From:	Thompson, Mark P.
To:	RaceToTheTop.Assessment
Cc:	Twing, Jon; Niyogi, Shilpi; Ostler, Wayne; Strain-Seymour, Ellen; Dolan, Bob
Subject:	Call for Public Comment: Framework for Interoperable Assessment Technology Standards
Date:	Monday, November 07, 2011 5:01:02 PM
Attachments:	1107_USED_Document_Feedback_Final.pdf

To: U.S. Dept. of Education
Attn: Susan Weigert, Ph.D
Education Program Specialist

Attached is Pearson's feedback to the Call for Public Comment: Framework for Interoperable Assessment Technology Standards.

Mark Thompson
Vice President
National Services
Assessment & Information
Pearson

D: (952) 681-4550
M: (651) 402-0766
E: mark.thompson@pearson.com

Pearson
Always Learning
Learn more at www.pearsonassessments.com

PEARSON ALWAYS LEARNING

PEARSON
2510 North Dodge Street
Iowa City, IA 52245-9945
T: (319) 339-6407
F: (319) 358-4224

www.pearson.com

Susan Weigert, Ph.D
Education Program Specialist
U.S. Dept. of Education
Office of Special Education Programs
550 12th St. S.W.
Washington, D.C. 20202-2600

Pearson Feedback: Development of Open Technology Standards

7 November 2011

Dear Ms Weigert:

Attached is Pearson's feedback to the recent post by the U.S. Department of Education (ED) concerning the development of open technology standards for managing and delivering student assessments and assessment results. Our feedback includes suggestions and recommendations on key areas to be included in the interoperability technology standards to achieve the desired goals of comprehensive industry standards to support next generation assessments by 2014.

Pearson is in a unique position to provide you with valuable insights regarding this topic for several reasons. First, due to our relatively strong position in the education services arena, we have acquired many disparate technologies and platforms which we had to make interoperable. As such, our expertise solving these internal issues is quite relevant for your mission. Second, our work with the various governing and oversight bodies (e.g., IMS, SIFA, and PESC—and to a lesser extent the efforts of the Council of Chief State School Officers and the Association of Test Publishers) have provided us with an external point of view into the various but often different concerns these groups have regarding interoperability and, as such, have helped us to develop an overarching understanding of strategies not wedded to any one particular point of view. Finally, since we support "cradle to career" learning systems we are delighted to see progress toward a ubiquitous lifecycle interoperability framework that will truly make education assets accessible for all.

If you have any questions or would like to discuss our response further, please contact me at 319-339-6407 or by email at jon.s.twing@pearson.com or my colleague Shilpi Niyogi, Executive Vice President, National Services, at 202-378-2128 or by email at shilpi.niyogi@pearson.com. We look forward to servicing this effort as it evolves in the future.

Sincerely,

Jon S. Twing, PhD
Executive Vice President and Chief Measurement Officer
Assessment and Information Group of Pearson
D: (319) 339-6407
E: jon.s.twing@pearson.com

Pearson Feedback: Development of Open Technology Standards

Introduction

Pearson appreciates the opportunity to provide feedback to the recent post by the U.S. Department of Education (ED) concerning the development of open technology standards for managing and delivering student assessments and assessment results. ED does an excellent job outlining the many attributes required for assessment interoperability standards to function as well as highlighting the fact that multiple standards exist today that address one, many, or all of those attributes. Based on the Request for Information on this topic released last December, ED has posted a framework for standardization at http://www.ed.gov/oii-news/educational-assessment-technology-standards. ED also has a view of the current standards landscape and how that landscape aligns with the assessment interoperability needs. Pearson believes that ED is driving the interoperability agenda with the Race to the Top Assessment (RTTA) grants in a very positive manner.

Pearson is a strong proponent for establishing precisely this level of dialog across the education and standards communities. By marrying the needs for assessment interoperability with the current interoperability standards we can create an overarching vision for assessment interoperability standards that defines the structures, attributes, and relationships for interoperability standards. Pearson does not believe that any single entity should undertake the effort of defining one single interoperability standard for all assessment needs—that would prove to be costly, untimely, and would ultimately fail. Alternatively, we believe that the community should come together in a collaborative way to leverage and enhance the standards that already exist.

In support of this vision, Pearson actively participates in all of the major educational technology standards bodies, including but not limited to IMS Global (IMS), Schools Interoperability Framework Association (SIFA), and the Postsecondary Electronic Standards Council (PESC). We provide support and resources for many of the working groups sponsored by these organizations as well as other organizations active in the educational technology standards community. Pearson representatives also serve on the Board of Directors for IMS, SIFA, and PESC.

Without a common and shared understanding of how interoperability standards should be utilized and how they relate to each other, the overlap and conflicts between them will continue to exist. Each standard will continue to expand to cover more of the interoperability landscape and their user's needs—only increasing the overlap and conflicts. New standards initiatives will be launched in an attempt to make common the

disparity between standards or to fill gaps in coverage, only further complicating the environment. This will result in each assessment consortium, state, vendor, or organization potentially utilizing a single standard or combining standards for their programs into unique implementations. This will drastically reduce the reusability of solution sets across programs. Today, there are many organizations that possess, are actively developing, or are planning to launch initiatives that fall under the banner of “standards.” We know that this web of initiatives (and associated acronyms) is confusing the vendor and user communities and decelerating the speed at which interoperability is achieved.

Conversely, with a common and shared understanding of the standards landscape, we can start to examine the landscape and identify natural interfaces between the standards. By collectively defining how the standards can be connected in seamless implementations, we will greatly increase the likelihood of the assessment solution sets being reusable as well as the ability for many solution providers to understand how they can connect their solutions within this environment. This will allow each of the standards to grow and innovate within their strengths, allow them to leverage the great work that is occurring in other standards, and reduce the motivation to proliferate standards.

While the ED draft document and the feedback provided in this document focuses largely on assessments, we must not lose sight of how assessments and assessment interoperability standards fit into the larger educational ecosystem. Assessments are a key component of the teaching and learning cycle and provide critical data for personalizing instruction for all students regardless of their learning styles or individual abilities. If the thread cannot be pulled through the educational fabric from when the assessment item is developed through how the feedback is delivered to the student (or instructor) to inform instruction, then the picture is not complete.

Specific Feedback—Interoperability and Innovation

The following text appears in the *Introduction* section of the draft document.

> Interoperability will spur innovation in technology in several ways, specifically by:
>
> - Encouraging seamless integration of assessments and other educational data in schools, districts, and States, while protecting students' personally identifiable information;
> - Providing opportunities for organizations to create new and varied approaches to developing and administering educational assessments without sacrificing compatibility with existing systems; and
> - Making assessment items and tasks portable across organizations, platforms, and States.

This discussion recognizes the natural *creative tension* that exists between strict adherence to standards and high levels of innovation in the market. These two ends do not need to be mutually exclusive but they do create tension. Most standards recognize that they cannot represent all possible scenarios within their standards so they generally provide for custom or user-defined extension points. While these extension points allow for the insertion of new ideas, they do limit the levels of interoperability when they are used. Most standards bodies recognize that the use of these extension points should be monitored as they provide a rich source for new requirements to be incorporated into future versions of the standards.

In the RTTT application guidelines, there were statements indicating that any custom or user-defined extensions to standards must be approved. While Pearson understands the drivers for such restrictions, we are concerned that this type of restriction will in fact stifle innovation. Some level of innovation must be allowed to occur partially, or in some more extreme cases, completely outside of the boundaries of the current interoperability standards.

In the cases where new innovations have been implemented outside of the standard's boundaries, the organizations implementing those solutions should provide the educational community:

- Clear articulation of when and where interoperability standards are not applicable so that users understand that they are accepting that risk.
- A plan for how the new innovations can be put back into the standards community or an explanation why the new innovation cannot be represented in existing standards.

In addition, significant new innovations may be closely held by the inventing organizations through IP policies, patents, or licensing restrictions. Organizations may have invested heavily in research and development and want to leverage that investment in their products or capabilities. Any measures to restrict such activity will also stifle innovation.

Pearson suggests that clear statements about the limits of interoperability standards and how innovation will be allowed to outpace the evolution of interoperability standards should be included in any documents that would drive policy or contract requirements.

Specific Feedback—Standards Elements and Priorities

The following text appears in the *Introduction* section of the draft document.

> Based on the RFI responses…interoperability technology standards should include the following priority areas:
>
> - Assessment items;
> - Assessment instruments;
> - Student information and data profile;
> - Student assessment results; and
> - Infrastructure, including data transportation, activation, and security.
>
> Standards for interoperability are important not only within these areas but across them. Standardization in other areas, such as the administration of assessments, scoring, and learning records management, may be beneficial but is not essential at this time.

This discussion provides two things:

1. How to identify the various elements that standards must address
2. A priority for related elements.

Pearson agrees that the high level priority elements identified are correct. We would also like to discuss the nature of the relationships among the priority elements. Understanding the elements and the relationships among them will ensure that we have a seamless flow of content and data throughout the assessment lifecycle as well as extending the data for use by the teaching and learning lifecycle.

The relationships among the elements will identify candidate interface points between standards and other potential elements for standardization to support the seamless flow of content and data. As an example—if we consider that assessment items and instruments represent two elements within the content layer, we must understand how all of the metadata at the content layer is used by the delivery and scoring process to generate results data that can be used by the consuming systems. The content layer will not only provide the item level scoring information but also how items are combined together to generate scoring and performance data by learning standard as part of the instrument's definition. If the content layer and data layer do not share a common representation of learning standards, then we have the potential for errors in translation/transition between standards. Also we must recognize that performance on learning standards may not be the only measure that is generated. Processes, mechanics, or other metrics may also be measured. In addition, the item and instrument scoring processes may also provide feedback (system and human generated) that will need to transition to the data layer. And while Pearson agrees that the scoring process itself is a lower priority standard candidate (as ED has indicated), scoring is a key part of the bridging process between content and data and therefore must be considered.

Pearson has been working closely with the standards organizations to develop structures (visual representations) of the assessment elements, the lifecycle of those elements (Assessment Lifecycle diagram below), and the relationships between categories of systems (see Assessment System Components diagram in later discussion) that we would like to become focal points for considering the standards landscape and the

likely interface points between the existing standards. Some of this material has been used in recent joint meetings between SIFA and IMS Global in Redmond, WA, and Washington D.C.

Pearson believes that our work and the concepts outlined in the ED document align well. However, these are all fairly high level discussions, and it will be critical to take this down to one or two more levels of detail before we can provide consistent and more detailed guidance concerning the application of standards for interoperability to the broader education industry as the RTTA programs are rolled out. Because both consortia have already released RFPs for system architecture and one has been awarded, it is likely that the organizations defining those architectures are wrestling with this topic. To some degree, the industry is behind the eight ball. However, if the users, standards organization, and industry can organize quickly to define the interoperability standards landscape, even at a high level, this would be very useful for those developing architectures and designing solutions to meet the needs of the market.

Specific Feedback—Endorsing Standards

The following text appears in the *Purpose* section of the draft document.

> Please note that ED does not endorse or recommend any specific standard or set of standards at this time. Examples of existing standards listed on the succeeding pages are based upon responses we received to the RFI and ED's knowledge of existing technology standards; they are not meant to be exhaustive.

While Pearson understands that ED cannot endorse specific standards, part of the dialogue that the industry must have is to identify the current standards that support specific elements of assessments, as discussed earlier, and select standards or combinations of standards that best fulfill the mission. We must also identify if those standards need updating or refreshing as well as identify gaps in coverage across standards. Once that has been accomplished, the discussions must shift to how the standards can interface so that seamless solutions can be created.

Specific Feedback—In the Seams

When would educators use assessment interoperability standards?

Priority Standards Area	Description/Use Cases
Assessment Items	Description: An assessment item is the content and material necessary to provide a stimulus that elicits a student response, taking into account student diversity.
	Use Case: An assessment bank or repository moves some or all items to another similar assessment bank or repository. Purpose: To permit the other system to undertake some or all the functions the original system was providing via the use or application of the items
Assessment Instruments	Description: An assessment instrument is a group of items collectively intended to provide information about student knowledge and skills, taking into account student diversity.
	Use Case: Assessment bank or repository moves some or all instruments to another similar assessment bank or repository. Purpose: To permit the other system to undertake some or all of the functions the original system was providing via the use or application of the instruments
Student Profile*	Description: A student profile provides sufficient information about a student to facilitate appropriate assessment administration (such as item, section, or instrument-wide accessibility needs as well as demographic data needed to administer, score, and report on an assessment activity).
	Use Case: An authorized educational system rosters and/or initiates an assessment for a specific student or group of students. Purpose: To permit an authorized system to start an assessment process for students where the assessment activity/process is located on a remote machine
Assessment Results*	Description: Assessment results are the raw scores, scale scores, and information about the relationship between raw and scale scores as well as the link between an individual student's score and the relevant content and achievement standards.
	Use Case: An assessment administration system transfers student work artifacts and/or raw scores to another system. Purpose: To score an assessment administered by one system, where scoring processes are located elsewhere or where scoring involves long-running and/or human intervention
	Use Case: Send scored assessment results to systems of record Purpose: To finalize the administration process by recording the score of a student or group of students with a system of record
	Use Case: Provide a series of reports/data about student assessment results to any systems that are authorized to display them Purpose: To provide flexibility in distributing results of the assessment to various systems, from publicly visible and searchable systems, to student systems that need access to individual student information (PII) consistent with privacy requirements
Infrastructure (Data Transportation, Activation and Security)*	Description: Data activation is the ability to invoke an action to or from one or more systems, and may involve reliable, multi-step actions. Data transportation is the ability to move data from one system to another. Security is the ability to undertake any action privately and with appropriate authentication (identity) and authorization (permission).
	Use Case: Enable independently managed systems to communicate with each other. Purpose: By permitting independence and communication among systems, systems can be designed to replace one another more easily, and building any given system can be simplified by defining a more narrow set of features that must be "exposed" to remote systems. Security can also be enhanced by narrowing the scope of transactions accepted across systems.

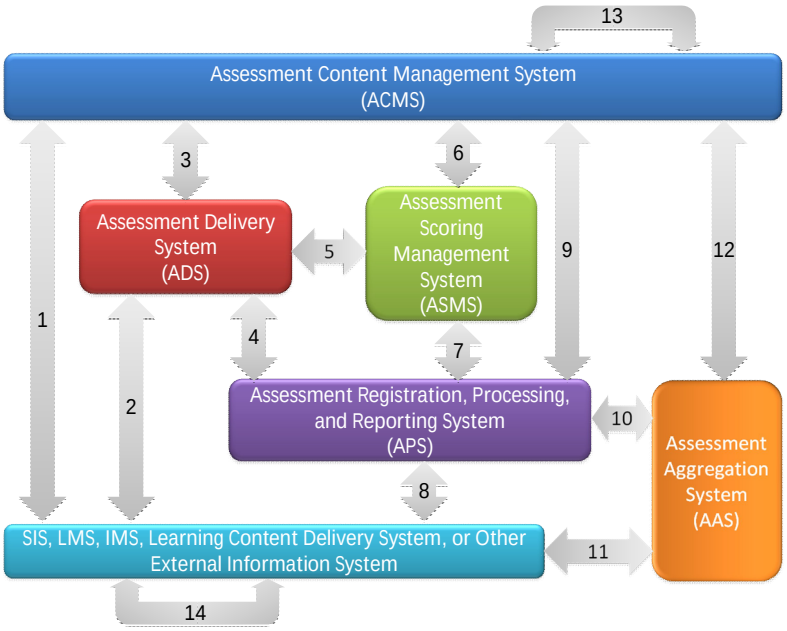
*Consistent with all applicable privacy statute and regulation for the security of personally identifiable information

Source: http://www2.ed.gov/about/offices/list/oii/assessment/technology-standards.pdf

As discussed earlier, we need to explore all of the seams between the rows outlined in this diagram. This will likely result in other priority items for standardization. The diagram below illustrates the typical components of an overall assessment solution. Today, these components are typically bundled together into an assessment platform. Each of the labeled arrows becomes a potential point for interoperability and therefore standards. In the outline above the arrows that are in play in this discussion are:

- Assessment Items and Instruments—Arrow 13
- Assessment Registration—Arrow 8 (upward)
- Assessment Results—Arrow 8 (downward), arrow 14 and possibly arrow 11 if we need to distribute summary results using standards (ex: class, school, district averages).

Assessment System Components



Based on the stated five priority items for interoperability, the following assumptions should be stated:

- Content interoperability (items and instruments) is between content repositories (ACMS to ACMS—Arrow 13), and not between a content repository (ACMS) from one provider directly to a delivery system (ADS) from another provider—Arrow 3. This direct exchange between providers (Arrow 3) is not in scope for phase 1.
- If multiple scoring solutions are provided for one assessment, Arrows 5, 6, and 7 can be implemented using proprietary techniques between scoring platforms. For example: if one vendor is doing all of the multiple choice scoring, another doing the open ended item scoring, and a third deriving the final scale scores and performance levels, those content and data exchanges are not in scope for phase 1.
- The ability to launch an assessment directly from any local platform (such as an LMS or grade book) is not in scope for phase 1—Arrow 2.

The following interoperability points are available with the five priority items to those choosing to implement.

- The ability for a platform to view an assessment item or instrument (Arrow 1) is possible given content interoperability standards.

IMPORTANT NOTE: By "not in scope for phase 1" we are specifically referring to the first wave of interoperability guidance. Defining interoperability will have to be a phased process over several iterations to complete the picture. We consider the priorities identified by the ED draft document as phase 1.

Specific Feedback—Infrastructure

All of the arrows in the assessment components diagram above require content and/or data transport. It is important to recognize that each of these arrows may have differing demands on infrastructure. For example, some interactions support asynchronous processing while others require more direct communications. Interactions that are not carrying personally identifiable information (PII) or secure assessment content may have less security concerns than those that do carry that information. Some interactions would benefit from being highly automated or scheduled while others may be event or "human" activated. Some interactions may require sophisticated sequencing or synchronization of information as part of the interaction, others may not.

As we look at each type of interaction (i.e. each arrow), we will need to identify the needs associated with that interaction and identify the appropriate technologies for implementation. Similar to the data and content standards discussion above, Pearson strongly suggests that each interaction type select the best technology and standard set as the solution that meets those needs and strongly suggest that all implementations adhere to those technologies and standards.

Specific Feedback—Standards Evolution

As with anything, especially technology, things evolve and become outdated over time. When considering today's existing standards, there may be aspects to those standards that need to be refreshed, refactored, or replaced. Several of the key existing standards have been altered and extended many times over the years and may be in need of significant updates or refactoring. Old approaches or technologies may now seem antiquated. All standards work hard to keep future enhancements backwards compatible with prior versions, which complicate the level and speed of change that can be attempted and increases complexity. In software terms, we often refer to this as technology debt, or in more colorful terms, spaghetti code. This phenomenon is virtually impossible to prevent. Organizations may make significant investments to overcome the buildup of technology debt.

As we consider establishing interoperability models and selecting standards, we must evaluate the level of debt we are inheriting and determine, if to be most effective, rework must occur. The groundwork we are laying today should have a shelf life of many years before the next wave of technology debt has been built up. We are not suggesting that we start from scratch, but we learn from the experiences we have today, make strong recommendations on where change needs to occur, and move forward.

Specific Feedback—Assessment Types

The five priority items listed in the draft ED document are largely, but not completely, focused on the large-scale or summative assessment environment. While this is a logical starting point, a potential risk is excluding the needs of formative and classroom use assessment types. With guidance from education industry experts we can make progress on the five priority items without boxing the discussion into corners that don't allow for future expansion into other assessment types.

Assessment items and instruments, as we know them today, may drastically change in the future. Today's interoperable assessment content describes the "inner-workings" of assessment content in order to make it interoperable across content banks and delivery, scoring, and reporting platforms as well as accessible to all students. Will gaming or simulation assessment scenarios make defining the inner-workings of assessment content untenable? As a potential path, will assessment content standards give way to "tool launch" models where the content stays put with its delivery platform—similar to those used in the instructional content space? Obviously this may result in some platform "lock-in" for assessment but it may be necessary.

Will the lines between instruction and assessment, such as gaming scenarios, be blurred to the point that assessment is no longer a separate, independently identifiable set of content or data?

While the outcome of these likely future scenarios remains uncertain, we should do what we can to standardize where we can and either allow for thoughtful extensions to or relief from strict enforcement of interoperability standards.

In Conclusion

The draft document provided by ED is consistent with how we recommend structuring the standards discussion. This draft document is a good starting point from which to elaborate more details about the standards landscape and the various dimensions that must be used to describe and discuss candidate standards. Pearson, is ready, willing, and able to help provide the vision and leadership that would be needed to complete this picture.

Pearson also suggests that other industries have likely gone through the same standards-based transformations. As examples, the banking, health care, and aviation industries have had to consolidate, connect, and prune standards to achieve very high levels of interoperability. While the education industry certainly has some very unique nuances, we should seek to learn from other's experiences.

Pearson also recognizes that we cannot (and should not) do this alone. Nor do we believe that any single organization, left to their own devices, can achieve the level of interaction required to develop a complete vision or model. In order to develop and define an optimal interoperability model that addresses the initial five priority items, and establishes a base from which future requirements can be addressed, it will require a combined force with broad, but targeted representation from the vendor, standards organizations, and user communities.

With each assessment consortium having released RFPs for the development of their architecture, with one being awarded, it is unclear how the service providers developing those architectures could participate in this activity during their contract periods. This would need to be reviewed in the light of each assessment consortium contract.

The timelines by which this strategy needs to form are tight. The window of opportunity that the current RTTA environment is creating may also close if not addressed in a timely manner. Like most standards development activities, Pearson would suggest that a small, very focused team (i.e. working group), develop a robust straw

man model by early 2012 for widespread distribution and comment. This small team should have a few representatives from vendors, standards organizations, not-for-profit organizations, and user communities to start the straw man development. The straw man model would also make recommendations on which standards should be adopted and for which elements of the assessment life cycle or system component interactions. The small team should utilize agile development techniques (common in the software industry) to quickly iterate and generate shareable information for community reaction and feedback. It would be important that the small team be endorsed by a cross section of the communities they represent and that their recommendations would be given strong consideration for adoption.