**APIP Project Response**

**Assessment Technology Standards Request for Information**

January 16, 2011

**Michael Russell**
Vice President of Innovation
Nimble Innovation Lab, Measured Progress
Associate Professor
Lynch School of Education, Boston College

**Dirk Mattson**
Director, Research and Assessment
Minnesota Department of Education

## Interoperability

Within an educational system, there are many types of data and resources that benefit from interoperability. As a few examples, student records (e.g., name, date of birth, address, grades, test scores, disability categorizations, etc.), curricular content (e.g., reading materials, assignments, learning activities, etc.), assessment content (e.g., items and test forms), and school demographic data (e.g., number of staff within departments, teaching experience, qualifications, etc.) are all sources of data and resources that are associated with a given educational system (e.g., school, district, state). When data and resources are to be shared across systems or data is to be aggregated across systems, storing that data or resources in an interoperable manner increases efficiency and accuracy of such transfer or aggregation.

While the term "interoperability" is used in a general sense to capture the desire to store and transfer data and resources in an efficient and accurate manner, interoperability standards developed must be specific to the type of data or resource one is focused on and the uses to which that data or resource will be put. As an example, the structure used to tag and store student records is different than the structure used to tag and store assessment content. While interoperability standards are necessary for both student records and assessment content, the standard employed for student records is distinct and independent of a standard designed for assessment content.

Although a different set of standards is required for each type of data or resources, it is important that touch-points are established when one type of data or resources may interact with another. As an example, while student records and assessment content are very different types of information, some elements of a student record may interact with some elements of assessment content. Consider a student record that indicates a child has been identified with a vision need (e.g., the student is blind). Also consider a test item that has been tagged with instructions on how to describe a graphic to a non-visual student. While these are two distinct pieces of information, a test delivery engine may

create an interaction between these two pieces of information to customize the presentation of a test item for students with vision related needs. In such a case, the information stored in the assessment content standard must be able to be used to interact with information stored in the student record. A touch point, then, exists between the way in which a visual need is stored by the student record standard and how information that prescribes how information in an assessment item is presented to a student with vision needs.

In short, it is important that discussions about interoperability are specific to the type of data or resources that are to be shared across systems. When standards for specific types of data or resources are developed, it is also important to consider how that type of data or resources may be used to interact with other types of data and resourcess, and to therefore develop touch-points between standards. Given this, questions such as "Why are the Accessible Portable Item Profile (APIP) standards needed when SIF already exists" should be replaced by questions such as "What touch points are needed to allow the APIP standard, which is designed for item content, and SIF, which is designed for school information, to interact in desirable ways?"

The following sections focus specifically on recent efforts to develop an interoperability standard for accessible test item content and for student access profiles. By way of background, an overview of Accessible Test Design is provided to help better understand why it is necessary to adopt an item information standard that intertwines item content and accessibility information, and why an item content standard must be developed in conjunction with an access needs profile standard. This response then ends by identifying next steps that are needed as an item content standard, such as APIP, is implemented and supported going forward.

**The Accessible Portable Item Profile (APIP) Standard**

The development of digital test content coupled with computer-based test delivery provides an important opportunity to improve the accessibility of test items. By applying principles of Accessible Test Design, next generation assessment systems will provide more valid inferences about student learning based on test scores for all students. Rather than developing assessment content for the general population of students and then making post hoc changes to that content in order to accommodate the needs of sub-groups of students, Accessible Test Design provides a framework for making careful decisions about the methods used to tailor test administration to maximize the measurement of targeted constructs for each student. In turn, the Accessible Portable Item Profile (APIP) Standards provide a tool for implementing Accessible Test Design. The APIP Standards empower next generation assessments to solve three challenges. First, APIP provides a structure for specifying and storing the access needs of each student. Second, APIP provides a structure for augmenting item content with a variety of supplemental and alternate accessibility information designed to ensure that a test item functions properly for students with a variety of access needs. Third, APIP provides specifications for developing test delivery systems that can use a student access profile to tailor the provision of access tools (e.g., magnification, color contrast, masking, etc.) and the presentation of supplemental accessibility information (e.g., audio, Braille, tactile, or signed versions of item content). Collectively, the tools provided by APIP enable next

generation assessments to capitalize on the flexibility of digital technologies to maximize test validity for all students.

For the past forty years, the educational testing community has struggled to accurately and validly measure the achievement of students in the margins, including English Language Learners and students with disabilities and special needs. It is only during the past decade, with the adoption of the No Child Left Behind Act, that students in the margins have been fully included in large-scale educational testing programs. Nonetheless, the equity and quality with which test accommodations are provided, the expense associated with the provision of test accommodation, and the effects that test accommodations have on the validity of test-based inferences continue to raise concerns.

The adoption of computer-based testing holds promise for addressing some of these concerns about test accommodations. In fact, the flexibility with which the delivery of digital content can be tailored for each individual student provides a unique and powerful opportunity to increase test validity for all students. This opportunity begins by addressing the challenge of designing tests that work for all students.

**Understanding how a test is designed to function:**

A test is designed to measure a specific construct or set of constructs, often referred to as an intended construct. To provide a measure of the intended construct, a test is composed of a set of items and/or tasks. Each item or task is carefully constructed to create a measurement experience that requires the student to apply the intended construct. Each measurement experience involves a carefully crafted three-step process. During the first step, the examinee is presented with information (or content) that is designed to establish a problem or question that stimulates the construct of interest. During the second step, the student is provided an opportunity to interact with content contained in the item or task as he applies the intended construct. Since the application of the construct of interest cannot be directly observed, the third step requires a student to produce a response that represents the product or outcome of the application of the intended construct. It is through this three-step process that a test item or task attempts to access the intended construct as it currently operates within the student. Through multiple observations of the student's application of the intended construct provided by multiple items or tasks, a test allows the outcome of each observation to be aggregated to produce a test score that is used to make an inference about the extent to which the intended construct operates within the student.

When creating a test item or task, item writers generally design items that function optimally for the general population of students, and in the process make several assumptions about students that will interact with each item. Among these assumptions are that students do not have any challenges with fine or gross motor skills, do not have any visual or auditory needs, are able to read near or above grade level, are proficient in English, have adequate executive functioning skills, are able to produce responses using either a pencil, keyboard, or mouse, etc. For many students, however, the knowledge and skills that item writers typically assume test takers possess present barriers to accurately and reliably accessing and measuring the intended construct. For these students, each phase of an item's functioning may present potential access barriers. As an example, for

students who are English language learners, have visual impairments, have difficulty decoding text, read below grade level, or are unfamiliar with words or phrases that appear within an item (e.g., names of people or objects), the text-based narrative format of an item stem may not stimulate the intended construct as designed. Similarly, some items and tasks that present multiple pieces of information, contain multiple steps, or mix the format in which information is presented (e.g., narrative text, tables, graphs, graphics, video, etc.) may be difficult to interact with for students who experience challenges with executive functioning, task prioritization, or information processing. Finally, items that require either written responses or the use of a mouse to select an option or manipulate digital objects may yield inaccurate responses for students with physical disabilities or limited fine/gross motor skills. Across these three phases of an item's functioning, inaccurate stimulation of the intended construct, difficulty interacting with item content, and/or difficulty producing responses that accurately reflect the application of the construct produce challenges to an item's ability to measure the intended construct. Digital test content and delivery provides an important opportunity to overcome these challenges. This opportunity results from two features of digital content and delivery, namely the ability to embed additional accessibility information into digital content files as an item is developed and the ability for a digital delivery system to selectively present sub-sets of that information to individual users based on their specific need.

**Default and Supplemental Item Content**

As Mislevy and his colleagues explain, several different representational forms can be used to present item content to a student.[i] To enable a student to recognize and process content, the form used to present that content may need to be tailored based on the student's representational form need. As an example, content presented in print-based form will not adequately stimulate the intended construct for a student who is blind. However, when that same content is presented in Braille, the content is able to access the intended construct within a student who is a Braille reader. Similarly, for a student who is deaf or hard-of-hearing, content presented in audio form may not adequately stimulate the intended construct. However, when presented in a signed form, the content is able to stimulate the intended construct within a student who communicates in sign.

Related to the notion of alternate representations are distinctions among default content, alternate content, and supplemental content. Default content is item information that is presented to an examinee who does not have defined access needs. Typically, default content includes text, graphics, and/or tables that form the item as developed for the general population of examinees.

Alternate content presents a different version of the item to students with specific needs. In essence, some or all of the original default content is replaced by alternate content. Examples of replacements that result in alternate content include a translated version of an item (e.g., Spanish instead of English) or simplifying the language contained in an item (i.e., simplified English).

In contrast, supplemental content supplements rather than replaces default content to address access needs. For example, text displayed as part of default content might be supplemented by presenting an audio, Braille, or signed form of the default content that is

displayed to the examinee. Similarly, to assist the examinee in identifying important aspects of default content, supplementary information may be presented that accompanies the default content (e.g., highlighting key words, translation or definitions for key words, flags that point the student to key information, etc.).

Digital content files allow item developers to specify default content and supplemental content for each item or task. In addition, a digital content file can also provide a pointer to separate files that contain alternate versions of the default content (e.g., Spanish translation). Within a digital file containing an alternate version of an item, default and supplemental content for that alternate version can be provided (e.g., default content presented in Spanish accompanied by supplemental content specifying how to read aloud that content in Spanish). Importantly, specifying default, supplemental, and alternate item information during the item development phase allows item writers to carefully consider whether supplemental and alternate information alters the intended construct measured by the item. When it is determined that the measured construct is altered by supplemental or alternate information, the item writer is then positioned to either modify supplemental or alternate information or to determine that supplemental or alternate information cannot be provided. By making careful decisions about providing supplemental and alternate information during item writing, a test program can ensure that the same high quality supplemental and alternate information is available for all students while assuring that the item still measures the intended construct.

**Digital Item Delivery**

Capitalizing on the flexibility of computer-based technologies, computer-based test delivery interfaces can tailor the presentation of default, alternate, and supplemental item content, interactions with that content, and response modes based on each individual's needs. To do so, developers should employ principles of universal design when creating systems that can personalize the testing experience based on each individual student's needs.

The concept of Universal Design focuses on "the design of products and environments to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design."[ii] Rather than creating a single solution, Universal Design has come to embrace the concept of allowing users to select from multiple alternatives. As Rose and Meyer emphasize, "Universal Design does not imply 'one sizes fits all' but rather acknowledges the need for alternatives to suit many different people's needs…the essence of [Universal Design] is flexibility and the inclusion of alternatives to adapt to the myriad variations in learner needs, styles, and preferences."[iii]

Technology allows developers to apply principles of universal design to educational assessments such that access improves for all users. When building a universally designed educational assessment, there are three important aspects that must be considered.

1. In order to access the intended construct within each student, it must be acknowledged that a single version of an item is not adequate. Instead, supplemental and alternate

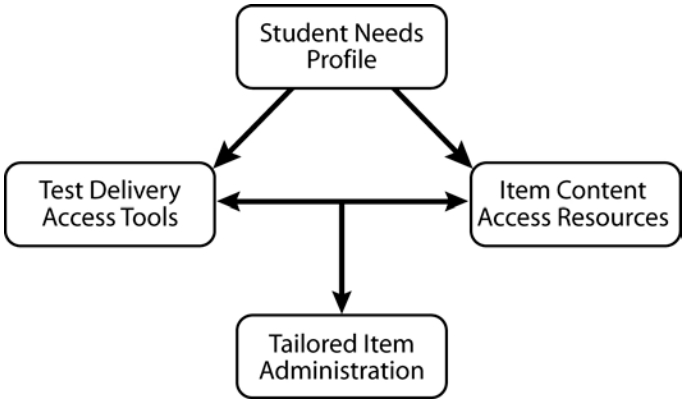
information that is carefully designed to access the intended construct within specific sub-groups of students must be embedded into item content.

2. A user access needs profile must be developed for each student and that profile must specify the method(s) of presenting, interacting, and responding to item content that are expected to best access the intended construct. More specifically, an access profile defines access needs for a given student and indicates which tools and/or representationl forms should be made available for that student. The profile might also specify specific settings, such as magnification levels, color contrasts, or default representational forms preferred by the student. Once defined, an Access Profile interacts with both the delivery interface and the item content.

3. The interface used to deliver items and tasks must be able to interact with each student's access needs profile and with the default, supplemental, and alternate information specified in each item. The interaction with the delivery interface focuses on specific tools or features embedded in the interface, activates those tools and features that are defined in the profile, and, in some cases, controls the exact settings for those tools and features. The interaction with the item content focuses on which of the specific representational forms embedded in the item should be presented and/or activated for a given student in order to meet his/her specific need.

As depicted in Figure 1, these three components must be designed to work together to tailor test delivery to meet the specific access needs of each student.

**Figure 1: Accessible Test Implementation Model**




## Accessible Test Design

Accessible test design addresses item content, representational forms, test delivery interface, and access profiles by specifying methods for flexibly tailoring an item such that the influence of non-targeted constructs is reduced for each individual examinee. Depending on an examinee's access needs, flexible tailoring may require an adaptation to the presentation of item content, the interaction with that content, the response mode, or the representational form in which content is communicated.

Adapted presentation may require item content to be presented in a manner that assists the intake of information, such as magnifying or adjusting the contrast level with which item content is presented to an examinee. Adapted interaction may require changes to the conditions under which an examinee applies the targeted construct, such as decreasing distractions by masking content, providing auditory calming, or highlighting key content within an item. Adapted response may require the examinee to use different types of tools to produce responses, such as a speech-to-text or an assistive communication device. Tailored representations may require item content to be presented using a different representational form, such as Braille, sign, audio, an alternate language (e.g., Spanish), or using simplified vocabulary.

Providing these adaptations and tailored representations in a consistent manner requires careful thought during the item development stage. In order to assure that adaptations and tailored representations do not negatively influence the validity of inferences based on the resulting test score, item developers must specify supplementary and alternate information associated with each item and assure that these alternate representations do not alter the measure of the intended construct. Providing adaptations and tailored representations in a consistent and valid manner also requires assessment programs to take a systems approach to accessibility. When designing a test delivery interface, an assessment program must specify the variety of accessibility tools and features that may be required for specific students. Finally, the test delivery system must be able to integrate examinee access information, item accessibility information, and interface accessibility tools and features to tailor item delivery to meet access needs for each individual examinee. Accessible test design requires a comprehensive model for test design and administration.

Since this approach of developing supplemental and alternate item content is new, a standard for tagging or coding that information is needed so that items are interoperable across delivery systems.

**Accessible Portable Item Profile Standard**

To meet many of these needs, the APIP Project has developed an open-standard called the Accessible Portable Item Profile (APIP) Standard. Led by the MN Department of Education, the APIP Project is a federally funded initiate that includes a consortium of 8 states, IMS Global Learning Consortium (an interoperability standard setting organization), and experts in testing, accessibility, and interoperability standards. Using the concept of accessible test design as a foundation, the APIP standard provides an open-standard for specifying default, supplemental, and alternate item content, and for identifying the access needs for each individual student.

A multi-step collaborative approach was employed to develop APIP. This process included:

a) reviewing existing options for item content standards and identifying QTI 2.1 as the most useful option;
b) reviewing current accommodation policies for all states and compiling a comprehensive list of potential accommodations and access needs;

c) reviewing research and development efforts that focus on meeting access needs not included in current policies (e.g., reduced answer options, scaffolding, etc.) and expanding the list of potential accommodations and access needs;
d) comparing the list of potential accommodation and access needs with the Access for All elements and identifying gaps in those elements;
e) working collaboratively with IMS technical experts to develop the APIP QTI profile and Access for All Personal Preference profile;
f) working with the QTI and Access for All working groups to develop extensions to those tools for the APIP elements;
g) sharing the APIP access elements with 8 participating states and acquiring input from all states on which elements should be required for APIP v1.0, which are optional, and which require further research before integration into future versions of APIP;
h) developing technical documentation of APIP;
i) sharing the APIP technical documentation with the larger community and receiving input on that documentation.

The APIP standard builds on the IMS Global Learning Consortium (IMS GLC) Question and Test Interoperability (QTI) and the Access for All Personal Needs and Preferences (AfA PNP) specifications. QTI is based on an Item model that allows item developers to specify a variety of information about a test item. The APIP standard expands the QTI item model into a comprehensive content accessibility framework. The AfA PNP defines the data model for the exchange of preferences for the use of learning systems. Again, APIP has expanded this specification. A draft of version 1 of the APIP standard was shared with the RTTT Assessment and Alternate Assessment Consortia, test vendors, and other interested parties on December 7, 2010 and has since been out for public comment. Currently, 2 Consortia have stated their intent to adopt APIP as their interoperability standard, one testing company has stated that it is developing APIP compliant solutions, and at least 4 states have or are referencing APIP compliant solutions for their state assessment programs. IMS GLC has also established an APIP working group to continue refining APIP and to create resources to assist in the implementation of APIP.

Version 1 of the APIP is designed to support three use-cases:

a) Importing of APIP Item(s) into an APIP compliant system/application/tool – the reading of an APIP interchange file (a form of 'zip' file) and the storage of the contents of that file in the system/application/tool;

b) Exporting of APIP Item(s) from APIP compliant system/application/tool – the creation of an APIP interchange file that can be stored in an external repository and/or which can be imported into another APIP-compliant system/application/tool;

c) Obtaining APIP personal needs and preferences – reading a set of personal needs and preferences, from a Preferences System, so that a system/tool/application can be configured to render APIP Items in a manner suited to the personal needs and preferences of the user.

A broader set of use-cases have been defined for APIP so that later versions of APIP can be extended without causing incompatibilities with version 1 implementations. For example, later versions of APIP will address test interoperability i.e. an APIP test consisting of several sequenced APIP Items.

The technical specification work has involved the development of extensions to the base IMS specifications and profiling of several IMS and non-IMS specifications. Profiling is the process by which, one or more specifications, are tailored and combined to provide a best practice solution. In the case of APIP, profiling has been completed for the IMS QTIv2.1, IMS AfA PNPv2.0, IMS Content Packaging v1.2, IMS AfA Digital Resource Description (DRD) and IMS Metadata v1.3. Profiling of the IEEE Learning Object metadata (LOM) has also been undertaken.

The technical documentation set for APIP consists of three other documents and the set of XML Schema Description (XSD) files and examples. Namely:

a) Accessible Portable Item Profile (APIP): Technical Specification – containing the detailed profiling of the various interoperability specifications and a description of how systems are expected to process the information contained within the APIP Item and AfA PNP data files. This document is relevant to organizations wishing to adopt the APIP;

b) Accessible Portable Item Profile (APIP): Technical Specification of New Features – containing the detailed description of the information model and XSD binding for the extensions to the IMS QTIv2.1 and IMS AfA PNP v2.0 specifications required by APIP. This document is relevant, in particular, to organizations wishing to implement the APIP and who require the detailed technical descriptions of the new features;

c) Accessible Portable Item Profile (APIP): Best Practice Examples – containing a number of annotated examples of the data files exchanged by APIP compliant systems for the import/export of APIP Items and access to the APIP preferences. These examples demonstrate how the core use-cases are supported by the technical solution. This document is relevant to anyone wishing to understand how the APIP can be used;

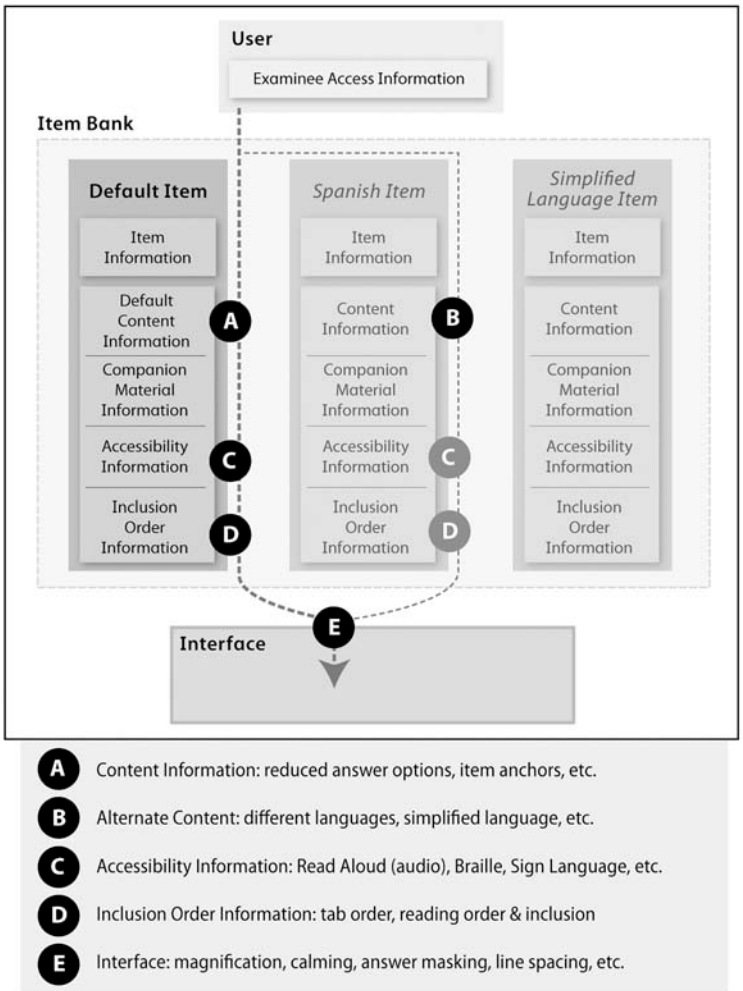
d) XSD files – these are the control validation files used by applications to confirm that the data instance files being exchanged by APIP compliant systems are syntactically correct;

e) Annotated examples – the set of example instance files that are described in the best practices examples document.

As depicted in Figure 2, the APIP framework is comprised of two information models that allow test delivery engines to tailor the presentation of items to meet the access needs of each individual examinee. The first information model focuses on Examinee Access Information. During test delivery, the Examinee Access Information model performs two functions. First, the Examinee Access Information model provides information that allows a test delivery engine to activate specific tools that tailor the presentation of item

content to the examinee. These embedded access features may include magnification, alternate contrast, increased white space, and answer masking. Second, the Examinee Access Information provides information that specifies which accessibility information embedded within the item model is pertinent to the examinee. APIP allows item developers to place a variety of types of access information within an item, including specifications for how an item is to be presented in auditory, Braille, sign, or tactile forms. In addition, the item information model allows an item developer to point to alternate versions of the item that are presented in an alternate language (e.g., Spanish) or in simplified English (e.g., with negatives removed).

**Figure 2: APIP Model**




As depicted in Figure 2, the APIP item model is composed of five components, namely Item Information, Content Information, Companion Material, Accessibility Information, and Inclusion Order. Each component is described briefly.

The Item Information component provides meta-information about an item. Examples of meta-information include the domain, sub-domain, intended construct, intended grade/age level, item difficulty, item discrimination, item exposure rate, etc. In addition, the Item Information component contains information about alternate versions of the item for which item content has been substituted to make the item accessible in a different language or as a simplified version of the item in the default language (e.g., simplified English).

The second component, Content Information, provides information about the contents of the item that are to be presented to an examinee assuming no access needs have been defined for that examinee. This information includes the item type, the prompt, media associated with the item (e.g., figures, tables, and graphs), response options, correct response, scoring rule, etc.

The third component, Companion Material Information, provides information about materials and tools that the examinee is expected to work with while performing the item. These materials and tools may include a reading passage, a primary document, a periodic table, ruler, protractor, calculator, etc.

The fourth component, Accessibility Information, provides information about alternate representations of default content. The types of alternate representations specified in the Accessibility Information component include audio, signed, or tactile presentation of item content. The Accessibility Information component may also include specifications for scaffold supports or key word translations.

The fifth and final component, Inclusion Order, specifies the order in which accessibility elements are to be presented to an examinee with a given category of access needs. The seven categories of access needs for which an inclusion order is specified include: 1) Text-based Audio Access; 2) Graphic-only Audio Access; 3) Text-based and Graphic Audio Access; 4) Blind Audio Access; 5) American Sign Language; 6) Braille Access; and 7) Other Sign Language. Text-based audio access is specific to examinees who are able to view contents displayed on a screen, but may need assistance accessing those contents. Often, examinees requiring text-based audio access read below grade level, have reading related disabilities, or the language in which item content is displayed is not their primary language. When audio forms are presented as core content, the item developer must specify the order in which content is presented. In addition, the item writer must consider whether all item elements are presented as core content in audio form. As an example, an item writer may opt not to present labels associated with a graphical element as part of the core content, or may opt not to read the contents of a table when the item is read from beginning to end. In these cases, item content that is not presented as core content may be accessed by the examinee on-demand.

More specifically, the tables in Appendix A summarize the supplemental accessibility information supported by APIP v1.0 and the access needs included in the APIP v1.0 access needs profile component.

In sum, the APIP model empowers item developers and testing programs with a standard method for specifying the tailoring of items to meet specific accessibility needs, and provides a foundation for accessible test design.

**Next Steps**

Developed over an 18-month period by experts in interoperability standards, assessment, computer-based test delivery, accessibility, and major state assessment stakeholders, APIP is the most comprehensive item content standard currently available. Rather than starting from scratch, APIP builds on several existing tools, including QTI 2.1, Access for All, and the Common Cartridge, each of which are familiar and already employed by different sectors of the industry. While APIP has not yet addressed the test model, by incorporating elements of QTI into the APIP item model, APIP is positioned to add the QTI test model. Clearly, performing this work is an important next step for APIP.

Additional next steps include the need to identify and develop touch points between the APIP student access needs model and other standards, such as SIF and/or Common Data Standards, so that the access needs of students is associated with other student information and that these access needs can be employed by test delivery systems to provide accessible test administration.

In addition, given the power that APIP provides item writers to specify supplemental and alternate item content, best practices for how to apply these tools must be developed. Doing so will enable the APIP interoperability standard to be implemented in the highest quality manner so that its goal of improving the validity of student assessment while enabling the portability of assessment items and tasks is realized.

Finally, efforts are needed to maintain and extend APIP. In particular, there will be a critical need to expand the item types supported by APIP as new innovations in item types are brought to the market. In it is current state, APIP supports item types that are typically found on state assessments and which have been demonstrated to support valid assessment of a wide variety of constructs associated with content standards. Currently, however, there are several efforts to develop new item types to measure constructs that are not measured well by the current item types. Clearly, it would be premature to add these new item types to the APIP model before they have been demonstrated to be of value and are ready to be used in a scaled manner. However, once a new item type reaches a mature state, it will be important to extend the current item types supported by APIP to include a mature new item type. While there is some concern about non-standard extensions, adding a mature new item type to APIP in a purposeful, thoughtful, and open manner would not constitute a non-standard extension.

---

[i] Mislevy, R.J., Behrens, J.T., Bennett, R.E., Demark, S.F., Frezzo, D.C., Levy, R.,

Robinson, D.H., Rutstein, D.W., Shute, V.J., Stanley, K., & Winters, F.I. (2010). On the Roles of External Knowledge Representations in Assessment Design. *Journal of Technology, Learning, and Assessment, 8*(2). Retrieved January 21, 2010 from http://www.jtla.org.

[ii] Center for Universal Design (CUD). 1997. About UD: Universal design principles. http://www.design.ncsu.edu/cud/about_ud/udprincipleshtmlformat.html (accessed February 13, 2009). Archived at http://www.webcitation.org/5eZBa9RhJ.
[iii] Quoted from page 4, Rose, D., and A. Meyer. 2000. Universal design for learning, associate editor column. Journal of Special Education Technology 15 (1): 66-67.