The McGraw-Hill Companies

Education

Bruce D. Marcus
Executive Vice President
Chief Information Office
Corporate Information Management

1221 Avenue of the Americas
New York, NY 10020-1095
212-512-3416 Phone
212-512-4484 Fax

January 17, 2011

Mr. Steve Midgley
Office of Educational Technology
Attention: Assessment RFI
U.S. Department of Education
400 Maryland Avenue, SW., Room 7E202
Washington, DC 20202-0001

Dear Mr. Midgley,

McGraw-Hill Education is submitting these comments in response to the U.S. Department of Education's December 20, 2010 "Assessment Technology Standards Request for Information ("RFI).

This submission includes our views on the broader development and implementation of objectives and strategies relating to education technology standards and our recommendation for creating and implementing such standards. We also provide direct answers to the questions outlined in the RFI. We thank you for to opportunity to participate and would welcome the ability to continue to participate in the discussion, creation and implementation of such standards.

McGraw-Hill Education is a premier provider of teaching and learning solutions for 21st century post-secondary and education markets worldwide. Through a comprehensive range of traditional and digital education content and tools, McGraw-Hill Education empowers educators and prepares professionals and students of all ages to connect, learn and succeed in the global economy. McGraw-Hill Education, a division of The McGraw-Hill Companies, has offices in 33 countries and produces content and digital tools in more than 65 languages.

For more than a decade the promise of technology is that it can transform education and positively effect students' outcomes. That promise until very recently has gone unfulfilled. Inadequate or non-existing infrastructure and bandwidth have been two key areas that have held back fulfilling that promise. Another inhibitor has been the difficulty in the creation and adoption of open, interoperable, and broadly adopted technology standards.

Recently, the Race to the Top Assessment (RTTTA) program and the Common Core State Standards (CCSS) being adopted by the state consortia have focused on the need for interoperability standards for the U.S. education system especially around assessment items and data on the results of the assessments.

As technology moves forward in education, standards and frameworks will be crucial to enabling innovation if done right and carefully. However, standards also may have the potential to curb innovation, so that the Department of Education, standards bodies, and the assessment consortia should be careful in the scope and specification of any standards. Interoperability standards have been used in many industries and have improved products for both providers and end-users. One relevant example is in higher education, where the IMS Global Learning Consortium provides a

single framework — Learning Tools Interoperability or LTI — for integrating rich learning applications with platforms like learning management systems, portals, or other systems.

A number of companies that serve the higher education market — from software providers to universities and institutions to publishers — have adopted these standards. From a publishers perspective McGraw-Hill is able to offer all course resources in a single, online environment, designed to significantly improve faculty and student workflows. In addition, common interoperability standards make it that much easier for the developers of infrastructure to embrace the next great innovation, thereby improving efficiency and lowering costs overall.

Assessment interoperability standards are necessary for testing components and platforms to work together. But that alone is not sufficient. McGraw-Hill Education wants to underscore and support strongly the Department's efforts through this RFI and as stated in the RFI, to work toward a set of "requirements, [that,] taken as a whole, give States the flexibility to switch from one technology platform to another, allowing multiple providers to compete for States' business and for States to make better decisions about cost and value. Use of technology standards that meet these requirements will help ensure that public investments in assessment instruments and related technology can be used in the education sector as broadly as possible and, at the same time, contribute to a competitive and innovative market place."

Today, there exist multiple standards that are used to varying degrees and for varying purposes in the K-12 assessment market. Each one has a specific focus but there is considerable overlap between some of the standards. Experience shows us that platforms and standards need to remain dynamic so that they can remain relevant to those the market they serve.

McGraw-Hill Education recommends that any standard entity be composed of and work with the K-12 assessment stakeholders be engaged to develop and maintain a common agreement that would be adopted widely and will ensure interoperability, at a minimum, of assessment items and data. The creation and adoption of such assessment interoperability standards will minimize the need for customized extensions, foster innovation, reduce costs, and ensure that States, schools and students have access to the latest technology improvements with the most beneficial features and benefits to best facilitate learning.

For additional information, please contact me, Bruce Marcus, at 212.512.3416 or bruce_marcus@mcgraw-hill.com or Bala Balachander, CTB/McGraw-Hill, 630.789.4168 or Bala_Balachander@ctb.com.

Sincerely,

Bruce Marcus

Attachment: Assessment RFI Response

Context for Responses

The primary intent of this RFI is to explore existing, in-process, or planned open technology standards, specification, and technology products that support the management, delivery, and exchange or assessment content and the capture and exchange of assessment results. While the focus of this RFI is assessment technology standards, the Department recognizes that assessment generally occurs within the context of broader learning activities (whether online or offline) and, therefore, does not wish to restrict the range of responses to assessment-0only approaches. The Department, therefore, also welcomes responses that address broader technology standards of approaches that are relevant to the handling of assessment management, delivery or reporting. As mentioned earlier, the Department has required RTTA grantees to adopt a technical standard (or standards) that permit interoperability o the assessments and technology developed by that program. To help focus our consideration of the comments provided in the response to the RFI, we have developed several questions regarding the development of assessment technology standard(s) and their application to the RTTA program. Because these questions are only a guide to help us better understand the issues related to the development of interoperable technology standards for assessments, respondents do not have to respond to any specific question. Commenters responding to the RFI may provide comments in a format that is convenient for them.

3.2 Questions about Assessment Technology Standards

General and Market Questions:

3.2.1 Current Landscape

What are the dominant or significant assessment technology standards and platforms (including technologies and approaches for assessment management, delivery, reporting, or other assessment interoperability capabilities)? What is the approximate market penetration of the major, widely adopted solutions? To what degree is there significant regional, educational sub-sector, or international diversity or commonality regarding the adoption of various technology standards and capabilities, if any?

Multiple standards dominate the assessment market. Each one has specific focus but there is considerable overlap between some of the standards. The principal standards are: *QTI, SCORM, AICC, APIP, SIF, and PESC.*

QTI

The Question and Test Interoperability (QTI) specification was developed by the IMS Global Learning Consortium. The specification is in its second version. The specification describes a data model which represents a question or item, a test, and the corresponding results data associated with the test and items. The specification enables the exchange of this information between authoring, assessment delivery systems, learning management systems, etc. The intent of the standard is to be applicable across a number of domains – including licensure and certification, K-12, and Higher Education. The specification does have some limitations – there is no set way to describe standards and alignment to the standards (an important requirement for K-12). The specification also does not contain information on the "look and feel" of the assessment item – an important requirement if the intent is to preserve the look and feel of an item and test across multiple delivery platforms.

SCORM

The Shareable Content Object Reference Model (SCORM) was developed by the IMS Global Learning Consortium. The intent was to promote the development of reusable learning or content objects that can be run in any platform / Learning Management System that supports the SCORM run time model. The specification enabled the decoupling of instructional content development with run time considerations. The SCORM standard does not address assessment items directly.

AICC

The Aviation Industry CBT (computer-based training) committee (AICC) developed a set of guidelines to define rules and interfaces that allow computer-based training content from a number of different sources to work with computer managed instruction systems. The intent of AICC is similar to that of SCORM. Neither standard has gained substantial traction in K-12. The AICC standard body collaborates with IMS standard body.

APIP

The Accessible Portable Item Profile (APIP) is a relatively new standards specification. The primary focus of the standard is on accessibility and accommodations – specifically for accessibility information to be built into items. The standard integrates two standards – QTI and Accessibility for All standards and extends the accessibility tags to include different types of accessibility needs. The specification was developed by stakeholders from a set of states (about 10) and two vendors well versed in accessibility. The standard is endorsed by IMS Global Learning Consortium.

SIF

The Schools Interoperability Framework (SIF) specification was created by the SIF Association, a consortium of vendors, school districts, departments of education and other entities active in pre-K–12 markets. The framework specification is a set of rules and definitions that enable software vendors to share information and data. The specification promotes interoperability across disparate systems. The initial focus was on student information, staff, academic records etc. but the specification has been extended to include assessment information. The focus is on the results of the assessment and not on the assessment items and test specification.

PESC

The Postsecondary Electronic Standards Council (PESC) predominantly focuses on standards in Higher Education but have links to K-12. The standards body focuses on interoperability and exchange of data and has specification on admissions, test score reporting, transcripts, etc.

The above standards focus on assessment information even if peripherally. There are other standards – specifically on data and reporting. Most recently, a new consortium, Common Education Data Standards (CEDS) was formed to formalize and organize the approach to state based longitudinal common data standards. The consortium consists of Council of Chief State School Officers (CCSSO), the State Higher Education Executive Officers (SHEEO), the SIF Association, the PESC, the National Center for Education Statistics (NCES) and the Data Quality Campaign. These partners will collaborate to advocate, communicate and promote common data standards and identify and develop standards as needed.

3.2.2 Timelines

Approximately how long would it take for technology standards setting and adoption processes to obtain a technology standard that meets many or all of the features or requirements described in this RFI? What are the significant factors that would affect the length of that timeline, and how can the impact of those factors be mitigated? More specifically, would the acquisition of existing intellectual property (IP), reduction or simplification of specific requirements, or other strategies reduce the time required to develop these technology standards and processes?

A number of factors influence the timelines associated with setting a standard. The number of participants, the diversity (vendors, school districts, standards bodies, etc) of the participants, the processes that are followed to solicit input and to reconcile multiple perspectives, the frequency of meetings, etc. each would have an impact on the timeline. The scope of the specifications is also a significant factor.

The timeline can definitely be shortened if the starting point for the standard specification is an existing set of standards that have been adopted by the industry. A well defined and narrow scope will also help the timeline.

The first version of a standard can take between 12 to 24 months. The following are illustrations of the typical timeline associated with setting a standard:

The first version of SCORM took about 12 months and APIP took about 20 months. These timelines are for the development of the standard and do not take into account the time needed to solicit comments and revise the specifications based on the comments. In the case of SCORM, two minor versions were released about 18 months after the first version was released. In the case of APIP, the specifications were built on top of existing standards.

Implementing a set of standards typically takes longer than setting the standard. The vendors participating in the standards setting tend to incorporate some of the aspects in the roadmap of their products and platforms but it takes a few years before a standard is completely adopted within a market place. For example, the Department of Defense was tasked with setting the specifications of SCORM in 1999 and it was in 2006 that DoD required the use of SCORM. In K-12, there is very little adoption of SCORM. In K-12, the SIF association developed a set of specification a few years ago but we haven't seen wide spread adoption of SIF in school districts. A number of factors influence the timeline associated with adopting a set of standards. The simplicity of the standards, the practical benefits of adopting the standards and the level of compliance of the standards in the vendor products all play a critical role in the implementation and adoption of standards.

3.2.3 Process

What process or processes are appropriate for the adoption, modification, or design of the most effective technology standard in a manner that would answer many or all of the questions in this RFI? We are interested in learning the extent to which the uses of one or another process would affect the timeline required to develop the technology standards.

Adoption, modification, or design of effective technology standards requires active involvement, input and support from key standards boards, industry organizations and stakeholders. Collaborative efforts amongst these participants are required to drive end-to-end implementation, from initial design to cross platform adoption with well defined timelines for phase-in and product support.

To date, many of the successes due to technology standards have been achieved through proprietary and standalone efforts. Processes that foster cohesive collaboration, centralized agreement, continued outreach to partners / stakeholders for input, and ongoing cooperation to evolve and adopt common standards are needed. In addition, these processes must discourage proprietary stakeholder interests while enabling competitive cooperation to exist.

The following actions / processes should be started (concurrently where possible) to facilitate development of widely accepted technology standards and ensure shorter industry adoption:

Establish joint governing body represented by stakeholders from existing standards bodies, leading industry organizations and solution partners

- Baseline existing standards set
- Develop vision and scope of desired technology standards
- Develop milestones and timelines for modification and adoption
- Pilot implementations with solution partners and stakeholders
- Beta implementations to help evolve standards
- Call for open source contributions to standards

3.2.4 Intellectual Property

What are the potential benefits and costs to the Federal Government, States, and other end-users of different IP restrictions or permissions that could be applied to technology standards and specifications? Which types of licensed or open IP (e.g., all rights reserved, MIT Open License, or Gnu Public License) should be considered as a government technology standard? How should openness relating to the IP of technology standards be defined and categorized (e.g., Open Source Initiative-compatible license, free to use but not modify, noncommercial use only, or proprietary)

The subject of intellectual property associated with standards is a complicated one and is one that merits further discussion. There are tradeoffs associated with control and openness in creating and setting standards (Simcoe, 2005). This trade-off becomes more nebulous as there are different interpretations of the term "openness". Openness can be associated with the standard setting process – i.e. anyone can participate in creating / developing the standard. It could mean that the implementation process is open i.e. anyone can implement a set of standards. It could also mean that a standard is open if one can freely adopt, implement and extend the standard (Simcoe, 2005).

Open standards doesn't necessarily mean that the implementation of these standards in products are open source. In fact, there have been numerous cases of commercial products that are built on open standards. For example, all network manufactures support Ethernet – a networking standard. Open standards can co-exist with commercial implementation and we believe that in order to promote innovation, especially in the next generation of assessments, it is imperative to have open standards but allow for innovation of commercial products, for which intellectual property rights are retained, that are based on the open standards.

Our recommendation is to make the standards implementation open – i.e. anyone can implement the standard but the standard is created by a consortium that is represented by state, district, vendors and the Department of Education and other standards bodies. As mentioned above, we recommend that the emphasis be on open standards and not on open source software. Assessment and education, in general, can greatly benefit from innovation by different vendors and organizations and it would not bode well to stifle innovation by mandating an open source approach to implementation.

3.2.4.1 Existing Intellectual Property

What are the IP licenses and policies of existing assessment technology standards, specifications, and development and maintenance policies? Are the documents, processes, and procedures related to these IP licenses and policies publicly available, and how could the Department obtain them?

Most of the standards appear to follow "open implementation" – i.e. they allow anyone to implement the standard. These standards are not "open creation' i.e. the creation of the standard is not open to everyone though they do solicit input from the public. It doesn't appear as though the intellectual property information is publicly available. The standards bodies can provide additional information in this area.

3.2.5 Customizing

Can assessment tools developed under existing technology standards be customized, adapted, or enhanced for the use of specific communities of learning without conflicting with the technology standard under which a particular assessment tool was developed? Which technology standards provide the greatest flexibility in permitting adaption or other enhancement to meet the needs of different educational communities? What specific provisions in existing technology standards would tend to limit flexibility to adapt or enhance assessment tools? How easy would it be to amend existing technology standards to offer more flexibility to adapt and enhance assessment tools to meet the needs of various communities? Do final technology standards publications include flexible IP rights that enable and permit such customizations?

What are the risks and the benefits of permitting such customization within technology standards? When would it make sense to prevent or to enable customization?

When addressing customization in assessment tools (i.e. any technology platform used to author items, delivery assessments, score assessment data, or report / analyze assessment results), it is important to consider the manner in which assessments and/or assessment content enters the tool. There are many ways to represent an item. At one end of the interoperability spectrum, there are items that exist in hardcopy formats only. From an assessment standpoint, these assessments require some amount of (automated or manual) reverse-engineering to provide the assessments in a standardized format. At the opposite end of the interoperability spectrum are assessments that exist in an industry recognized standard format. Examples of such formats are XML based representation of assessments like QTI and SCORM. Customization of assessment tools may occur to consume / expose assessments in a standard format without sacrificing the functionality of the assessment tool itself.

In order for assessment systems to support an interoperable assessment, assessment vendors need to recognize a standard representation of assessments and meta-data associated with assessments. Currently, there are some XML (mentioned above) based standards that encapsulate the ability to represent an assessment. While these standards do a good job of representing what an assessment is, they do not indicate how to display assessments and assessment items to individuals. In most cases, rendition of the assessment is left up to the assessment vendor. For low-stakes situations, rendition is not as critical, but in high-stakes situations, differences in the rendition of an assessment may throw off the psychometric models used to construct the assessment. Another drawback to current technology standards is the lack of academic standards alignment available within the standard. Most standards support meta-data around the assessment, but there is no specific method available to align assessment content to to content or performance standards.. Typically, to do this, the assessment / content vendors leverage flex fields in the standard for this purpose, but the practice is not standardized.

In order to create a truly interoperable standard, the standard must define the structure of the assessment and items, rendition of the items, and contain a structured way to represent meta-data related to the items and assessments. Current standards have their strengths and weaknesses. For instance, QTI does a great job of describing the structure of the assessment and assessment content, but does not contain specific information as to how to render the assessment content on or offline. Another alternative would be HTML, but that only focuses on the rendition of content, it does nothing to describe content structure and meta-data. A 'perfect' standard would be one that contains all three concepts for an assessment.

Most technology standards do not account for IP considerations relating to content. While IP issues must not impede interoperability, mechanisms must exist to protect the author's copyrights and IP. Most vendors leave the IP issue up to the person using the toolset. In most situations, the user bears the burden to license the content and the tools do not have enforcement mechanisms built in. One exception is the IMS Common Cartridge specification. It does contain specification for a web service that may be exposed by a content provider to validate a proper license exists.

As mentioned above, customization comes with some risks. Any customization that alters proprietary content or changes the psychometric basis for the assessment must be prevented.

3.2.6 Conformance and Testing

Do existing technology standards or technologies include specifications or testing procedures that can be used to verify that a new product, such as an assessment tool, meets the technology standards under which it was developed? What specifications or testing procedures exist for this purpose, e.g., software testing suites, detailed specification descriptions, or other verification methods? Are these verification procedures included in the costs of the technology standards, or provided on a free or fee-basis, or provided on some combination of bases?

Conformance and testing are an integral step in rolling out a successful technology standard. Most standards include a certification and testing process that allows a vendor to publish compliance to a standard. Typically, the conformance is managed by a governing body that is part of or a delegate of the standards setting body.

Most standards bodies include a test suite or testing service that a vendor must pass before being certified. In some cases, the certification may contain levels of compliance. The process typically follows the following pattern:

- Vendor implements the technology standard
- Vendor executes a set of test cases provided by the standards body as part of the QA cycle
- Vendor applies for certification or compliance by notifying standards body and providing needed documentation
- Vendor executes certification test scenarios and provides results to the standards body
- Standards body reviews results and documents and approves/denies certification
- If approved, standards body provides vendor an image / logo that the vendor can display in their product indicating they have been certified.

Costs for certification are the responsibility of the applicant. In most cases, standards bodies are non-profit / independent organizations that are funded by charging membership fees or by a government entity. In some cases, these bodies will not grant certification to non-member organizations. Typically, the certification is open to all vendors (members or non-members).

3.2.7 Best Practices

What are best practices related to the design and use of assessment interoperability technology standards? Where have these best practices been adopted, and what are the general lessons learned from those adoptions? How might such best practices be effectively used in the future?

Standards (technical or non-technical) must be very well documented and easy for a Subject Matter Expert to understand. Additionally, technology standards walk a very fine line between being flexible and easy to implement. If a standard is too flexible, vendors struggle to implement it in a consistent manner which leads to partial implementations or vendors loosely adhering to the standard. In cases where a standard appears too flexible, interoperability suffers because vendors do not implement it consistently. Standards that are to concrete or inflexible do not translate well for a wide audience. Therefore, inflexible standards are not widely adopted because the standard does not meet the requirements of a majority of the vendors.

Another key best practice is to ensure standards are independent of the underlying technology stack. When standards dictate technology must be used to implement the standard, they tend to alienate a large portion of the vendor community. Additionally, the standards described typically require some specialized knowledge to implement which increases the implementation costs for vendors. Finally, standards that dictate technology

do not adapt well to technology changes. If a standard defines the technology to be used to implement, it will die out when the technology it depends on becomes obsolete.

Finally, standards must be governed by an independent, vendor-neutral, non-government organization. Standards governed by a particular vendor tend to serve the best interests of that vendor, while standards governed by a membership driven body tend to be more flexible and transportable because of the different inputs into the standard. The governing body must have a strong governance structure in place to prevent one or two major contributors pushing agendas that are not in the best interests of the community. Some good examples of bodies that manage standards are IEEE and ISO.

Technological Questions Regarding Assessment Technology Standards:

3.2.8 Interoperable Assessment Instruments

What techniques, such as educational markup or assessment markup languages (see also http://en.wikipedia.org/wiki/Markup_language), exist to describe, package, exchange, and deliver interoperable assessments? How do technology standards include assessments in packaged or structured formats? How can technology standards enable interoperable use with resources for learning content? How can technology standards permit assessment instruments and items to be exchanged between and used by different assessment technology systems?

There are many different markup languages used in to describe, deliver, package, and exchange assessments. Typically, the markup languages fall into one of two broad categories.

- Presentation – Markup tags describe how to render text or graphics to an end-user
- Descriptive – Markup tags contain meta-data about the contents encapsulated by the tags

Presentation markup languages promote interoperability by describing how to present an item / assessment in a consistent manner regardless of the technology used to present it. Typically, these markup languages are limited to a specific technology domain. Examples of presentation markup languages are HTML, XHTML, Postscript, and TeX. Presentation markup contains no descriptive information about the content. It only tells a consumer how to present the content. Therefore, the content consumer receives no information describing how to use the content or the purpose behind the content.

Descriptive markup languages promote interoperability by giving a content publisher the ability to describe how to use the content. The most widely used descriptive markup language is XML. There are numerous standards based on XML. In education, some examples of XML based standards are QTI, MathML, SCORM, SIF, and APIP. Descriptive markup only describes content. The rendition of the content is left up to the content consumer.

Currently, technology standards for assessments provide the basics need to describe the items in the assessment and sequence of items on the assessments. Additionally, the standards contain a lot of flexibility around assessment meta-data. The meta-data may include instructions, psychometric information about items, assessment type, publisher information, etc. Currently assessment data is not standardized and each standard represents data in a different manner. For a technology standard to support a high degree of interoperability, the format and contents of the assessment / item meta-data need to be standardized.

In addition to data standardization, the information about the rendition of items on the assessment must be a part of the standard. Currently, there are ways to embed presentation information within a descriptive XML tag. One easy example is placing XHTML content with XML tags. This provides both descriptive information about the assessment / item and rendition information for the consuming system. By embedding presentation

information in the tags, the publisher can be assured that the assessment items will be rendered to the user in a consistent manner across multiple assessment platforms.

3.2.9 Assessment Protection

For this RFI, ''Assessment Protection'' means keeping assessment instruments and items sufficiently controlled to ensure that their application yields valid results. (See also paragraph below, ''Results Validity.'') When assessment instruments or content are re-used or shared across organizations or publicly, are there capabilities or strategies in the technology standards to assist in item or instrument protection? What mechanisms or processes exist to ensure that assessment results are accurate and free from tampering? Do examples exist of public or semi-public assessment repositories that can provide valid tests or assessments while still sharing assessment items broadly?

"Validity is the ongoing trust in the accuracy of the test, the administration, and interpretations and use of results. According to Messick (1995), "validity is not a property of the test . . .as such, but rather of the meaning of the test scores . . . (that) are a function not only of the items or stimulus conditions, but also of the persons responding . . . (p. 741). ..Validation must encompass the full testing environment: (content) standards, test constructs, items, persons, characteristics, administration, and interactions of each" (Barton, 2007, p. 82).

Technology standards can help assure interoperability across technologies for:

- the item authoring environment (i.e. capturing standard types of information across environments during item creation that lend to cross-reviews of alignment to construct and standards, tagging data, scoring philosophies, accommodation influence),
- how and which data are stored (item data, test form data, examinee response data, etc.)
- administration or delivery modes and accessibility features (online, on paper, with or without technology-enabled accommodations, on portable devices, etc.)
- reporting interfaces for and/or access to assessment results

What technology standards should not address are the purposes and valid interpretations of the full assessment; that is how the items collectively represent an assessment form for a given purpose. However, if the given purpose of an assessment is standard across technologies (for example, a single assessment delivered in multiple modalities), the assessment community should provide standardization documentation across technologies involved in the assessment development, delivery, scoring, and/or reporting to attend to confounding variables that may potentially influence the valid interpretations of the assessment across technologies. For example, if a single assessment is to be delivered across devices, then the assessment entity should provide strict standards for maximizing comparability of results across those devices so that validity of results remain comparable.

Technological attention coupled with adherence to the constructs intended to be measured should drive the assessment entity's decisions for acceptable devices or platforms and relevant standards, since the technology should be agnostic to constructs measured. <u>Technology can facility increases in validity by opening the options for assessing a construct, but it is the utilization of the technology in light of the construct that affects the validity. It is not the technology alone that influences validity.</u>

It is akin to addressing accommodations made available when certain constructs, such as calculators or "read alouds," and evaluating whether or not the accommodation is in conflict with and invalidates the measures intended. Technology standards can influence how well an accommodation is provided, but not whether or not it should be provided. The assessment entity makes the decisions about constructs measured and threats to validity that accommodations may pose through accommodation guidelines (see for example CTB's

Guidelines for Inclusive Test Administration (www.ctb.com/media/articles/pdfs/general/guidelines_inclusive.pdf) .

3.2.10 Security and Access

In what ways do technology standards provide for core security issues, such as access logging, encryption, access levels, and inter-system single-sign-on capabilities (i.e., one login for systems managed by different organizations)?

All software-based systems can be vulnerable to security attacks and a number of security protocols have been developed as safeguards. Security standards are critical for systems that require data exchange and interoperability. Secure software must be dependable, contain little vulnerability, and also be resilient. For interoperability and maintenance, the following standards (at a minimum) must be addressed for all software systems/applications:

- User access control standard
- User naming standard
- User password management standard
- Privileged user account management standard
- Service account standard
- Information classification standard/guideline
- Network security standard
- Remote access security standard
- Patch management standard
- Data security and encryption standard

When utilizing the technology standards listed above, systems defining access levels, for instance, will require less effort to implement. With standard security protocol, systems will also provide better interoperability, delivery or exchange of content and assessment results.

One example of a standard that facilitates access across systems is Single Sign-On (SSO). SSO uses centralized authentication servers for controlling access to multiple, related, but independent software systems. A user logs in once and gains access to all systems without being prompted to log in again. Single sign-off is the reverse property whereby a single action of signing out terminates access to multiple software systems.

As different applications and resources support different authentication mechanisms, single sign-on has to internally translate and store different credentials compared to what is used for initial authentication. There are many commercial and open source based SSO systems available.

Benefits of SSO include:

- Reduces phishing success, because users are not trained to enter password everywhere without thinking.
- Reduces password fatigue from different user name and password combinations
- Reduces time spent re-entering passwords for the same identity
- Supports conventional authentication such as Windows credentials (i.e., username/password)
- Reduces IT costs due to lower number of IT help desk calls about passwords
- Provides security on all levels of entry/exit/access to systems without the inconvenience of re-prompting users

- Centralizes reporting for compliance adherence.

Another recommended security standard is encryption. All secure, confidential, personally identifiable information that must travel through the Internet should be encrypted at 128-bit or higher for data protection.

Batch file transfer – use SFTP so that data is encrypted during the transport or encrypt data manually and FTP if SFTP is not viable.

Online data/system access – use HTTPS instead of HTTP for secure transaction. HTTPS will encrypt from the source and decrypt at the target. HTTPS creates a secure channel over an insecure network. This ensures reasonable protection from eavesdroppers and man-in-the-middle attacks, provided that adequate cipher suites are used and that the server certificate is verified and trusted.

3.2.11 Results Validity

For this RFI, ''Results Validity'' means protecting the statistical validity and reliability of assessment instruments and items. How can interoperable instruments be managed to ensure they are administered in a way that ensures valid results? Are solutions regarding assurance or management of validity appropriate for inclusion in technology standards, or should they be addressed by the communities that would use the technology standards to develop specific assessments?

Standards around the statistical validity and statistical reliability of the assessment instruments and items should be managed by the psychometric and assessment communities using the technology standards. The communities should adhere to best practices to assure validity throughout the assessment system, from item and form development, throughout administrations across administration modalities, and to results. However, technology-related standards should be considered in order to reduce measurement error, such as due to variability in examinee experience, and to increase the likelihood for greater validity and reliability. Note that there are a multitude of factors that can influence both validity and reliability that are not fully covered in technology related discussions.

A key component to validity that should be well attended to in technology-related standards is accessibility and consistency in administration. For example, examinees should have consistency in item presentation format within an administration, and equitable access (including assistive and personalized technology) to the administration environment of the assessment, content within the environment, and navigability through the environment. One mechanism for assuring maximum and equitable access across examinees, would be for technology standards, as recommended in Section 3.2.8, to provide consistency in the markup language around renditions to maximize interoperability and standardization, thereby minimizing systematic error (differences in examinee experience) that diminish reliability and thus validity of results.

Even with greater access, some examinees will still need accommodations provided. There should also be standards around minimum availability of accommodations within the assessment environment, so that across environments, examinees have equitable access to the tools or accommodations need. Again, this provides equity in the assessment and does not introduce systematic error in the measurement of the constructs being assessed, resulting in increases to validity.

Validity does not stop after administration. Score interpretation and use is part of validity. To interpret scores in light of accommodations provided CTB has developed standards around accommodations in testing, *Guidelines for Inclusive Test Administration (www.ctb.com/media/ articles/pdfs/general/guidelines_inclusive.pdf)* that provide a framework categorizing accommodations for examinees with disabilities and who are English Language Learners. The standards and categories for the framework are based upon extensive independent expert reviews and external research. Stakeholders in assessment

results, such as states and districts, utilize the categories to qualify the validity of score interpretations based upon the accommodations provided. The framework can readily adapt to include additional accommodations and/or be reflective of system-specific standards.

In addition, technology standards should emphasize pilot testing around examinee-technology interactivity in various scenarios to assure access is maximized and equitable, leading to increased reliability and validity, as recommended in Section 3.2.6 Conformance and Testing. The interactivity and possible threats to examinee access and navigation should be incorporated into the assessment (item, form, administration, scoring) development process.

3.2.12 Results Capture

How can technology standards accurately link individual learners, their assessment results, the systems where they take their assessments, and the systems where they view their results? How do technology standards accurately make these linkages when assessments, content, and other data reside across numerous, distinct learning and curriculum management systems, sometimes maintained by different organizations?

In order to accurately track results and accurately link information back to students across a heterogeneous environment, two things must exist:

1. Results data structured in a manner that can be exported / imported by different systems
2. Identifiers for key data elements must be globally unique and understood by all systems

There are many technology standards that describe how to send student response data between systems. Typically, the response data is limited to textual elements. Responses that require graphics (i.e. when a student draws a graph for a response) are not widely supported, if at all.

Currently, technology standards do not mandate or promote a globally unique ID set. Standards only provide data structures to store identifiers. The origin of the identifier is not specified. Therefore, it is up to the systems exchanging data to agree upon the ID set used when exchanging data. Several methods exist to solve the identifier problem. One way involves a central authority that manages the identifier for certain data elements. Another method, the most common way of addressing the identification issue, involves designating a 'system of record' that 'owns' specific data elements. In the 'system of record' case, the system of record generates the IDs and maintains changes to the data elements associated to that ID.

3.2.13 Results Privacy

How do technology standards enable assessment results for individual learners to be kept private, especially as assessments results are transferred across numerous, distinct learning systems? How can such results best be shared securely over a distributed set of systems managed by independent organizations that are authorized to receive the data, while still maintaining privacy from unauthorized access?

Privacy contains several dimensions. One dimension of privacy is anonymity of students. Another dimension is data protection of sensitive data. Finally, user authorization to view data must be enforced to ensure that proper privacy exists. In the end, the goal of privacy is to ensure that data is protected and not available unless proper authorization exists.

Many technology standards exist to encrypt data. In situations where data is being transferred over the Internet, SSL, using X.509 certificates, is an industry adopted standard. The degree of encryption varies depending on the certificate used. Additionally, many standards exist to encrypt data in data stores that contain sensitive information.

Encrypting may be one-way or two-way (using public / private key encryption methods). There are many standards used in the technology industry to support data encryption. Currently, no educational specific technology standards specify encryption requirements.

While data may be encrypted in the data store, it is important to consider who has access to view the data. Therefore, a rigorous provisioning framework must exist in applications that allow user access to results data. These applications must ensure a user only sees sensitive data they are permitted to view. Capabilities that expose data not privy to a user or hide data from a user are not considered secure. It is difficult to for a technology standard to mandate user data access. Typically, the provisioning is left up to the application exposing the data. If multiple systems access data for a master repository, a Single-Sign On or Trusted authentication mechanism can help maintain data privacy across and 'eco-system' of applications.

In cases where results data is exposed publicly, student identity must be preserved via anonymity. Identifying information pertaining to individuals must be protected and not exposed publically. Anonymity goes beyond simply hiding information that may be used to identify an individual. The population of individuals where data is being made publically available also must be considered. If there is a very small population of individuals, a person familiar with the situation may be able to 'back into' individual identities based on some analysis of the data. Therefore, anonymity must not only consider hiding identifying information, it must also ensure the population of individuals is diverse and large so that identities are protected because of the volume of data available.

3.2.14 Anonymization

Do technology standards or technologies permit or enable anonymization of assessment results for research or data exchange and reporting? How do various technology standards accomplish these tasks? For example, where a number of students take a test, can their answers be anonymized (through aggregation or other techniques) and shared with researchers to examine factors related to the assessment (e.g., instructional inputs, curriculum, materials, validity of the instrument itself) without revealing the identity of the learners? Is this an area where technology standards can help?

There are a few options when considering standards to anonymize student identity. It is a common practice to assign a unique identifier to each student and to maintain identity related information in a separate table from other types of metadata in database. Using a unique key or system-generated globally unique identifier, identity of learners can be maintained in a separate information store. Including a GUID instead of student identity in the data to be shared is an efficient and effective way of anonymizing PII (Personally Identifiable Information).

Barcode

The most widely used identification system for electronic tagging is the barcode. The recommended standard is Code 128 because it provides excellent density for all-numeric data and good density for alphanumeric data. It is often selected over Code 39 in new applications because of its density and because it offers a much larger selection of characters. The Code 128 standard is maintained by AIM (Automatic Identification Manufacturers). The Code 128 character set includes the digits 0-9, the letters A-Z (upper and lower case), and all standard ASCII symbols and control codes. The codes are divided into three subsets A, B, and C. There are three separate start codes to indicate which subset will be used; in addition, each subset includes control characters to switch to another subset in the middle of a barcode. Subset A includes the standard ASCII symbols, digits, upper case letters, and control codes. Subset B includes standard ASCII symbols, digits, upper and lower case letters. Subset C compresses two numeric digits into each character, providing excellent density.

RFID

Another emerging standard is related to RFID tagging. RFID tags are intelligent bar codes that can talk to a networked system. The advantage of RFID over barcodes is that RFID readers are much faster and do not require a direct line of sight to active or passive RFID. Barcode readers require a direct line of sight to the printed barcode. This feature; however, makes RFID processing less secure since any reader within range (generally up to 300 feet) can potential decode an RFID tag.

Standards are critical for RFID applications, particularly for interoperable systems with data exchange requirements. A great deal of work has been going on over the past decade to develop standards for different RFID frequencies and applications.

There are existing and proposed RFID standards that deal with the air interface protocol (the way tags and readers communicate), data content (the way data is organized or formatted), conformance (ways to test that products meet the standard), and applications (how standards are used on shipping labels, for example). The International Organization for Standardization (ISO) has created standards for the air interface protocol (11785), conformance to readers (18047) and for testing the performance of RFID tags and readers (18046). But RFID standards are still emerging and a number of security issues must be resolved before RFID can be considered a viable standard for secure assessment data.

Encryption

If PII must be part of data exchange or information to deliver, encrypting each identity field can effectively block such information. In online systems, it is a very common practice to store passwords or other highly sensitive information in encrypted format.

3.2.15 Scoring and Analysis of Results

How can technology standards be used for the scoring, capture, recording, analysis or evaluation of assessment results?

Hierarchy Standard

Establishing proper hierarchy is vital for data capture, aggregation and reporting. Data built around the standardized hierarchy is easily portable amongst the parties involved and essential to aggregate data.

Barcode Standard

In paper & pencil tests, using a barcode that identifies a student is highly desirable to improve accuracy, speed of scanning and data capture. Documents with standardized barcode can be scanned in multiple systems and organizations. Barcode standards are described above.

Data Retention

For security, cost and efficiency in data processing, it is desirable to set specific retention period for each data type early on.

3.2.15.1 Results Aggregation and Reporting

How can technology standards enable assessment results to be aggregated into statistical or other groupings? How can technology standards provide capabilities for results (aggregated or raw) to be reported across multiple technology systems? For example, if a learner takes an assessment in one system, but the results are to be displayed in another, how do technology standards address transferring results across those systems? How do technology standards address aggregation of results for a number of learners who are assessed in one system and whose results are displayed in yet another technology system? Can anonymization controls be included with aggregation and reporting solutions to ensure individual data privacy and protection (see also 3.2.14 above).

ETL tools

There are many commercial (Informatica, BusinessObject, Oracle, SAS, etc) and open source based (Pentaho, Talend and others) that facilitate data exchange amongst heterogeneous systems and organizations in secure and systematic way.

Data Mart and Warehousing

Centralized, data mart and warehouse can provide data to many heterogeneous systems easily along with ETL tools as well as flexibility in data aggregation.

Web Service (SOAP/REST)

Web service can be used in providing necessary data to other parties for aggregation or reporting purpose. Student BIO information can be easily anonymized by providing student-identifier or GUID instead of BIO (see 3.2.10)

3.2.16 Sequencing

How do technology standards enable assessment items stored within an assessment instrument to be sequenced for appropriate administration, when the assessment consists of more than a single linear sequence of items? For example, how do technology standards address computer-adaptive assessments? How are the logic rules that define such sequencing embedded within a technology standard?

Technology standards related to item selection algorithms enable computer-adaptive testing to be offered. An iterative algorithm that repeatedly searches for and presents the optimal item from an available item pool based on updated estimates of the examinee's ability from prior responses forms the basis for basic computer adaptive testing.

Item adaptive technology specifications typically describe paths through a collection of items within a computer adaptive assessment. The specification could define the relative order in which questions are to be presented to a student and the conditions under which an item is selected, delivered or skipped. Such specifications and standards may rely on the concept of an activity tree, comprised of learning / assessment activities that are associated with other activities in a hierarchy.

Sequencing rules and limit conditions can be associated with these activity hierarchies and clusters. The sequencing rules influence the order of items presented while limit conditions (attempts, durations), are used by the sequencing rules to further define when an activity (item) is sequenced next to a user.

3.2.17 Computer-Driven Scoring

How do technology standards permit, enable, or limit the ability to integrate computer-driven scoring systems, in particular those using ''artificial intelligence,'' Bayesian analysis, or other techniques beyond traditional bubblefill scoring?

Technology standards can enable interoperability across scoring methodologies by addressing the meta-data available for each item. Standards should encourage competitiveness in the actual algorithms to support computer-drive scoring systems.

Assessment entities concerned with reliable scoring across systems should be responsible for assuring the meta-data provided is accurate and consistent across scoring systems. For example, the technology standards should provide the framework for the data attached to each item, such as the tagging of examinee responses. The assessment entity should be responsible for developing the scoring rubric or standards for the high-level constraints around more intelligent scoring logic. For example, the assessment entity might say for assessments in Grade X, only items from grades X-1 and X+1 should be made available to the computer adaptive algorithm. How the algorithm utilizes the response data and adapts accordingly should be open to a competitive algorithm development environment.

3.2.18 Formative, Interim and Summative

What technology and technology standards exist that support formative, interim, and summative assessments? What technology standards support nontraditional assessment methods, such as evidence, competency, and observation based models?

The rate of technology change always out-paces the adoption process for interoperable standards. Therefore, standards must evolve to incorporate new advances in technology. As the standard evolves, though, there is always a period of time where the latest innovations are not supported by the technology standard. In essence, the technology standard is in a constant 'catch-up' mode. Practically, the 'lag' is a good thing because it gives the industry time to vet innovation before it becomes standardized. If technology standards get updated, too quickly, they run the risk of standardizing an unproven technology. Once an innovation is proven, though, it needs to be incorporated in the standard as quickly as possible. Therefore, technology standards must be easily adaptable to support technology innovation.

3.2.19 Learning and Training

What applications or technology standards exist that can apply assessment results to support learning and training? Are there technology standards or applications that support more than one of the following: Early learning, elementary/secondary education, postsecondary education, job training, corporate training, and military training?

Much effort has been put into the technical re-use of electronically-based teaching materials and in particular creating or re-using *Learning Objects*. These are self contained units that are properly tagged with keywords, or other metadata, and often stored in an XML file format. Creating a course requires putting together a sequence of learning objects. There are proprietary and open, non-commercial and commercial, peer-reviewed repositories of learning objects such as the Merlot repository.

A common standard format for e-learning content is SCORM (discussed above) whilst other specifications allow for the transporting of "learning objects" (Schools Framework) or categorizing metadata (LOM).

In the U.S. pre-K-12 space there are a host of content standards that are critical as well. The NCES data standards are a prime example. Each state's content standards and achievement benchmarks are critical metadata for linking e-learning objects in that space.

Federation of American Scientists: FAS Learning Science and Technologies Roadmaps Overview

Table 1: A Model for Integrating Creation and Delivery Tools

Tasks	Milestones			Measures	Needed Expertise
	3-years	5-years	10-years		
Development model	Development of high level workflow model for instructional design Initial mapping to existing and emerging infrastructures	Development of tool implementation guidance and enablers	Definition of an open "plug and play" environment for content development and deployment	Tools designed and built in parallel and in tiers based on common structures and formats	Instructional Design Cognitive Science Computer and Network Science and Engineering

Enabling formats and standards	Determine key formats and standards for tool development Demonstrate standards-based tool development	Design and build middle to upper tier tools based on emerging instructional interchange formats	Define large scale eLearning delivery "platform" in terms of specifications and practices	Broad and parallel development of standards-based tools	Instructional Design Cognitive Science Engineering

Table 2: Research Challenges in Integration Tools

Research Priorities	R&D Outcomes
Course Building Tools	▪ Extensible model for how tools and services might interconnect and self-discover. ▪ Enabling Formats and Standards.
Shareable Content Objects to Simplify Use	▪ Content creation tools designed for instructional designers that hide technical implementation. ▪ Tools that seamlessly integrate varied content types for non-technical authors. ▪ Seamless search and access to digital assets. ▪ Tools that support merging content formats including: static, interactive, stream-based, and active; and examine the authoring, integration and deployment issues. ▪ Integration tools for combining disparate media types.
Tools and Services to Assist in Application of Metadata	▪ Implementation guidelines for developers in different domains. ▪ Tools to map semi-automatically across domains and determine impact on content developers. ▪ Tools to automate the application of metadata to all levels of content, perhaps through intelligent analysis by agents. ▪ Methods to connect current and emerging intelligent search and retrieval services that use learning metadata with increasingly complex services and information.
Tools for Collaborative Building and Maintenance of Learning Environments	▪ Documented requirements of tools that support various pedagogical and theoretical approaches; tool examples that support the models. ▪ Rules-based sequencing approaches capabilities for non-technical designers. ▪ Strategies for creating "mini context" templates for reusable compound learning objects that can support many different communities of practice (e.g., Higher Ed, Training, Performance Support, etc.). ▪ Search strategies to enable "real time" assembly of content based on learner profiles, mastery, subject, etc.

3.2.20 Repositories

What technology standards-based assessment instruments, questions, or item banks (or repositories and learning management systems) are used to manage and deliver assessments?

Technology standards for assessment content management and item banks must supported an integrated, repository-based system for item authoring, metadata and test delivery. The standards will ensure one-time entry of the assessment framework structure specifications (content standards, and content and process strands), test items, rubrics, passages, and associated stimuli, and tagging, which can be tracked, versioned, searched, and accessed by authorized users. The system must adhere to security standards as outlined in Section 3.2.10.

Establishing standards that provide interoperability across all assessment products and markets might be ambitious; however, throughout the lifecycle of a test item, standard data fields and formats can be defined to maintain both a complete item history and all item attributes. QTI, SCORM, APIP all address item authoring standards and have been described

above. Conformance to any standards requires specifications including data fields, formats, and tagging.

A web-based, template-based approach to item authoring will ensure ease of use and monitoring of the creation, editing, review, and approval of items during the development cycle.

Item templates for a specified assessment product would be an extension of the technology standards and would be pre-populated with content standard, strand or sub-skill information. The templates would also contain additional specifications to support the writing of high quality items. Following is an example of a standard template used for item authoring. This is not the complete standard but serves as an example of standard data fields that should be set across all assessment components.

- Multiple Choice (MC) or Selected Response (SR) Item
- Full text description of learning standard
- Reporting category
- Thinking skill
- Grade
- Item Specification
- Unique item identifier
- Version
- Manipulative (Calculator, formula, ruler)
- Art required
- Answer-choice specifications
- Samples items (PDF)
- Word count
- Use restrictions
- Approval

The authoring templates should utilize standard editing tools to assist the writing process: bold, underline, cut and paste, undo, and redo, as well as mathematical symbols including graphs, equations, and inequalities. The items must have a preview option available in a standard display format such as HTML.

3.2.21 Content Lifecycle

How can technology standards be employed to support an assessment content lifecycle (creation, storage, edit, deletion, versioning, etc.)?

Technology standards associated with the content repository must include a workflow framework such as Documentum to manage the authoring and test assembly process. Quality gates, review and approval steps must be part of the workflow standard. To ensure that cognitive demands and readability were appropriate, standards such as graded word lists, such as *EDL Core Vocabularies* and *The Living Word*. Our goal is to be sure that:

- Items include only information necessary for assessing the skill or knowledge.
- Syntax is grade-level-appropriate
- Items that contain detailed directions or large amounts of text are divided into steps, sections, or bulleted lists to help students understand the task
- Key words or phrases in the items are presented in a consistent style to make the task clear for the student
- Items address significant content-specific knowledge or skills

3.2.22 Interfaces and Services

What interoperability specifications for application program interfaces (APIs) or Web services interfaces to assessment management, delivery and tracking systems have been developed? How are they organized? What are the best practices related to their design and usage? How broadly have they been adopted, and what are the lessons learned from those who have designed or implemented them?

Right now, most technology specifications focused on the education space only define the format of data, they do not mandate the transport mechanism used to move data between systems. Typically, the transport mechanism is negotiated by the systems exchanging the information. Transport may be file-based (e-mail, manual data load, FTP, SFTP, etc) or transactional (web services). One underlying assumption is data must be exchanged over the Internet via a common Internet transport protocol.

SIF does specify some transport mechanisms. SIF includes a SOAP based web services interface as part of the specification. Additionally, SIF defines the concept of a Zone and Zone Integration Server (ZIS) that marshals communication between a student information publisher and a discrete set of subscribers. While there a lot of SIF compliant / compatible systems in existence, the adoption of the ZIS has not gained a lot of momentum.

3.2.23 Internal Transparency and Ease of Use

Are there technology standards and communication protocol implementations that are ''human readable?'' What are the benefits and risks of ''human readable'' technology standards? Some technology standards are not comprehensible without tools to unpack, decode, or otherwise interpret the implementation data resulting from use of the technology standard. Other technology standards, such as HTML, RTF and XML, are largely readable by a reasonably sophisticated technical user. RESTful-designed Web services are often specifically intended to be readable by, and even intuitive to, such users as well. We ask commenters to consider the extent to which various technology standards possess native ''human readability'' and comprehensibility.

The phrase "Human readable" may need further clarification. As it is described in the RFI, most "human readable" standards are typically readable only by a sophisticated technical user. If this is the definition of "human readable", then yes – there are benefits to standards that are human readable. However, there is a tradeoff between human readability and performance. In ensuring interoperable systems, one of the practical considerations is the performance of the systems – in terms of exchanging data or performing transactions that are distributed across systems. A human readable format tends to be verbose and may affect performance. Technical standards focused on interoperability should balance the performance needs with the human readable needs. In cases where standards are not human readable, it is fairly trivial to convert the machine friendly format to a human friendly format.

3.2.24 Discovery and Search

How is the discovery of items or instruments (or other elements) handled within a technology standard or technology? For example, are there search APIs that are provided to permit a search? How are metadata exposed for discovery by search engines or others?

Discovery and search typically have two components – either one or more registry that different vendors and assessment providers would use to list themselves and a mechanism to register and search different assessment entities. Typically the standard body provides the registry or influences the entity providing the registry. The registry provider, in

conjunction with the standard body and members of standard bodies would develop the attributes needed for registration and search.

The search for items and instruments can get complicated – typically the discovery search is done on pre-created indices on a corpus and the technology used to build the index and the frequency with which the indices are built would affect the search results. If a full text search is not used, then the metadata specification becomes critical and the discovery and search will have to be aligned to the metadata specifications.

3.2.25 Metadata

What kinds of metadata about assessments (i.e., information describing assessments) are permitted to be stored within technology standards or technologies? How do technology standards accommodate structured data (such as new State curriculum standards) that were not anticipated when the technology standard was designed? How are metadata describing unstructured (such as free-text input) and semistructured data incorporated within assessment technology standards?

Technology standards focused on assessments support meta-data needed to render assessments and the items within assessments. The meta-data for items tends to be more robust than meta-data around assessments. When looking at standards there are some gaps that are common across the board. Current standards do not support or have very limited support for the following items:

- Standards framework alignments – Some standards support flexible meta-data fields that may be used for standards alignments, but support is not standardized.
- Computer Adaptive Testing – Some standards have limited support for CATs, but most of today's standards only support sequential assessments. Some limited support exists to define a decision map for the next item to render, but the definition must be based on a decision tree. No support for algorithmic CATs where the application must send item meta-data to an algorithm for the next item to render.
- Advanced Scoring support – Most standards support mechanisms to machine score an item and attach a raw score based on a student's response. Standards do not support structures or interfaces to support advanced or algorithmic scoring techniques like scaled scores, Item Response Theory, Norms, etc.

3.2.26 Recommendation, Rating and Review

Do technology standards or technologies permit rating, review, or recommendations to be incorporated within an item, instrument, or other element? If so, in what ways? How are conflicting ratings handled? Do technology standards or technologies permit ''reviews of reviews'' (e.g., ''thumbs up/down'' or ''Rate this review 1–5'')? Is the rating or review system centralized, or are multiple analyses of the rating data permitted by distributed participants?

3.2.27 Content and Media Diversity

What types of diverse content types and forms of assessment content exist that extend beyond traditional paper-based assessments translated to an electronic delivery medium? We are interested in learning more about electronic delivery and interaction media, such as performance-based assessments, games, virtual worlds, mobile devices, and simulations.

When looking at the landscape of assessments beyond traditional paper based assessment, there is a lot of diversity and potential. As technology continues to improve, technology

standards will lag behind modern technical capabilities. Technology enabled assessments have evolved rapidly over the years. Early on, technology enabled assessments simply took paper-based assessments and put the paper items online. Now, technology allows educators to evaluate students in many new / unique ways. Additionally technology allows more frequent assessment deliveries.

Assessment delivery capabilities have advanced over the recent years. Online technologies allow for a richer online experience for students. Several large technologies offer platforms that support a rich media interaction. Adobe's Flash and Microsoft's Silverlight are examples of these platforms. Unfortunately, these platforms are controlled by a single entity and are not supported on all technology platforms. A more recent innovation has been HTML 5. HTML 5, while it shows a lot of promise for a standardized rich Internet platform, is still relatively new and support by different Internet browsers is not standardized currently. Regardless of platform, current technology supports items with streaming media, links to other content, manipulatives, drag/drop, highlighting, graphing/drawing, read out loud, and many other features.

Content and media diversity also needs to consider assessment results capture. With the growth in tablets / pads, touch-based interfaces, clickers, other hardware technologies, the support for more innovative items is increasing. Currently, technology will support capturing student responses as text, a voice recording, a graphic/drawing, or allow students to manipulate objects to generate a response.

With new technology, there will always be some limitations. An important issue with newer content and media is the capability of technology standards to keep pace. Most standards support traditional item types and assessments, but standards support for more innovative item types and/or adaptive assessments is non-existent or minimal. Another issue is consistent platform support. New content and media depends on a specific technology platform to serve the content / media. Typically, support for the platform requires hardware or middleware vendor support. Therefore, platforms only run on certain permutations of hardware / middleware. Also, the platform may behave differently from implementation to implementation. Finally, network infrastructure must be considered. Newer innovative items generally use more bandwidth than traditional text-based items. Therefore, a fragile or low-bandwidth network may degrade the student's experience and jeopardize the validity of the assessment.

3.2.28 Accessibility

How do technology standards ensure that the platforms are accessible to all persons with disabilities? How can technology standards ensure the availability of accommodations based on the individual needs of persons with disabilities? What factors are important to consider so that accessibility capabilities can be included within an interoperable technology standard, both for end-users, as well as operators, teachers, and other administrators? How are issues related to Universal Design for Learning (UDL) relevant to standards for accessible use? How can technology standards provide for, improve, or enhance Section 504 and 508 of the Rehabilitation Act compliance for assessment technology?

Assessments should be designed to provide maximum access for students. Universal design principles guide us in designing and delivering products and services that are usable by people with the widest possible range of functional capabilities—products and services that are directly usable (that do not require assistive technologies), as well as products and services that require assistive technologies.

The Standards for Educational and Psychological Testing state that, "all examinees (must) be given a comparable opportunity to demonstrate their standing on the construct(s) the test is intended to measure. Just treatment also includes such factors as appropriate testing conditions and equal opportunity to become familiar with the test format, practice materials,

and so forth. Fairness also requires that all examinees be afforded appropriate testing conditions" (AERA, APA, NCME, 1999, p. 74). However, fairness is not the only justification for universal design. Universal design becomes even more critical in the face of federal laws requiring the inclusion of all students in large-scale assessment. Such laws include NCLB, IDEA 97, ADA, and the *Civil Rights Act* of 1965.

The use of universal design strives to provide:

- Equitable use,
- Flexibility in use,
- Simple and intuitive use,
- Perceptible information,
- Tolerance for error,
- Low physical effort, and
- Size and space for approach and use.

Addressing the full range of differently-abled students requires attention to how the needs of target populations may affect test conceptualization, test construction, test tryout procedures, item analysis procedures, and test revision.

Accessibility in the context of this RFI centers on persons with disabilities. In other contexts, accessibility may include the ability to provide assessments in multiple display formats (online, print, tablets, phones, etc). For a technology standard to gain traction, it must consider both of these factors in order to be adopted. With regards to accessibility, standards must consider the fact that an item may need to be presented in different ways based on the accommodation and/or display format chosen. For instance, a non-disabled person may read the item on a screen in a particular order, but when a visually disabled person is read the item, the order in which the item is read may differ from the visual representation of the item.

Within the current educational technology landscape, there are a few standards that address accessibility. One example is Universal Design (UD) which specifies a process to create assessment content that remains consistent across various delivery platforms/modalities. While UD focuses on assessment content development, the ideas and principals of UD must be considered when determining how to address accessibility with standards.

APIP, while not yet a standard, defines a how to represent accessibility information within a QTI assessment item. The APIP definition does have some limitations, though. APIP does not stand alone, it is part of QTI. Also, APIP, as currently defined, requires that all items specify accessibility information for Braille and Audio representations of the item. If an item does not have either of these representations, it will not have a valid APIP representation. For wide-spread adoption, standards should specify how to represent items in different modalities to facilitate accessibility. Standards should not enforce which modalities are required in order to create an accessible item. Definition of required accessibility formats should be defined as part of the item specification not a technology standard.

3.2.29 English Learners

How do technology standards ensure that assessment platforms support the assessment, reporting of results, and other capabilities related to the assessment of English learners?

As with differently-abled examinees, English learners may need specialized accommodations and access tools. The availability of accommodations should be equally available across systems and provided in a standardized manner. English learners should be included in any usability or pilot testing. Translations of test items can be highly varied, even within

assessments. Therefore, it will be highly difficult for technology standards to address the content of the translations. However, other kinds of accommodations related to language, such as the availability of English glossaries that do not impede upon the construct being measured, might be required to be available. In other words, the technology standards should focus on assuring equitable offerings or accommodations and access features across assessment systems and platforms, leaving the content of the accommodations to the assessment community.

Questions about process and IP for technology standards development include:

3.2.30 Transparency

How do the organizations that develop assessment technology standards approach development and maintenance activities? Is it common for such work to be performed in an unrestricted or open public forum? Are there examples of organizations conducting technology standards development through private (e.g., membership-driven) activities? Are the final work products produced through standards-development activities made publicly available in a timely manner? If not, when or for how long is it necessary to keep these products private? What circumstances require, justify, or benefit from protecting trade secrets or intellectual property?

Transparency depends on the model adopted by the standards body. As discussed earlier in the Intellectual property section, it depends on the "Openness" model adopted. Typically, in the education space, the model that has been successfully followed is the membership model. Standard bodies, districts and schools, departments of education and vendors are members of a body that develops the specifications. The draft of the specifications, once completed, are typically posted in a public forum to solicit feedback from the general public. The discussions during the specification development processes tend to be private though some of the standards bodies publish the minutes of the meetings. In general, intellectual properties or trade secrets are not discussed in these meetings.

3.2.31 Participation

Does the development of assessment technology standards depend on membership fees from individuals and organizations who wish to contribute to development and maintenance activities? Are there requirements for ''balance'' within membership across different constituencies? What are the cost and structure of such memberships? Are there viable alternative methods for generating revenue necessary to conduct the work? What are the most realistic and useful ways to generate participation, fund work, and ensure public access to a technology standards setting process?

A number of these sections, transparency, participation, etc, depend on the model adopted by the standards body. A number of the standards bodies discussed earlier, SCORM, SIF, amongst others, have a membership fee and the members collaborate to develop the standards. Typically the standards bodies strive to have a balanced view and have the right mix of representation from the different stakeholder groups – schools / districts, standards body members, vendors and departments of education (state and federal) representatives. The cost structure can be a tiered cost structure with different levels of sponsorships. Leveraging existing standards bodies, as opposed to creating one, may be the most efficient, realistic and useful way to fund, develop and maintain standards activities.

3.2.32 Availability

What are the costs associated with final publication of technology standards, and with all supporting materials for those standards, and can these assessment products be made available at nominal or no cost to users? Do technology standards require restrictions for use or application, including limitations on derivation, resale, or other restrictions? Is it appropriate to obtain patent, copyright, or trademark protections

for assessment technology standards? Are the publications for technology standards and materials provided in a machine-readable, well-defined form? Are there restrictions or limitations on any future application of the publications and materials after initial release? Are developer-assistance materials (e.g., Document Type Definitions, test harnesses, code libraries, reference implementations) also made available free under an open license? In what circumstances should technology standards-setting organizations retain rights or control, or impose restrictions on the use of publications, derivations, and resale or developer-assistance technologies, as opposed to open-licensing everything? When should materials be made freely available (that is, at no cost to the consumer) while still retaining most or all copyright license rights?

3.2.33 Derivation

For technology standards, do copyright licenses for publications and all supporting materials and software licenses for software artifacts permit the unrestricted creation and dissemination of derivative works (a.k.a. ''open licensed'')? Do such open licenses contain restrictions that require publication and dissemination of such works in a manner consistent with the openness criteria described by, for example, a GNU Public License (a.k.a. ''viral licensed'') or an MIT Public License (a.k.a. ''academic licensed'')? Are there policies or license restrictions on derivative works intended to prevent re-packaging, re-sale, or modifications without re-publication for assessment technology standards?

3.2.34 Licensing Descriptions

(for materials contained within the standard, not for the standard's licensing itself). How do technology standards address licensing terms for assessment resources described within the technology standard? Are there successful technology standards or approaches for describing a wide variety of license types, including traditional per-use licensing, Web-fulfillment, free (but licensed), open (but licensed, including commercial or non-commercial use permitted), and public domain status. Are there other resource licensing issues that should be addressed within a technology standard as a best practice?