-Enhanced Items)
    - Follow specifications as found in the Technology-Enhanced Items (Part VI) of the Smarter Balanced Style Guide.
    - Any answers that require user interaction, functions as intended (e.g., clicking and dragging on a bar chart moves the bar to the value specified by the student)
5. Accommodations
    - Some item characteristics (like background color) are less important unless it's part of a given accommodation (like high contrast).
    - Shapes/colors of buttons are less important than the ability to read the text on the buttons
    - If a student has an accommodation that is related to screen colors like high contrast, the color choices must all have enough contrast so that the student can easily read text and answer choices.
    - If a student has access to foreign language glossaries, appropriate fonts are available for the words to appear correctly.
    - Correct video codecs are available on all devices that are used to render the video for prompts and accommodations (e.g., American Sign Language videos)
    - For read aloud accommodations, need to have speech to text capability with properly working headphones, or an adult reader.
    - All of the Embedded Accommodations listed in Table 5 of the Smarter Balanced Usability, Accessibility, and Accommodations Guidelines Document, located at: http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/08/SmarterBalanced_Guidelines.pdf
6. Rendering and interactions must be supported on the following test-taking devices / secured browsers:

Test-Taking Devices and Approved Secure Browsers

For latest version, go to http://www.smarterbalanced.org/test-taking-devices-approved-secure-browsers/

Smarter Balanced requires that testing devices meet certain minimum requirements and that they run an approved secure browser. The official Smarter Balanced policy regarding testing devices is encompassed in the Technology Strategy Framework and Testing Device Requirements. Below is a summary of universal requirements, the current list of approved secure browsers and the corresponding devices. Following those is a list of sunset dates after which certain operating system versions will not be supported.

Universal Requirements

The following requirements apply to all testing devices regardless of format, manufacturer or operating system.

Screen	Display must measure a minimum 9.5 inches diagonal (sometimes described as “10-inch class”). Resolution must be a minimum of 1024×768.
Headphones / Earphones	Headphones or earphones must be available to students for use during the English language arts test and for students who require text-to-speech features on the mathematics test.
Keyboard	A physical keyboard is required to avoid consuming screen space that must be available for test content.
Pointing Device	A pointing device must be included. This may be a mouse, touch screen, touchpad, or other pointing device with which the student is familiar.
Network	Must be connected to the internet with a minimum of 20Kbps available per student to be tested simultaneously.
Security	It must be possible to secure the device so that the student does not have access to unauthorized web sites or applications. This is accomplished through the use of an approved secure browser.

Secure Browsers and Scope of Support

The following secure browsers will be released in the fall of 2014. Following release, this page will be updated to include download and installation instructions for each.

Secure Browser	Scope of Support
Secure Browser for Microsoft Windows	Windows XP Service Pack 3 **(See sunset date below)** Windows Vista; Windows 7; Windows 8 and 8.1; Windows Server 2003; Minimum Hardware: Pentium 233 Mhz processor; 128 MB RAM; 52 MB hard drive free space. Recommended Hardware: 1Ghz processor; 1GB RAM; 1GB hard drive free space. *Virtual desktop systems running versions of Microsoft Windows require additional assurance that the virtual desktop client (the device in front of the student) has been secured to prevent unauthorized access to applications or websites.*
Secure Browser for Apple Mac OS X 10.5-10.9	Mac OS X 10.5 to 10.9 Minimum Hardware: Macintosh with Intel 300Ghz Processor; 256 MB RAM; 200MB hard drive free space.

Secure Browser	Scope of Support
	Recommended Hardware: Macintosh with 1Ghz processor; 1GB RAM; 1GB hard drive free space.
Secure Browser for Apple Mac OS X 10.4-10.5	Mac OS X 10.4.4 to Mac OSX 10.5 Minimum Hardware: Macintosh with Intel x86 or PowerPC G3 300Ghz Processor; 256 MB RAM; 200MB hard drive free space. **(See PowerPC sunset date below)** Recommended Hardware: Macintosh with 1Ghz processor; 1GB RAM; 1GB hard drive free space.
Secure Browser for Apple iOS (iPad)	iOS 7.1 iPad 2, iPad 3rd Generation, iPad 4th generation, iPad Air As with all devices, a physical keyboard is required for use with iPad devices. The iPad mini is unsupported due to inadequate screen size.
Secure Browser for Google Chromebook	ChromeBooks running Chrome OS 31-34
Secure Browser for Linux	Ubuntu 9 to 12 Fedora Core 6 to 16 Minimum Hardware: Pentium II or AMD K6-III 233Mhz Processor; 64MB RAM; 52 MB hard drive free space. Recommended Hardware: 1 Ghz Processor; 1GB RAM; 1GB hard drive free space.
Secure Browser for Android	Select Android devices running Android OS 4.0.4 to 4.2 The secure browser for Android tablets requires that the secure browser keyboard be selected before students access the login page. This is to disable the keyboard predictive text which would unduly aid students when entering written responses.

Sunset Dates

Support for certain operating systems and associated devices will expire at the end of the school year according to the following table.

Operating System	OS Release Date	Smarter Balanced End of Support Date
Mac OS 10.4.4	January 2006	**End of 2014-2015 School Year**
Mac OS 10.5 (Power PC)	October 2007	**End of 2014-2015 School Year**
Mac OS 10.5 (Intel)	October 2007	End of 2016-2017 School Year
Mac OS 10.6	August 2009	End of 2018-2019 School Year
Mac OS 10.7	July 2011	End of 2020-2021 School Year
Mac OS 10.8	July 2012	End of 2021-2022 School Year
Mac OS 10.9	October 2013	End of 2022-2023 School Year
Windows XP (SP 3)	April 2008	**End of 2015-2016 School Year**
Windows Vista	January 2007	End of 2016-2017 School Year
Windows 7	October 2009	End of 2019-2020 School Year
Windows 8	August 2012	End of 2021-2022 School Year
Windows Server 2003	April 2003	End of 2015-2016 School Year
Windows Server 2008	October 2008	End of 2018-2019 School Year
Fedora Core 6	November 2007	End of 2016-2017 School Year
Ubuntu 9-12	October 2009	End of 2018-2019 School Year

Operating System	OS Release Date	Smarter Balanced End of Support Date
iOS 6	June 2012	TBD
iOS 7	September 2013	TBD
Chrome OS	Rolling Release	TBD

Scoring

The Scoring component of IRP focuses on an Assessment Delivery System's ability to synthesize: a) student data; b) students' responses to assessment items (item scoring); c) test scoring*; and d) results packaging (for the Data Warehouse).

The overall idea is that a set of *sample students* each complete *one or more tests*, with associated *item responses* that are then processed through the Assessment Delivery System's scoring engine to yield a *Test Score Batching xml file*, suitable for transmission to the Data Warehouse, as specified in the Smarter Balanced document: Reporting Data Specification Format Test Score Batching, located at: http://www.smarterapp.org/documents/DataWarehouse-Spec-ScoreBatching.pdf.

In this case, though, rather than sending to the Data Warehouse, the vendor will upload the Test Score Batching xml file to a web-based, IRP Test Scoring module, where it will be analyzed for compliance with the following:

I. Item scoring accuracy
II. Test scoring accuracy*
III. Completeness and accuracy of student data
IV. File structure (xml schema and data) for integration with the Data Warehouse

*NOTE: Test Scoring will not be available in IRP Version 1.0.

Figure 2, below, summarizes the Test Scoring process, and what follows is description of each of the steps listed.

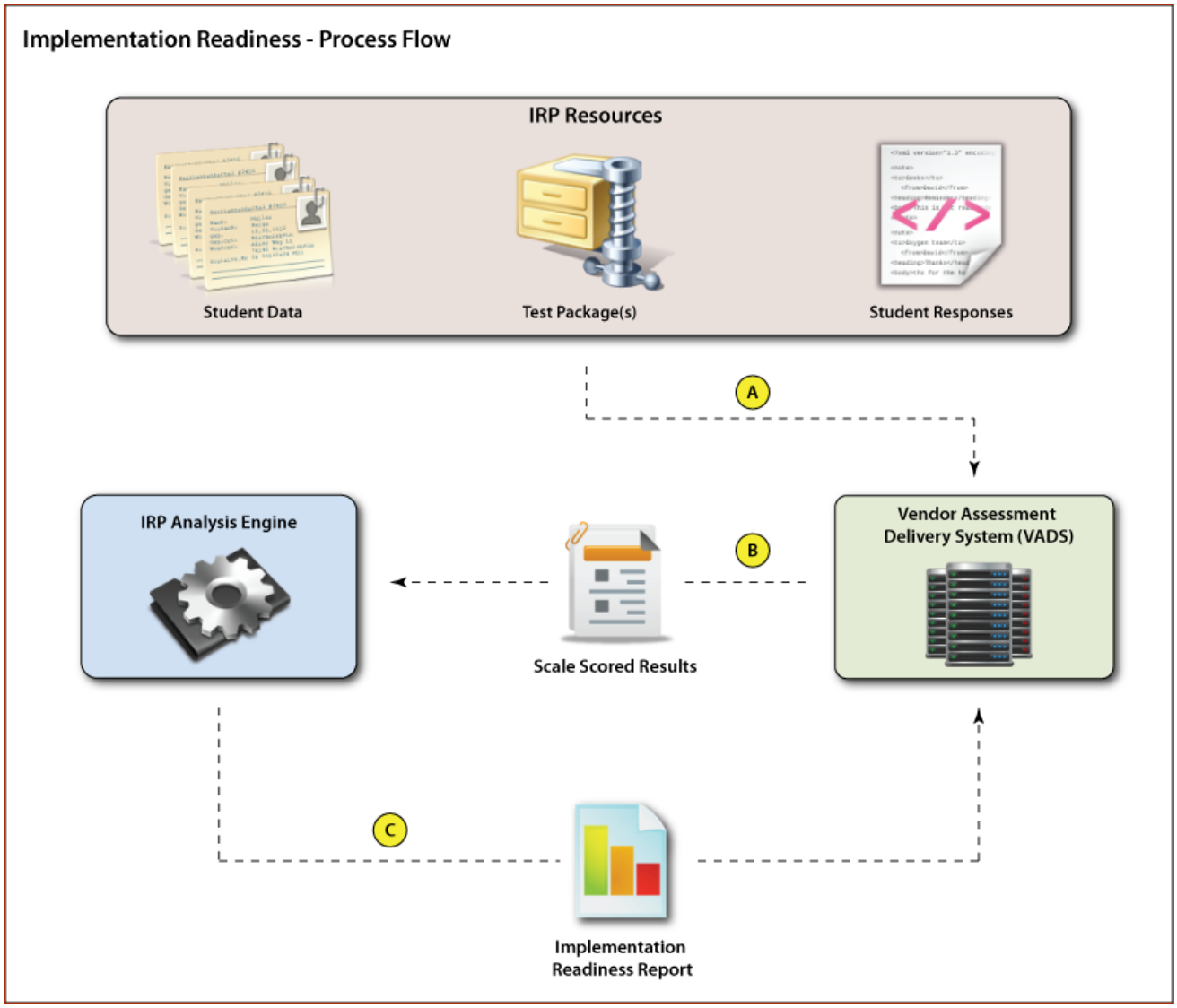

Figure 2: Implementation Readiness - Test Scoring Process Steps

Step A: Data Acquisition

From an IRP-hosted website, the vendor will manually download the following:

- Sample student data
- One or more test packages
- One or more scoring packages (corresponding to each test package)*
- An xml file of student responses to each of the assessment items.

*NOTE: Test Scoring will not be available in IRP Version 1.0.

Sample Student Data

The sample student data is a .csv file that conforms to the information/data fields specified in the Smarter Balanced document: *Administration and Registration Tools User Guide*, Table

16, pages 64-68, which can be found here:
http://www.smarterapp.org/documents/Administration_and_Registration_Tools_User_Guide.pdf .

This .csv file contains a list of sample students to be used in the IRP Test Scoring analysis. The requirement for vendors is that these virtual students must be integrated into their Assessment Delivery System in such a way to satisfy the following requirements:

1. Each student is regarded by the system as having *completed* the one or more tests downloaded from the IRP-hosted website.
2. The *captured assessment item responses* – i.e. the item response data the Test Delivery System receives from the student (client) machine – must be artificially inserted such that each student's responses match those contained in the *Student Item Responses* xml file downloaded from the IRP-hosted website.
3. The vendor Assessment Delivery System must process each student's item responses through its internal scoring engine, and provide overall Test Scoring for each student.
4. These overall Test Score results must get reported out in a *Test Score Batching* xml file, suitable for transmission to the Data Warehouse

NOTE: The IRP does NOT dictate how the vendor accomplishes the above tasks, as each Assessment Delivery System will be different, with different methods and dependencies for integrating and processing students and their test results.

Test & Scoring Package(s)
The test and scoring packages downloaded will be consistent in structure and formatting of actual, in-service test and scoring packages, and will contain all required information for processing in an Assessment Delivery System. It is expected that each of the sample students (described above) complete each test of the downloaded test package(s).

Student Item Responses
An xml file of student responses to assessment items will be provided. For each student (present in the Sample Student Data file), this file will list their response to each presented assessment item. It is up to the vendor to associate these item responses (presumably via manual entry simulating each student at their client machine) with the sample students, and to process these responses in their system for Test Scoring and Test Score Batching.

Step B: Data Processing
With the IRP data received and processed, the assessment delivery system will next synthesize the student item responses and complete Test Scoring* and Test Score Batching. The Test Score Batching xml file will include all of the information required for storage in the Data Warehouse, as spelled out in the Smarter Balanced hosted document:
http://www.smarterapp.org/documents/DataWarehouse-Spec-ScoreBatching.pdf.

*NOTE: Test Scoring will not be assessed by IRP version 1.0.

Once ready, the Test Score Batching xml file will be manually uploaded to the web-based, IRP Test Scoring module for analysis.

Step C: Data Analysis

The Test Score Batching xml file will be analyzed by the IRP system for compliance with the following:

1. File format. Is the xml file properly formatted to meet the requirements for integration with the Data Warehouse?
2. Consistency & accuracy of student information. Has the student data been properly preserved, and is all appropriate student data provided (i.e. identifiable vs. de-identifiable information), according to the Data Warehouse requirements specifications?
3. Item Scoring. Was each student's response to each assessment item scored correctly?
4. *Test Scoring. Has the assessment delivery system properly scored the tests as a whole, according to existing rubrics and weightings?

*NOTE: Test Scoring will not be assessed by IRP version 1.0.

The results of this analysis will be disseminated to the vendor in a downloadable report, summarizing the Assessment Delivery System's degree of compliance across the four categories listed above.

Phase 2: Planned Future Development: Architecture & Process Flow

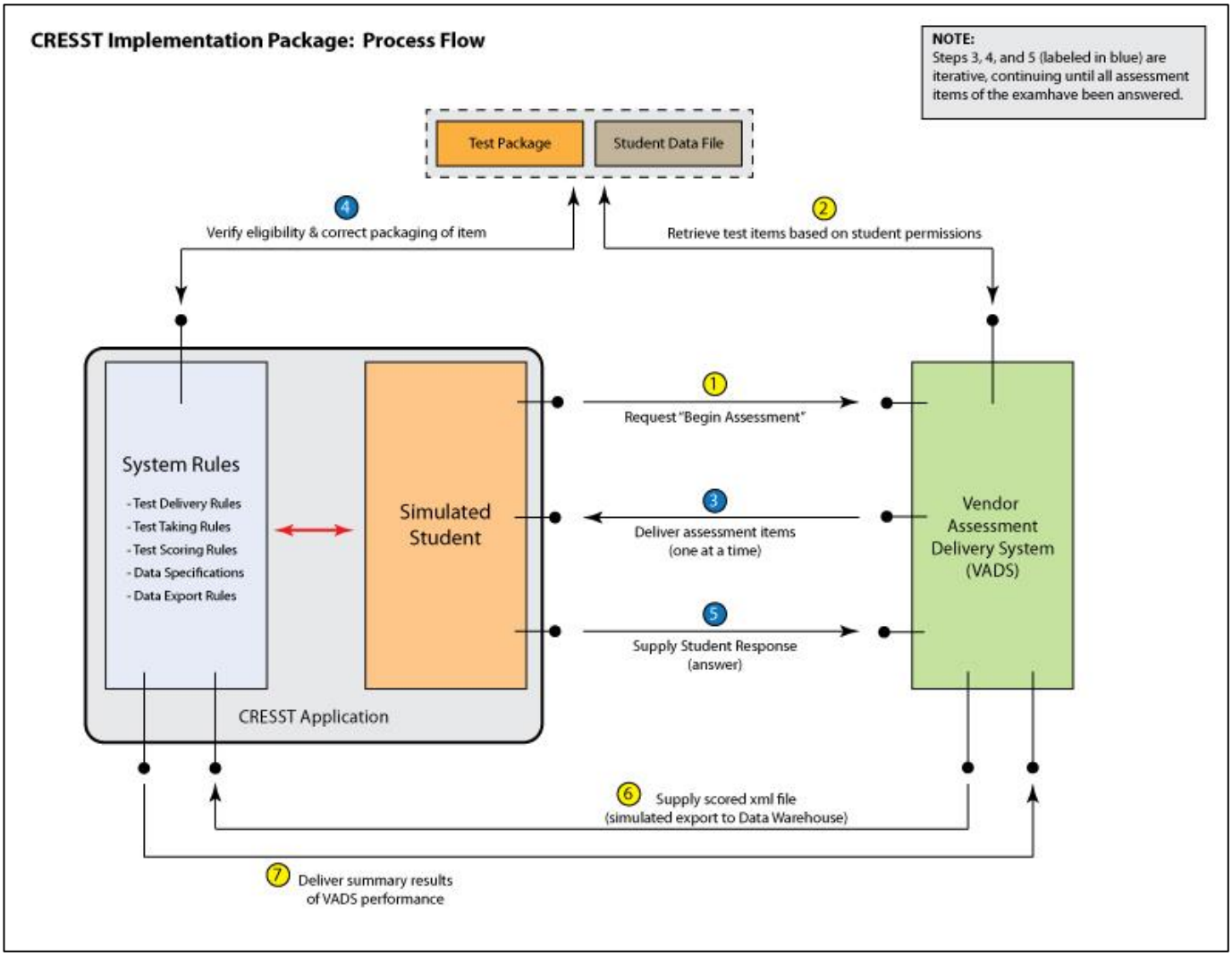

Figure 3: Implementation Readiness Package: Architecture and Process Flow

Figure 3 above shows the architecture and process flow of the planned future IRP model (projected for late 2015), which is a more automated version of the current IRP. This system will include a collection of web services through which the Vendor Assessment Delivery System (VADS) can interface. The purpose of this system is to simulate and test the functions required for administering an assessment, beginning with the initial request from the student (client) machine to start the test, to the final data export to the Data Warehouse. Prior to commencing the simulation, the VADS will be supplied with a test package and student data file for use in the test simulation. These same data files will be used by the CRESST Application to evaluate the performance of the VADS, ensuring that all source data are consistent.

Each step of this simulated assessment administration is numbered in the diagram, and is described in detail below. Note that steps 3, 4, and 5 are repeated iteratively as the simulated assessment is in process, with each occurring for each assessment item delivered.

Description of Process Steps

Step 1: After the VADS has initiated connection with the CRESST Application, and has received the test package and student data file, the CRESST Application will initiate a request (simulating a student at a client machine) to begin the assessment.

Step 2: The VADS will receive the "Begin Assessment" request and internally reference the test package and student data file to determine the items (and sequence for delivery) to be administered to the student. Note: this will require the VADS to internally process student eligibility, accessibility, accommodations, etc., according to the SBAC Test Delivery Specifications.

Step 3: The VADS will deliver (one at a time) the assessment items to the CRESST Application.

Step 4: The CRESST application will analyze the received assessment item for item eligibility, item packaging completeness (i.e. are all associated images, text, etc. present that should be), and item accommodations (i.e. language, audio, etc.). NOTE: The CRESST Application will NOT be able to determine the item's adherence to SBAC rendering requirements. Rather, this is a human-conducted, visual test that the vendor must perform separately.

Step 5: The CRESST Application will provide a student response (i.e. answer) to each assessment item, as fully correct, partially correct, or fully incorrect, based on the specific test case being run, and will internally keep track of all supplied responses.

Step 6: At the conclusion of the simulated assessment, the VADS will internally score and package the test results, and export them as an XML file to the CRESST Application for analysis (NOTE: this simulates the VADS export to the Data Warehouse). This file will include all pertinent student data, individual item scoring, and aggregate item scoring, as well as all pertinent test parameters and session statistics, as detailed in the SBAC TEST INTEGRATION TO DATA WAREHOUSE specification.

Step 7: The CRESST Application will analyze the test results file (described in Step 6) for adherence to SBAC specifications regarding formatting, content completeness, and content accuracy. The results will be packaged into a summary performance report that will present the VADS compliance with SBAC specifications for item selection, item delivery, scoring, and results packaging.

Test Cases

The CRESST Application will enable a variety of test cases to be run, ranging from different student types (i.e. grade, accommodations, etc.), to different content areas, different item types, and different blends of correct vs. incorrect student responses within a given assessment. Variations in test cases will provide contrasts to enable the CRESST Application to evaluate various VADS capabilities regarding IRP Modules A, B, C, and G.

The Implementation Readiness package will provide an interface for the VADS to configure test cases prior to the commencement of a simulated assessment.