Assessment Interoperability Framework

Definition and Requirements Document

June 2012

Table of Contents

Background

The Race to the Top Assessment (RTTA) Grant Program presents some grand challenges for assessment systems. This historic funding has the promise to better enable the entire teaching and learning process through a variety of assessment types; enabling the formative assessment process; supplying better data to make more informed instructional decisions; providing insight as to professional development needs and providing data to inform decisions around teacher and administrator effectiveness.

One of the critical components when building comprehensive assessment systems involves the various technical considerations and the advantages of utilizing open technical standards. These standards, when used effectively, can provide a common technical manner to exchange data and information between all of the software systems within the education enterprise – schools, districts, state departments of education, Learning Management Systems (LMSs) and Test Publishers.

The SIF Association (SIF) and IMS GLC (IMS) communities, in partnership with the SBAC and PARCC consortia, have joined forces to develop a standards-based technical solution in support of the Race to the Top Assessment Grant Program for deployment in states and schools.

The AIF work is being undertaken to directly support the needs of the RTTA Consortia and the RTTA Program requirements. Ideally this work will inform the continually maturing CEDS work, but AIF work inclusion in subsequent CEDS versions is determined by the established CEDS approval mechanisms. Conversely, CEDS suggestions for AIF will be considered in all AIF deliverables, but not a mandated component in AIF decision-making processes.

Purpose of this Document

This document outlines an initial blueprint for the implementation of specific educational technology and assessment standards referred to as the "Assessment Interoperability Framework" or AIF. The AIF working group sponsored jointly by SIF and IMS and supported by NCES's CEDS program will use this initial blueprint to detail out specific implementation plans and documentation that can be used by any assessment implementation providers supporting RTTA.

While the collaboration effort focuses on RTTA and the assessment program architectures being developed by the two consortia, PARCC and Smarter Balanced (SBAC), it is expected that the AIF can be applied to a wide variety of assessment programs and applications. It is also the hope of the AIF members that this discussion also does not end at assessment. There are many opportunities for broader collaboration of standards groups across the education market. Ideally the AIF work will become the model or the incentive for greater collaboration.

It is important to note that this implementation will be a phased approach. It is recognized that the current standards, in their current form, do not support all possible use cases that RTTA programs will require. However, by focusing standards on specific interoperability use cases, we can leverage the specifications and focus future development energies with minimal overlap, thus eliminating redundancy and market confusion.

Scope

What the AIF will include:

- A high-level interoperability architecture for an Assessment Platform and how the Assessment Platform integrates with the broader education systems enterprise.
- Identification of cross-standard interoperability alignments or transformations necessary for data and content to flow through the assessment platform and to other consuming or providing systems.

What the AIF will not include in Phase 1:

- Recommendations for assessment standards interoperability for record exchanges or transcripts.
    - There is typically a very small subset of assessment result data elements that are used for these purposes.
    - It is expected that existing standards can support these activities.
- Recommendations for SLDS standards implementations.
    - Currently outside of the scope of the AIF.

Timelines/Deliverables

Phase 1

Phase 1 is expected to take 6 to 9 months to complete. Some of the primary deliverables for this phase are:

Deliverables/Activities	Time Frame	Who
An agreed upon AIF diagram, definitions, scope, and standards recommendations	1 – 2 months	AIFWG
An independent collaboration site to be used by the AIFWG for documentation, discussions, and reviews.	1 month	SIF
Use cases developed – the AIF will adopt the SBAC and PARCC as a starting point, will perform a gap analysis of the use cases and extend as necessary.	SBAC – 1 month PARCC – 1 month	AIFWG

Identify collaboration efforts (most should be identified in this document), develop a strategy, and identify teams to work on them. Some likely candidates are:
• PNP and SIF for registration capabilities
• Content to Data
    ◦ Learning Standards
    ◦ Sub-test definitions
	1 – 2 months	AIFWG

Clarify the use of APIP terminology to the market	1 month	IMS
Identify changes to existing standards to support AIF and provide recommendations to the existing working groups	6-9 months	AIFWG
Develop new standards, as applicable, to support AIF	6-9 months	AIFWG
Set of documentation that describes/defines how a solution should be implemented.	6-9 months	AIFWG

Deliverables/Activities	Time Frame	Who
Develop a working prototype that demonstrates key interoperable components.	6-9 months	TBD – Solution Providers are needed

Note: A preliminary schedule is available on the AIF collaboration site.

Phase 2

While Phase 2 has not been defined in any detail, it is expected that phase 2 will start to address the lower priority items that were not addressed in Phase 1.

Phase 2 could also include more clearly defining data, profiles, content, or transport standards that were exposed as weaknesses during Phase 1 implementations.

Phase 2 may extend scope beyond SIF and IMS standards.

Interoperability versus Comparability

The AIF focuses on interoperability of content and data so that solutions providers can exchange content and data effectively and connect system components together seamlessly. Comparability generally refers to assessments being taken across various "administration variables" and using psychometric data to ensure that the results can be compared. Some of the various types of comparability are discussed in Appendix A. It is important to understand that interoperability does not directly translate into comparability.

From an interoperability standpoint, the interoperability standards must support the required data (item performance data, scoring data/rules, and outcome data/results) to support psychometric analysis and comparability studies. As part of the AIF, we will profile the existing standards and identify gaps in this area.

It is important to note that the interoperability standards themselves will not ensure assessment instrument comparability.

Assumptions

1. There is an expectation through the RTTA program that major components be interoperable to allow end users to make different solution provider decisions over time. As such, interoperability is most applicable when multiple solution providers are contributing to the solution. In some instances, a specific implementation may leverage two or more components from the same solution provider and the connections between those components may continue to be proprietary to optimize performance or functionality and specific interoperability not defined.
2. While the focus of this framework is on RTTA, it is expected that this framework will support all types of assessments including summative, interim, and formative.

3. For the first phase of this framework, it is assumed that an assessment registration function will exist in all cases. This implies that student demographic data and personal preference data is collected and stored within the overall assessment platform solution set. In the future, formative platforms may leverage local information systems more extensively to provide this functionality.

4. For the first phase of this framework, the transport layers of the implementation are not dictated. However, it is expected that where transport standards exist (such as SIF transport) that implementations will leverage them where applicable.

Risks/Issues

1. No current standards meet all use cases. Industry players may be motivated to create new standards to "bypass" existing standards and potentially invalidate this document's recommendations.

    a. Because of this fact, will partial successes (wins) be acceptable to the market? This group should be clear about current limitations and communicate those clearly to users to set expectations. The collaboration will continue to help guide/push existing standards forward to overcome limitations.

2. Backwards compatibility is a strong motivator for updating existing standards. This effort should be conscious of the limitations this may impose on standards to move quickly or make breaking changes. However, if breaking changes are necessary to enable new innovations, technologies, or approaches, we should not hesitate to suggest them.

Assessment Interoperability Overview

The following diagram illustrates the highest level assessment platform components and the external local, regional, and state systems that provide information to or consume information from the assessment platform. This AIF diagram also indicates the recommendations from the working group for selected industry standards to connect the components together.

It is important to note that all diagrams or illustrations in this document are "functional" in nature and do not represent (or recommend) a physical implementation. The functional models show the logical components of any system. These do not represent how the components may be physically bundled or packaged into a specific implementation. In fact, not all components may exist in any given implementation. In addition, components may be combined into a single element of a physical implementation for performance or optimization considerations.
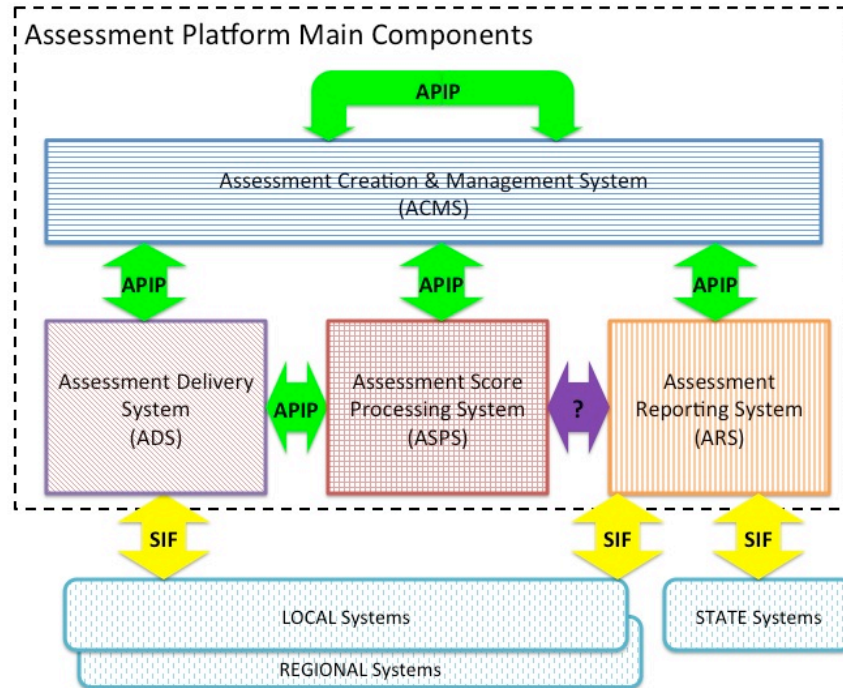
Note: Arrow color legend is provided in the Interoperability section below.

From Content to Data

As illustrate by this diagram, we are generally utilizing two main interoperability standards, APIP and SIF. At the highest level, we are using APIP to represent all assessment content and operational scoring information and we are using SIF to collect registration data and to represent scoring results for reporting and distribution. Therefore, it will be critically important that all content structures that are used to drive delivery and scoring can generate the expected results for representation by SIF. While SIF provides a data model and transport, the focus for the Phase 1 work is on the data model and the transport is optional.

The Assessment Platform Components

The Assessment Platform will consist of four super-components:

1) Assessment Creation and Management System (ACMS),
2) Assessment Delivery System (ADS),
3) Assessment Score Processing System (ASPS), and
4) Assessment Reporting System (ARS).

These super-components are needed to complete an Assessment Platform. These components and their sub-components are discussed in the definitions section below. The Assessment Platform will also interact with various State and Local information systems as part of the complete solution.

External Components

The Assessment Platform will interact with many state, local, and/or regional systems. While it is highly likely that many common external system components will exist across all states and districts, this interoperability specification does not attempt to dictate which are used to complete the assessment enterprise. A well-formed interoperability standard should not require specific types or categories of systems to be in place as long as the end points can adhere to the interoperability standard data or content format. For example, while it is likely that all districts will have some form of a student information system (SIS) and that system will likely be the source of student assessment registration data, this framework does not require the use of an SIS for registration. Registration data could also come from an LMS or reporting solution within the district or be provided by the region or state on behalf of the district.

Wiring Diagram

The following diagram is provided to identify the next level of detail of the assessment platform and how interoperability standards will be used to "wire" the sub-components together.

The arrows between the components represent the focus areas for interoperability. It is expected that users will ultimately select components from multiple solution providers to meet specific needs and to leverage solution sets.

It is expected that the inner-workings of each component will be where solution providers can distinguish themselves in the market by providing unique and innovative functionality. However, with well-defined interfaces, solution providers can continue to innovate while supporting the desired interoperable (plug-and-play) environment.

AIF Wiring Diagram
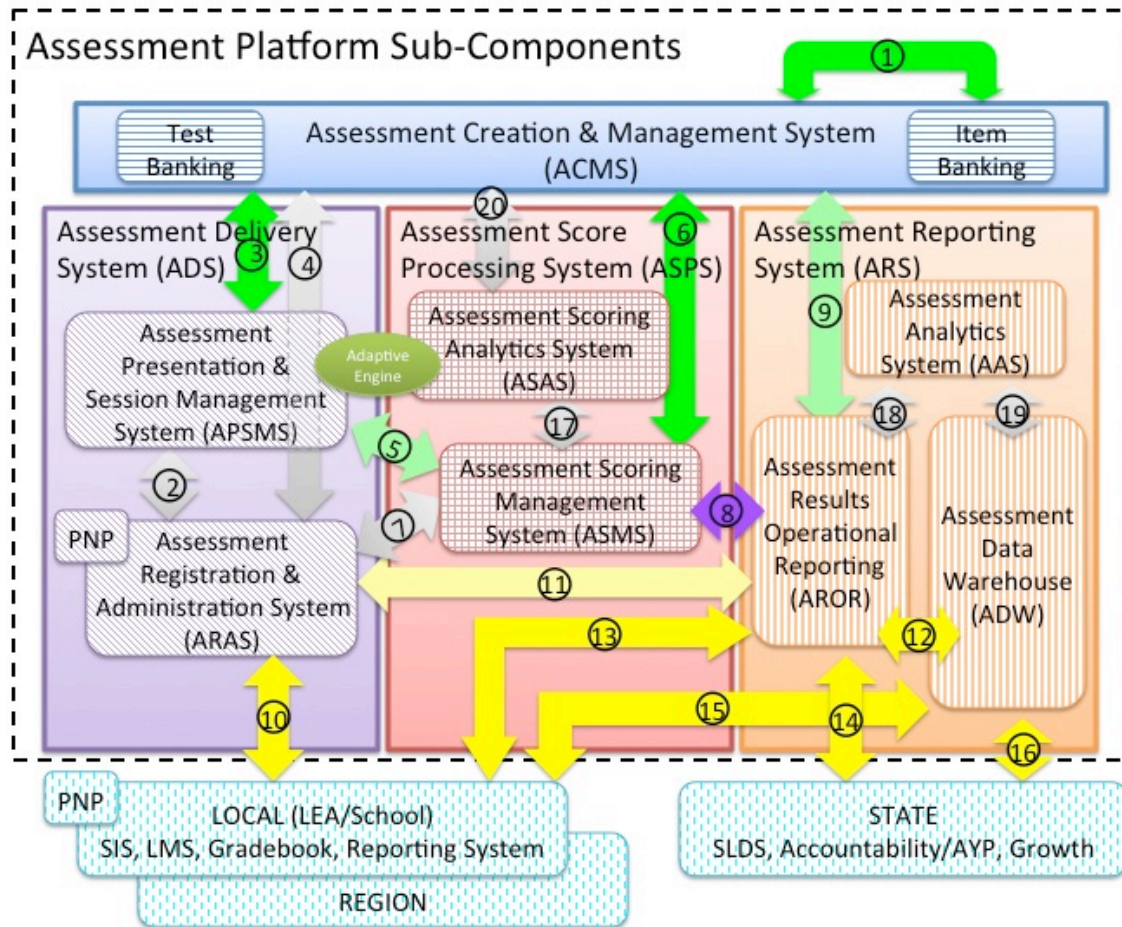
Interoperability (the Arrows)

The arrows between the components represent the areas where interoperability standards will likely be applied. All arrows are represented in the diagram as bi-directional for illustration only. The arrow may actually represent multiple interactions and may be one-way. The arrows will be discussed in the Interoperability Requirements Statements section below.

The arrows are color coded and numbered to represent which current industry standard best fulfills the known use cases of that connection (in the opinion of the AIF working group).

Bright Green Arrows (#s 1, 3, 6) – Where IMS APIP Assessment Content standards are to be applied for content portability. The transport layer for these exchanges will likely use existing techniques and technologies. For arrow 1, this will likely be manually triggered and will utilize sFTP or other batch oriented transport technologies for Phase 1.

Light Green Arrows (#s 5, 9) – Where IMS APIP Assessment Content standards are likely to be applied but may not be a priority for phase 1 of this framework. Transport techniques/technologies will be determined in subsequent phases.

Bright Yellow Arrows (#s 10,12,13,14,15,16) – Where the SIF data model and the SIF transport standards are to be applied for data interoperability. Application of the SIF transport will be optional.

Light Yellow Arrows (# 11) – Where the SIF data model and optionally the SIF transport standards are likely to be applied but may not be a priority for phase 1 of this framework.

Bright Purple Arrows (# 8) – Determined to be a priority interoperability point in the framework but the small working group did not achieve consensus. Possible candidate standards for this interoperability point are:

1) The IMS QTI Results Reporting Standard
   (http://www.imsglobal.org/question/qtiv1p2/imsqti_res_bestv1p2.html) and
2) The SIF Assessment Reporting standards
   (http://specification.sifassociation.org/Implementation/US/2.5/html/AssessmentWorkingGroup.html#AssessmentWorkingGroup).

For a discussion on these options, see arrow 8 in the Interoperability Requirements Statements section of this document.

Light Gray Arrows (#s 2,4,7) – Not a priority for phase 1 and was not discussed in enough detail to determine best interoperability standard options.

Personal Needs and Preferences (PNP)

The PNP provides the detailed student needs and preferences for specific assessments. Assessment preferences may differ between subject areas, i.e. a student may have different preferences for Math and Language Arts. At this point, it is not expected that existing local information systems can provide the level of detail and variability by subject that is truly needed by PNP and assessment. It is expected that Phase 1 implementations will leverage the IMS PNP specification for designing specific registration systems but it is unlikely that this information can come directly from local information systems using the interoperability standards. It is also expected that the registration system will provide methods for importing or a user interface to collect the information. It is also important to recognize that PNP data as it relates to which accommodations where used by a student will likely be required output from the reporting system as part of the assessment results.

Assessment Scoring

Scoring is likely where some of the translation from APIP to SIF will occur – where we move from mostly a content representation to data representation. For this section, we are limiting the scoring discussion to scoring an individual assessment and not considering scoring a battery of assessments, calculating growth scores from multiple assessments over time, or any aggregation of scores such as class or school averages. Scoring is also where consensus on which standards to apply is more difficult to achieve by the working group.

It is important to note that the interoperability specification will not define how the internal workings of a scoring process will work but will define the input and output parameters (data) necessary to perform

the scoring process. As an example, the interoperability specifications would not define the specific algorithms used in an artificial intelligence scoring engine but they will define how the item (or prompt) is identified, the scoring rubrics are defined, and student response structures are provided as input to the scoring process. How the scoring process consumes or uses this information is entirely up to the solution provider to design. Scoring may include a variety of scoring methods including machine scoring (lookups or matching to keys), algorithmic or AI scoring, and/or human or professional scoring services.

In order to understand this in more detail, let's consider the specific inputs and outputs from the scoring process. The following diagram illustrates the major interfaces to a scoring process.
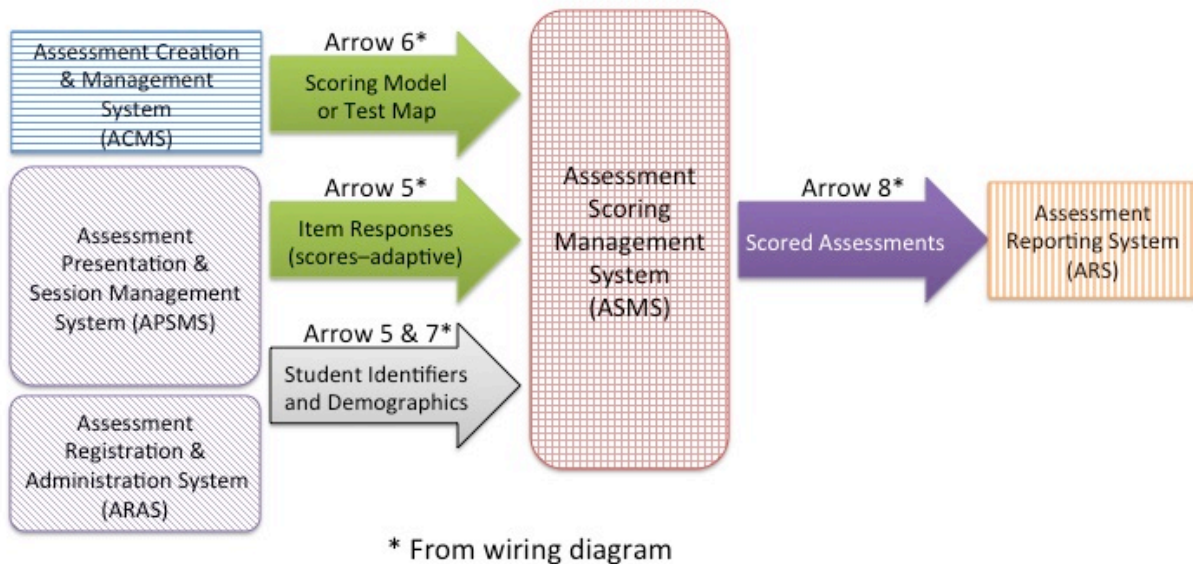
The three main inputs to the scoring process are:

1. Scoring Model or Test Map (note there may be other terms used to describe this structure) – This defines the "rules" by which items, sub-tests and total scores are calculated for an individual assessment. Generally, the psychometricians or the authors of the assessment define the scoring rules. It is expected that the scoring rules "live" with the assessment definition in the ACMS although in legacy systems, the scoring rules may be kept offline in other forms, such as a spreadsheet application. We will explore this structure further in the next section.

2. Item Response and Scores – After (or as) the student has tested, various response information is collected. This may be simply the response (ex: A/B/C/D of a multiple choice item) or other data about the interaction, such as time-on-task, tools used, clicks, etc. In addition, if the assessment is an adaptive assessment, then some amount of scoring occurs while the test is being delivered. This could include item scores, ability levels, confidence intervals, standard errors, etc. depending upon the specific algorithm. This information will come from the APSMS delivery system and be delivered to the scoring process.

3. Student Identifiers and Demographics – Prior to testing, the student data is likely collected through a registration process. Alternatively, some student data may be collected at time of test. Generally, the registration system will provide basic student data to the APSMS delivery system (generally enough to identify the student and which test form to administer). This information is typically passed through the delivery process to the scoring process. However, the scoring system may need additional information about the student that may come directly from the ARAS registration component.

   a. In addition to student data, the APSMS or the ARAS may also provide data about the administration or session, such as testing interruptions (ex: fire drill), suspicion of cheating, or other indicators about testing status that may influence the scoring process.
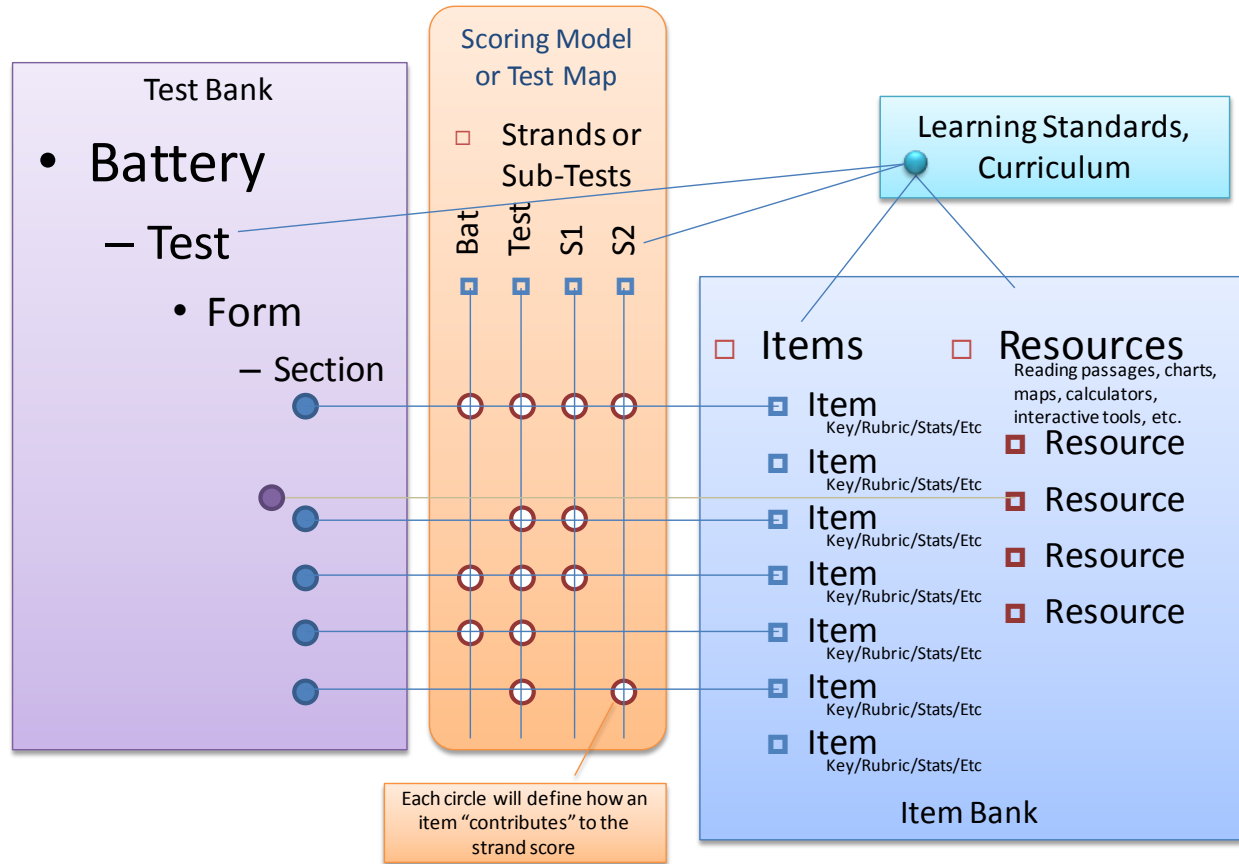
The output from the scoring process is the scored assessment. This will include:

- Total test scores – generally an assessment will provide an overall or total score. This may include a raw score, percent correct, scale score, percentile rank, grade equivalent, performance level, mastery, or many other types of scores.
- Sub-test or strand scores – it is common for an assessment to measure more than one standard or psychometric construct. In this case, the scoring system may derive one or more scores for each sub-test or strand. These scores can include most of the types listed in the total test score item above.
- Item responses and scores – It is common for the assessment to return each item response and the associated set of scores for that response. Some items may provide multiple scores. For example, an essay may return an overall score and one or more trait scores.
- Feedback – Some assessments may generate feedback to the student based on performance on the test, sub-tests, and/or individual items. Feedback may be "system generated" based on scores and standards being measured or it may be provided as part of a human scoring process (i.e. the scorer may provide comments or annotate responses). Feedback may be text based or references to supporting instruction or remediation materials.

Scoring Model or Test Map

The scoring model or test map defines how items are "grouped" to compose a total test or sub-test score. Items may contribute to scores in a variety of ways. Items may be weighted for each sub-test. Item scores may contribute to sub-tests based on correct or incorrect responses. The following diagram illustrates how a scoring model may appear.

Test Bank

- **Battery**
  - **Test**
    - **Form**
      - **Section**

Scoring Model or Test Map

□ **Strands or Sub-Tests**

Bat | Test | S1 | S2

Each circle will define how an item "contributes" to the strand score

Learning Standards, Curriculum

□ **Items**   □ **Resources**
Reading passages, charts, maps, calculators, interactive tools, etc.

■ **Item**
Key/Rubric/Stats/Etc

■ **Item**
Key/Rubric/Stats/Etc

■ **Item**
Key/Rubric/Stats/Etc

■ **Item**
Key/Rubric/Stats/Etc

■ **Item**
Key/Rubric/Stats/Etc

■ **Item**
Key/Rubric/Stats/Etc

■ **Resource**

■ **Resource**

■ **Resource**

■ **Resource**

Item Bank

In general, as a form is developed, items are selected from the item bank. Each item will "come with" its content (default and accessibility), answer key, scoring rubric, performance statistics and many other data or meta-data elements. Each item is then included in one-or-more strands or sub-tests for scoring purposes. As each item is included in a sub-test, addition information may be provided about how the item "contributes" to the sub-test score.

Historically, these three constructs; items, test definitions, and scoring models/test maps, come in different forms. Legacy systems may have:

- Exchanged items as PDFs (or even worse…hard copy) with item cards or spreadsheets to provide meta-data (item keys/stats/etc.)
  - Open-ended item scoring rubrics and supporting scoring materials were also delivered separately.
- Test definitions that would have provided the "structure" and order to the items, rules about navigation, resources (passages, tools, etc.) to include, etc. would be provided in spreadsheets or data files.
- Scoring models that defined how items are used to derive total and sub-test scores, lookup tables for scaled scores, performance levels, etc. may have been in spreadsheets or delimited data files.

While these strategies do work and continue to be used today, they are very error prone and require significant effort to ensure everything is "in sync" and quality is maintained. Ideally, the three constructs would be "integrated" with the content such that as each element is updated that all affected and related elements can also be updated automatically or identified/tagged as needing revision so that each consumer can be ensured that they have a complete and up to date representation of the assessment instrument.

Interoperability Requirements Statements

The following table lists the interoperability requirements for the AIF that will be addressed in Phase 1.

Identifier	Short Description	Details
Global 1	Profiling Standards	With most standards, there can be flexibility to interpret the standard in different ways or to implement the same structures in multiple ways. In order to achieve the highest levels of interoperability, consistent implementations and representations of common elements or structures must be considered. Generally referred to as "profiling", it is expected that the AIF will provide standard profiles that should be adopted by all RTTA participants who are contributing interoperable components to the Assessment Platform.
Global 2	Student Identifiers	The AIF will need to identify how student identifiers will be used across platforms and standards. For example, the SIF RefID works well within a SIF zone but loses value across zones or across platforms that have not implemented SIF. All [or most?] states have unique student identifiers assigned by the state. Today, these provide the main linking mechanisms for exchanging information about a student today. While error prone, this method is effective.
Global 3	Item/Test Identifiers	The AIF will need to identify how items and test forms are identified across platforms and standards. Generally, items are uniquely identified by the publishing organization. A customer may also have a semi-intelligent numbering system to identify item subject, grade, standard, etc. While potentially error prone, this method is effective and is widely used today.
Global 4	Learning Standards	The AIF will need to identify how learning standards will be identified across platforms and standards. In order to be able to report scores or performance against a specific standard by item or sub-test, the standard measured will need to be identified uniquely. Today, some standards publishing organizations or authoring organization (states, benchmark organizations, common core, etc.) generally provide an outline numbering scheme to identify line items within the standard. We must ensure that all levels within the outline structure can be identified and associated with a publishing organization or authoring organization.
Global 5	Code Sets	The AIF will need to identify common codes sets that are used across platforms and standards. Code sets are important for data elements that have a defined set of values (ex: gender may be male/female/unspecified). Alternatively, the AIF could dictate how code sets are cross-walked between standards if established code sets are already in place. Code sets have historically been problematic with all education standards.

Identifier	Short Description	Details
Global 6	Versioning (not Phase 1)	The AIF will need to identify a strategy for versioning content (items, tests, and scoring information). It will be necessary to be able to know what version of content and scoring information was used to evaluate any individual student assessment.
Arrow 1a	ACMS to ACMS item content portability	In a multi-provider item content development environment, one provider may send items to another provider for processing. The items may include content, scoring information, standards alignment information, and share content (reading passages). The items may include their default content as well as accessibility extensions to the content.
Arrow 1b	ACMS to ACMS test content portability	Similar to 1a but will include instrument definition. In a multi-provider content environment, one provider may build the items and tests to be delivered by another provider. The content development provider will send items and all test definitions and scoring models to another provider for processing.
Arrow 1c	ACMS to ACMS content bank portability	At the end of a contract, one provider may send an entire item and/or test banks to another provider as part of the contract transition between providers.
Arrow 2a	ARAS to APSMS student authorization (not Phase 1)	A student accesses the APSMS to take a test. The APSMS will access the registration system to verify the student is eligible to test and to determine which form of the assessment to administer. Student profile information is retrieved from the ARAS to ensure proper accessibility content and features are available to the student.
Arrow 2b	ARAS / APSMS testing status (not Phase 1)	Generally the ARAS will monitor the overall administration to provide testing status to administrators. This likely would include number of students tested and how many of those have been scored. In general, the APSMS will likely provide testing status to the ARAS so that administrators can ensure the testing is progressing according to schedule.
Arrow 3	ACMS to APSMS test publish	After the assessment provider has created an assessment form and that form is ready for administration, the provider will request the ACMS to package the assessment form and provide it to the APSMS for delivery to students. The content will be published for specific delivery platforms (paper/online/mobile) The package will contain the content (both default and accessible), the assessment structure, and potentially the scoring information and performance statistics of the items if the form contains an adaptive section. After receiving the content package, the APSMS may store or alter the packaged assessment to optimize it for that delivery platform. NOTE: An alternative is for the APSMS to request the packaged assessment form from the ACMS at student login time.
Arrow 4	ACMS to ARAS form administration controls (not Phase 1)	When a form is ready for administration, the assessment provider may also provide administration controls to determine when and where the form is available (sometimes referred to as a form sampling plan). Specific forms may be available only to certain districts within a state or may only be available during certain administration (Fall, Spring, etc.). The ARAS and/or the APSMS will ensure the form can be administered when and where appropriate.
Arrow 5	APSMS to ASMS scoring a delivered assessment	After an assessment form has been administered to a student, the responses and the adaptive scores from an adaptive test will be sent to the ASMS for evaluation and scoring. All response details, including interactions, tools used, timing, comments, etc. should be passed to the scoring system for evaluation. Any student identifying or demographic information will also be passed with the response and score data.

Identifier	Short Description	Details
Arrow 6a	ACMS / ASMS score definitions	When a form is ready to be scored, the ACMS will pass the scoring model or test map to the ASMS for item scoring as well as total and sub-test scoring. All tables used for derived scoring, such as raw-to-scale, scale-to-performance level, etc. must also be passed to the scoring system. Scoring rubrics may also be included for open ended items.
Arrow 6a	ACMS / ASMS performance statistics (not Phase 1)	During or after scoring is complete for all assessments, resulting scores are typically used for psychometric evaluation. The resulting statistics generated from the analysis are sent back to the ACMS to be stored with the items (in the item bank).
Arrow 7a	ARAS / ASMS student data (not Phase 1)	Generally the APSMS will pass through student identifying information to the scoring process. However, there may be additional student demographic data that is not passed through the delivery system but is required to complete the scoring process. In these cases, the ARAS will provide the required demographic data to the ASMS to complete scoring an assessment.
Arrow 7b	ARAS / ASMS scoring status (not Phase 1)	Generally the ARAS will monitor the overall administration to provide testing status to administrators. This likely would include test and scoring status for number of students tested and how many of those have been scored. In some cases, if the scoring is occurring across multiple solution provider platforms, then the ARAS may orchestrate those interactions. In general, the APSMS will likely provide testing status to the ARAS and the ASMS will likely provide scoring status to the ARAS.
Arrow 8	ASMS / AROR scores to reporting	After each assessment has been scored, the resulting data is generally sent to the AROR for reporting the results to all users. The AROR will typically store the results. This interaction must support the ability to receive preliminary scores as well as final scores after scoring is complete. In other rare cases, an assessment may be rescored based on an issue with the instrument or after a scoring appeal. In these cases an updated score may need to be transmitted and processed by the AROR.
Arrow 9	ACMS / AROR content display on reports	For online reporting capabilities in the AROR, it may be beneficial to show items to the users. For example, if the reports are telling a teacher that a large percentage of his/her student missed question 14 on the test, it would be beneficial to display the item content along with any other diagnostic, standard measure, or feedback information from the item to help inform instruction.
Arrow 10	LOCAL / ARAS registration	It is typical that student registration data is collected from the local districts or regions to support a centralized administration of an assessment. The student registration data is typically collected from an SIS. This data will likely include student identifying data, demographic data, enrollment data, teacher data, and program data. Generally representing a point-in-time for accountability purposes.
Arrow 11a	ARAS / AROR registration data	It is typical that not all registration data is passed through the delivery system so that it can be handed to the scoring/ reporting systems after administration. Therefore, the registration system will likely provide the complete registration detail to the AROR as required.
Arrow 11b	ARAS / AROR record changes / re-roster	There are cases where registration data may be modified post-administration. In these cases, the updated registration data must be shared with the AROR.
Arrow 12	AROR / ADW populate data warehouse	If the ARS provides for a longitudinal or analytical data warehouse solution, the detail results will generally be provided by the AROR to the ADW.

Identifier	Short Description	Details
Arrow 13a	LOCAL / AROR return results	After a test (or an administration is complete), the results will be returned to the local information systems for various uses. For example, test scores may be used to determine course grades, matriculation, or graduation. In addition, the results can be used to personalize instruction for individual students.
Arrow 13b	LOCAL / AROR preliminary results	It may also be necessary to return preliminary results (preliminary pass/fail) to the schools prior to all scoring being complete. This is generally used when matriculation/graduation requirements are affected by test outcomes.
Arrow 14a	STATE / AROR preliminary results approvals	After a test administration is complete, the results will be returned to the state for review and approval.
Arrow 14b	STATE / AROR final data	After final results have been generated, the resulting data is sent to the state for use in their systems.
Arrow 15	LOCAL / ADW return summary data	If the ARS provides a data warehouse, the data warehouse may return various summaries or analyses to the local information systems for tailoring instruction or reviewing program effectiveness.
Arrow 16	STATE / ADW return summary data	If the ARS provides a data warehouse, the data warehouse may return various summaries or analyses to the state information systems for tailoring instruction or reviewing program effectiveness.
Arrow 17	ASMS / ASAS scoring analysis	It is typical for scoring results to be analyzed by psychometricians prior to release scoring results for reporting.
Arrow 18	AROR / AAS reporting analysis	Assessment results will likely undergo various analyses to provide consumable information by the user. For example, analysis may be performed to determine student performance on standards, teacher or program effectiveness, or other research.
Arrow 19	ADW / AAS reporting analysis	Very similar to arrow 18 but may also include longitudinal analysis.

Definitions

Note: The use of LOCAL here could be substituted with REGION for some programs.

Terminology	Definition/Description/Meaning
AIF	Assessment Interoperability Framework. This document defines AIF.
AIFWG	AIF Working Group. A small team that represents SIF, IMS, vendor and user community representatives that develop the AIF.
APIP	APIP is a specific and testable profile of the QTI umbrella assessment specification that is provided for the US K-12 market in support of RTTA. The APIP profile extends the base QTI specification to include all accessibility elements to enable assessments for all students.
Assessment Analysis System (AAS)	A sub-component of the ARS, the AAS is provided to perform various analyses on assessment results for reporting or data warehousing purposes. Analysis may be provided to determine student performance on standards or for teacher or program effectiveness.
Assessment Creation and Management System (ACMS)	The ACMS is responsible for managing all assessment content, metadata, business rules, and tools necessary to create (author) assessment items and forms. In laymen's terms, this type of system may be referred to as an item banking or test banking system. The ACMS must support the content development lifecycle for the assessment program including initial content creation, edits and reviews, and collection and analysis of item performance data (statistics) after an assessment has been administered. The ACMS must be able to define all scoring rules and meta data necessary to score the individual items, sub-tests (or strands), and total test scores and performance data. The ACMS must be able to manage individual content elements (items, art work, reading passages, tools, etc.) as well as test and form structures necessary to build a deliverable assessment. The ACMS must be able to package and provide access to assessment content so that it can be delivered, scored, and reported by other assessment services and external systems.
Assessment Data Warehouse (ADW)	A sub-component of the ARS, the ADW will provide long term data storage and aggregation of data to support business intelligence type of reporting.
Assessment Delivery System (ADS)	The ADS provides all registration, test delivery (or test presentation), session management, and scoring functionality. In short, the ADS is the traditional assessment system. Given package assessments from the ACMS and test taker registration information from an external source (such as an SIS), the ADS is responsible for ensuring that all test takers are registered and accounted for, that proper security credentials are available to the delivery system, that the appropriate tests and forms (including accessibility options) of the tests are available for delivery, that all scoring is complete, and that results are available and ready for use by the reporting systems.
Assessment Platform	When all assessment system components are combined or integrated together, the sum of the components completes the assessment platform.

Terminology	Definition/Description/Meaning
Assessment Presentation & Session Management System (APSMS)	A sub-component of the ADS, the APSMS is responsible for presenting the assessment to the test taker and collecting the response information. Given test taker credentials and profile data from the ARAS and a packaged assessment form from the ACMS, the APSMS must be able to present the assessment and collect response data based on the specific delivery platform or media. A paper-based APSMS is likely much different from on online ADS. The APSMS may also integrate closely with the Assessment Score Management System (ASMS) in the case of an adaptive test where delivery and scoring must occur in real time. The APSMS is responsible for ensuring the assessment content is presented in the correct sequence and that all supplemental content and tools are available to the test taker as appropriate. The APSMS will also ensure that all accessibility options are available. The APSMS may also provide monitoring tools and controls to a teacher or test proctor to oversee and manage the session as well as record information about the session as needed. For example, if a student is caught cheating, then the proctor may be able to intervene and stop the test.
Assessment Registration & Administration System (ARAS)	A sub-component of the ADS, the ARAS will provide methods for loading, identifying and profiling individual students that may be testing. For large scale assessment programs, the ARAS must account for all students including not tested students. The ARAS will implement all rules concerning test administration windows, form assignment, ensuring all students are accounted for and all scoring is complete.
Assessment Reporting System (ARS)	Given completed (delivered and scored) assessment data, the ARS is responsible for formatting and presenting assessment results to all users or other consuming systems. The ARS may also provide longitudinal tracking or aggregation data from the individual student results in the form of a longitudinal assessment data store. The ARS can be used to provide the historical information to a scoring process that will perform the calculations concerning multiple assessment cycles such as growth or progress data. It is expected that the ARS can analyze detailed results data and convert that to meaningful and actionable information for the specific user's needs.
Assessment Results Operational Reporting (AROS)	A sub-component of the ARS, the AROS is intended to provide immediate results to users for a test administration. The AROS is not expected to derive any score data but may use score data to calculate summary information such as class/school averages as an example. The AROS may provide print and online versions of the reports. The AROS will provide data to other consuming systems as needed. This may require some reformatting or filtering of information (such as de-identification for research).
Assessment Score Processing System (ASPS)	The ASPS is responsible for processing all assessment results post-delivery, applying all scoring rules/models, and generating score data.
Assessment Scoring Analytics System (ASAS)	A sub-component of the ASPS, the ASAS is commonly used to analyze the scoring results prior to the data being released to reporting. Specifically for high stakes testing, analysis may be performed to established scale or cut-scores. Additional analysis may be performed to ensure the resulted generated by scoring are as expected (accurate).

Terminology	Definition/Description/Meaning
Assessment Scoring Management System (ASMS)	A sub-component of the ADS, the ASMS is responsible for deriving score, performance and feedback information for individual responses as well as collections of items in strands/sub tests and for the total assessment. The ASMS may use algorithms or lookup tables to derive alternate scores, such as scale-score or normative data. The ASMS may also determine student performance level information or mastery data. The ASMS must be able to score a variety of item types such as multiple-choice, true/false, short response, etc. The ASMS must be able to implement various scoring algorithms and techniques such as matching, lookup, algorithmic (i.e. artificial intelligence), and human or distributed scoring in order to perform all scoring necessary. The ASMS may have to rely on a longitudinal data store (possibly from the ARS) in order to calculate growth scores.
AYP	Annual Yearly Progress
CEDS	Common Education Data Standard
Item Banking	A sub-component of the ACMS, Item Banking provides all the management functions for creating and editing item level content including items and shared content elements (such as reading passages). Item Banking functionality allows for the inclusion of accessibility extensions to all content elements. All item content can be aligned with learning standards. The Item Bank will provide all scoring information for each item including scoring rubrics. The Item Bank will house any item statistics for each item and each item/form/administration used.
Learning Standard	An academic content standard that expresses the content knowledge, skill level, or process knowledge a student should have at a particular grade level. Each state has these standards established and there are other authoritative sources for different standards. These standards are represented by the SIF Learning Standard set of objects.
LMS	Learning Management System. Generally manages the delivery of learning content to students based on lesson plans or teacher directed instruction. An LMS generally can tailor instruction based on learning outcomes, including assessment results.
PARCC	Partnership for the Assessment of Readiness for College and Careers. One of two consortia selected to develop a Core Assessment program (www.parcconline.org)
PNP	Personal Needs and Preferences. An IMS standard for defining student preferences for assessment.
QTI	An IMS specification for defining assessment items (questions) and tests (forms). This is the umbrella specification that APIP profiles for RTTA assessments. In this document we generally refrain from using QTI in favor of using APIP.
SBAC	SMARTER Balanced Assessment Consortium. One of two consortia selected to develop a Core Assessment program (www.k12.wa.us/smarter).
SIF	Schools Interoperability Framework (Association). SIF provides for a data and transport standard for managing student and other educational data.
SIS	Student Information System. Generally manages all student data, enrollment, schedules, health records, contacts, etc. Typically integrated with or includes grade book functionality. Can be locally or centrally hosted. Many industry players in this space.
SLDS	State Longitudinal Data System.
Test Banking	A sub-component of the ACMS, Test Banking provides all the management functions for creating, editing, and publishing forms of tests that can be delivered to students. Test Banking will define all scoring information in order to derive total test and strand/sub-test scores including raw, scale, percents, norms, performance levels, etc. as well as which learning standards scores are being reported against.

Appendix A – Assessment Comparability

There are many types (or administration variables) that can impact comparability. Some of those are discussed below.

1. Assessment form comparability – It is common practice for assessment publishers to build multiple forms of an assessment (each form containing some or all unique items). Generally speaking multiple forms are used to mitigate cheating, item exposure, and longer administration windows as well as a method to inject larger quantities of field test items into the administration process. Assessment form comparability ensures that regardless of which form of an assessment is taken by a student, the results can be compared. Generally this is mitigated through a scaled scoring system.

2. Cross media or device comparability – The comparability between media (paper versus online) or device (laptop versus tablet versus mobile device) that the student is using to take the test. Much research has been conducted concerning paper versus online comparability but less on various electronic devices. Similar to form comparability, differences are generally mitigated through a scaled scoring system.

3. Cross vendor platform comparability – Generally limited to online assessment, this comparability would be between two assessment platforms that have different presentation or navigation styles as well as built in tool differences (ex: different highlighter functionality). No research for this is immediately available.

4. Comparability between states using the same instrument – With RTTA several states will likely be using the same assessment instrument. This practice is common with norm reference tests and formative/benchmark assessments but is less common for standards-based or criterion referenced testing. One example may be the American Diploma Project for Algebra. If all states are not using the same platform, then cross vendor platform comparability may also play a role.

5. Comparability between states using the same instrument with different adaptive algorithms – With RTTA, it may be possible for different states to be using the same assessment content (or item bank) yet utilizing different vendor platforms for adaptive testing. If given a common set of adaptive rules, will each adaptive algorithm generate comparable results? No immediate research is available.

6. Comparability between consortia – With RTTA the two consortium (PARCC and SMARTER Balanced) are developing different assessments (different content banks). It is not expected that the assessments are comparable across consortia. However, some "bridging" may need to be developed for students that move between states and consortium.

7. Longitudinal comparability – Actually a misnomer, this is really referring to the ability to track growth with the assessments from year-to-year. The RTTA assessment designs should address these requirements.