Open Research Online

Assessment Interoperability

Produced by the European IMS Network (EIN)

1. Introduction

Assessment is a term covering a range of educational activities. The broadest interpretation includes the processes of testing and any consequent judgements and decisions. In ths chapter, the focus is on the processes of computer based testing for summative and formative purposes.

Assessment interoperability has made measured progress over the last five years, since international specifications first became available. As with many reform activities, early attention has been focused on the most tractable issue - the lowest hanging fruit. In this case, the main interest has been on achieving basic interoperability at a granular level of test items based on electronic question types that have been in vogue for over a quarter of a century. The specification developed is capable of supporting present practice in this area.

Complementary developments are starting to create new imperatives. With the first broad adoption of content interoperability, practitioners and policy makers have begun seeking greater detail and reporting of summative assessment data in order to facilitate accountability and to underpin the provision of customized learning programs for individual learners. Demands for efficiency have brought to the foreground the formative role of assessment. Innovation prompted by research or the desire for differentiation has led to new assessment models and techniques that are not currently supported by open specifications, e.g. where the assessment is integrated within the learning process with input from the learner and their peers. Additionally, new online learning environments, such as online gaming, simulation, and collaborative spaces, require different approaches to assessment so that we assess in context. Assessment methods and assessment interoperability in these areas are underdeveloped.

At present electronic assessment is most likely to be encountered as a set of discrete tasks presented within the context of an adopted Virtual Learning Environment VLE. The task is based on either the facilities provided by the VLE or a compatible but discrete plug-in system specialized for delivery of assessment. Although the most common VLEs have limited electronic assessment capabilities, market pressure is requiring VLE providers to augment their internal functionality so that customers can present their programs within a monolithic VLE context. However, within the evolving assessment marketplace, new models for the delivery of assessments are coming to the fore. These are based on integrated Web services provided by a variety of organizations rather than vertically integrated offerings from a single vendor.

The assessment practices described above have conspired to limit interest in operability of test items across delivery platforms. The institutional tendency to adopt a specific VLE, the limited availability of central assessment resource of perceived value, the vendor interest in locking in customers to their product, and the growth of Web delivered assessment services - have all impeded the drive towards assessment interoperability. Without intervention, the effectiveness of drivers for interoperability in future marketplaces may well remain limited.

2. State of the art concerning standards and interoperability

The de-facto standard for assessment interoperability is provided by the IMS Question & Test Interoperability (QTI) specification. This describes a model for the representation of questions, test and the reporting of their results. Therefore, the specification facilitates exchange between item and test authoring tools, item banks, learning and assessment systems. QTI V2 has been released and V2.1 exists in public draft form.

There has been an ambiguous level of acceptance of QTI by the developer community, particularly by developers of major tools such as Questionmark, [WebCT], Moodle, etc., and, to date, interoperability has not been as high a priority as public statements suggest. The market evidence is that customers are most comfortable with monolithic systems that offer the end-user integrated functionality and that developers respond to this by aspiring to but not prioritizing interoperability.

It is helpful to look at key parts of the lifecycle of an assessment to see where "standards" fit in and where interoperability might be an issue.

Assessment creation and re-use. To exchange, edit, or disaggregate assessment items in a variety of different tools, the items need to be described in a common format. The IMS Question and Test Interoperability Specification is used to describe the content of common assessment types, such as multiple choice, true/false, etc. A number of vendor's tools implement the import and/or export of assessment in "QTI format." The IMS QTI v2 specification addresses many issues raised by implementers over the last 3-4 years of use of the QTI specification. This includes but is not limited to more complex question types and support for mathematical representations and processing. The current V2 specification addresses Item interoperability. The next addition to the V2 specification, currently in development, will address updates to QTI's Test interoperability. IMS QTI has been designed for multiple language settings. Whether its language support is sufficiently adequate or not requires more investigation.

Assessment Discovery. If you maintain a searchable collection of assessments for internal or outside use, then the assessments need to be tagged just as a library assigns card catalog information to books to enable their discovery. The IEEE Learning Object Metadata Standard, also known as "LOM," defines the names and meanings of the tags used to describe electronic information used in learning and training, including assessments. A LOM application profile for QTI V2 has been developed as an adjunct to the Colleges Open Learning Exchange Group COLEG Online Assessment Project that generated an item bank of formative assessment instruments for Scottish colleges. The Dublin Core metadata specification, developed by the Library community, is also commonly used in e-learning.

Usage data is an essential element of assessment discovery. It links items with the validation process and is a necessary element in the construction of eportfolios and reflective learning. However, caution is needed as such data, e.g. statistics of student responses, are context sensitive. QTI V2 recognises this issue.

If there is a need to describe the competency definition or educational objective, then the IMS Reusable Definition of Competency or Educational Objective (RDCEO) specification can be used. This is discussed further in the Chapter on Learner Informaton where RDCEO's relevence to prtfiolio assessment is highlghted.

Assessment ordering. Once an assessment is discovered, the HR-XML Assessments and Assessment Catalog specifications support ordering assessments in a standard way that a tool could facilitate. This would allow the automation of parts of the order fulfillment process.

Assessment distribution. Distributing an assessment to a teaching or training organization or business partner is much like sending a physical product. You need a way to package up all of the related files, label the assessment, and send it in a way that the receiver knows how to inspect and use the contents. The IMS Content Packaging specification is designed to support the electronic distribution of learning and training materials. Most e-learning tools support this specification, and it is the basis for work from the US Department of Defense's Advanced Distributed Learning (ADL) effort most commonly known as SCORM (Sharable Content Object Reference Model).

SCORM does not support IMS QTI, meaning it is not possible to "look inside" a SCORM-compatible piece of content and see individual assessment items; however, IMS has described the appropriate way to include QTI-based assessments in an IMS Package in the QTI V2.1 specification.

Assessment reporting and status. Standards for reporting assessment status and outcomes have generated the greatest level of interest. They enable the management of training without requiring assessment developers to expose the intellectual property that makes up the assessment. There are a number of standard ways to report status and/or results. Only simple results and status are described. This includes; IMS QTI, the Aviation Industry Computer based training Committee AICC standard, the Sharable

Content Object Reference Model SCORM, the Human Resource Extensible Mark-up Language HR-XML, and the Schools Interoperability Framework SIF (from the K-12 community). In the final case, the reporting supports the US Government No Child Left Behind initiative.

Besides defining "what" to exchange, the aviation industry has also defined "how" to move assessment results and status between systems. SCORM uses a modified version of the AICC approach and it is a cornerstone of SCORM's ability to provide an interoperable content model that separates content from Learning Management Systems. The SCORM version of the AICC work has been standardized within the IEEE.

For content developers using simulation-based content, the IMS Shareable State Persistence (SSP) specification was created to work in conjunction with the AICC (IEEE CMI) specification to enable more detailed reporting of simulation data that might include assessment information or data that could be used for assessment purposes. The specification does not provide any semantics about the detailed data and therefore the utility of the specification is limited to single companies or trading partners that share the internal structure of their information that is exchanged according to the SSP approach.

3. Demands, Requirements and Use Cases

The assessment perspectives and the consequent interoperability needs of different stakeholders are described below.

Governments and their agencies. At the policy level, the outputs of assessment and public perceptions are the dominant concerns. Assessment for summative purposes must be efficient, discriminating, robust, and trusted by the public. The outputs must be recognised as broadly as possible and be shareable between the educational and emloyment domains.

Greater reliance on computer based testing is seen as a means of increasing efficiency and objectivity, and exhortations and directives to adopt such methods are appearing in national policy papers, e.g in the drive to adopt 'screen-based assessment' by the UK's Qualifications and Curriculum Authority.

It is clear that effective interoperability could assist in meeting policy aspirations but interoperability has been seen as an operational matter and there has been little policy level involvement its promotion.

Teaching and training organizations. These groups are seeking greater reusability of items and tests and access to item banks from both internal and distributed collections. However, there are varying degrees of confidence in the relevance and applicability of such collections, with use most developed in areas with unified curricula and centralized control over summative procedures. Use of those banks that do exist has been hampered by the difficulties, real and perceived, of importing items into institutionally specific tests.

There is a demand for more advanced types of assessments and a need to develop greater detail in and a shared understanding of the reporting on the competencies that will be tracked within the ePortfolios or learning portfolios required for mobile workforce development and customized learning and training.

Teaching and training organizations seek to have more options with respect to the vendors and tools that they choose from to provide a learning experience for their learners. However, to date, interoperability has not been a foreground issue in the attempt to exercise flexibility.

Subject Communities. Communities of practitoners have a shared interest in promoting their subject and their colleagues. Assessment is seen as an essential means of ensuring professional standards but, also, as a burden. So, alongside all other groups, subject communities are seeking efficiency in assessment delivery. The re-use of assessment resources is recognised by some as a means of developing professional skills in assessment and oas away of harmonising standards within the subject.

Content and Test Developers. Content developers are seeking a common framework for the structuring and delivery of content. They see assessment interoperability as a part of that common approach. Recently, adoption of item level interoperability has been motivated by the promise of internal operational efficiencies, e.g. by sharing assessments across book and online course production processes.

The large test publishing firms have begun to respond to internal needs for greater efficiency and to acknowledge customer demands for open specification support. In a recent Association of Test Publishers Annual Conference, a major conference program element "Interviews with the Titans of Assessment", a senior Vice President of one of the world's largest publishers listed 3 major trends in 2005, one of which was the use of an open specification (IMS QTI) for assessment interoperability.

Test publishers have historically competed by providing vertically integrated services and technology and have offered advanced assessment types, e.g. spreadsheet assessment tasks, advanced drag and drop, variable replacement. Major publishers have moved towards more open specifications though to date this move has been limited in the main to earlier and less advanced interoperability specifications. This might imply that customers have happily or grudgingly accepted lesser functional assessments that are more broadly interoperable versus higher function, lower interoperable assessments. However, it is not clear that interoperability has been a significant factor in the purchasing decision and the customer and vendor behaviours may reflect a trend to a broader and larger market having more limited assessment strategy aspirations.

Test Authoring and Delivery Tool Vendors. Test authoring and delivery tool vendors, like most tool vendors, are caught between the common need to differentiate their product and an increasing call for interoperability from consumers. The open specifications and standards do not support the full range of advanced assessment types, or advanced assessment authoring, or advanced delivery, e.g. local, hosted, and hybrid assessment delivery models where an outsourced company might provide essay grading services. The advantage from a requirements perspective is that there is substantial practice instead of wishful thinking that can inform requirements collection.

4. Relationship with services/components

There is a need for harmonisation of assessment standards and specifications with related specifications; for example, standard results formats for sharing with eportfolios, semantic interoperability with Learning Design, etc. The reason is obvious. Data exchanges within a system which are using two different specifications (for example QTI and learner information profiling) may re-interpret data in different manners, with differing assumptions. The most relevant of these issues are dicussed in Section 2.

The following paragraphs describe areas of specific present concern and future relevance in the development of assessment interoperability.

Collaboration between vendors and publishers. The IMS Common Cartridge Specification has not been released yet but signals the willingness of publishers and software vendors to work more closely together. It will offer a standard format for the reliable distribution of rich content and assessments across platforms. It will facilitate the integration of diverse content, assessments, discussion groups and third-party services into a unified packaging structure. Content providers (both commercial publishers and the open content community) will benefit from substantially reduced production, testing and distribution costs. Education and training providers will benefit from a broader choice of content on offer and being able to mix-and-match content from different sources. VLE/LMS providers will benefit from a broader complement of content for their platform and reduced production and testing costs. Once a "cartridge" is imported into a VLE/LMS, additional local assessment types can be added to the ones supported by the specification.

Granular Learning Objects. There is an increasing level of interest in learning object repositories, either local or distributed. Today's bulky (low granularity) content objects assume an integrated content and assessment model with summative assessment reporting to an LMS/VLE. Should e-Learning content repositories become wide spread, then more detailed assessment of learning may lead to separation of assessment from content with the reduced need for complicated pedagogical sequencing and thereby assume a more just-in-time and targeted assessment for learning or remediation approach.

Gaming, Simulation, and Collaboration. It is a common refrain that it is a good thing that the basic training model of twenty years ago has, to a sufficient degree, been captured in open specifications. At the same time, most would say that it is critical to create learning opportunities that deliver additional learning outcomes. Current open content and assessment models do not support the approaches that are being

adopted to cope with these new aspirations. For example, gaming and simulation, whether single or multi-player, are increasingly looked at as trend setters and as key to the next generation of learning strategies. With a new generation of learners completely at ease with Short Messaging Service SMS and online gaming, learner demand and their ability to use these types of environment for educational purposes can be expected. Assessment approaches must keep up to accommodate testing within a context sufficiently similar to the learning context. It is an open question as to whether such developments will promote the sharing of assessments or items. However, initially at least, it is doubtful as any innovation is unlikely to be captured in specifications until there is wider uptake.

Personalization. With the advent of ePortfolios, the much awaited goal of personalized learning becomes feasible in a much more interoperable manner. Likewise, the corporate organizations who need to be efficient in re/training workers in a fast moving economy demand greater understanding of a worker's competencies. Again, ePortfolios provide an important means of collecting information to make these business outcomes a reality. What is missing is agreement on ontologies and vocabularies. Assessment approaches need to be able to sufficiently integrate with the demands of ePortfolio applications in these personalized learning and training contexts.

5. Trends and Issues

The long term trends and issues in assessment interoperability will be governed by the evolving role and nature of assessment itself. Therefore, this section provides a general account of those assessment trends and issues that will affect the demand for interoperability.

Much of the present discussion of interoperability is framed in terms of technical specifications relating to a limited part of the existing repertoire of assessment practice, i.e. automated computer-mediated testing. However, assessment that is authentic will continue to be based around a much broader range of diverse tasks. These include; the production of specified documents (e.g. essays, reports etc.), engagement in real or simulated professional roles (e.g. laboratories, fieldwork, training roles in organisation, virtual practice environments, etc.), public performance (e.g. in giving presentations, playing music or sport, etc.), self reflection (e.g. in portfolio construction), and reporting on prior experience. This isn't an exhaustive list. Given the imperative for personalised learning and authentic assessment, it is likely that assessment will reflect increasingly local and personal context. Interoperability can be thought of as an expression of the need for such pedagogical imperatives to be reconciled with the needs for broad metrological validity and accountability.

There is an intimate relationship between the summative and formative functions of assessment. This is evident in the tendency for students to adopt a pattern of study that is strategic in the sense that it is geared towards meeting the summative requirements. It follows that assessment should be designed so that it encourages effective learning and interfaces appropriately with support functions. Mechanisms for interoperability should support this integration of assessment within a broader learning process.

It is interesting to consider where a desire for assessment interoperability is visible in the present or future market place. Assessment is profoundly culturally related with nations, institutions and individuals insistent on retaining their control over summative assessment content, process and accreditation. The degree to which this is exerted tends to vary with level and is most acute at the tertiary level, where the concept of the self directing university and academic is dominant, at least in areas where there is not a strong external driver (e.g. a professional accreditation agency such as a professional body or a powerful vendor, e.g. Cisco). It follows that there is major reluctance to recognise or engage with the exchange of assessment content, process or even information.

A related issue is the sharing of understanding or documentation of levels of achievement. In spite of long standing attempts to codify local norms and generate broadly recognised benchmarks, genuine progress has been limited. The extent of the failure to arrive at operationally effective understandings is often concealed as it is potentially embarrassing at political, institutional, directorial and individual levels. How can interoperability at summative level operate effectively without a shared understanding of achievement and the description of achievement?

Similar arguments surround the pedagogy within which the assessment lies. There are major differences in expectation of and judgements about teaching and learning styles.

It is possible to have greater optimism about interoperability in assessment for learning, not least because there is a long history of interoperable assessment use in end of chapter questions and other prosaic resources. Formative assessment has fewer issues of ownership, reuse and security. The main issues are perceived relevance and benefit, not least in increasing the efficiency of the teaching process.

Within the narrower CBA/CAA domain, there are several possible future models of assessment practice with different issues for interoperability. It is unclear whether a model in which assessment is distributed and implemented on a variety of platforms is sustainable. Instead we might expect assessment to be a service or solution delivered via networks from a central source, just like a map or an auction. This model, in limited but widely adopted form, is common in the United States where it is seen as an adjunct to publishing. For example the uptake of Wiley Plus with its well developed formative assessment capability is growing rapidly. The challenge may be to offer a richer service, which, for example, includes the wider functions of a community of learning, while retaining cultural diversity and flexibility. Moodle is an example of an open source initiative that embodies such an aspiration.

The roles and levels of competence of those involved in generating and delivering assessment will affect interoperability. As standards, specifications and authoring aspirations increase in complexity, in order to reflect more complex assessment strategies, it becomes increasingly difficult to envisage a process in which either the 'teacher' has the technical competence or a 'learning engineer' has the pedagogical skills to deliver an appropriate solution. The community will be driven towards team approaches.

6. Stakeholders Concerns/Issues

The issues, concerns, drivers and obstacles surrounding the adoption of assessment interoperability were discussed by a diverse group of assessment experts. The following sections summarise their views and, where appropriate, the consensus reached.

There is a concern expressed by many actors across all sectors about the market penetration and operational value of the existing QTI standard and the demand for the imminent QTI V2.1. It is broadly acknowledged that the specification provides a comprehensive account of practice, as well as the means to transform most specified task into interoperable items and items into tests. As already noted, the context here is restricted to the CBA/CAA domain. However, there are a range of worries about operational uptake and the perception of a gulf between the specification and practice.

It has been suggested that the pace of development of specifications tends to be slower than that of commercial products - arguably this is inevitable as there must be a practice before one can attempt to standardize it - and that QTI is not sufficiently comprehensive in the range of question types to satisfy those devising commercial systems. This is difficult to sustain in regard to QTI 2.1 which offers means of encoding a formidable range of potential tasks.

At present, compliance is not seen as adequate. The development of international standards is on-going with varying interpretations of the existing standards. There is no definition of compliance and no formal accreditation procedure. This allows products to claim standards compliance without true interoperability. Although, in some cases, interoperability can be recovered by the intervention of interfacing tools, this complexity and/or shortfall devalues the notion of compliance.

It has been argued that this problem would be eased by increasing the stability of standards and specifications such as IMS QTI 2.1, Content Packaging v1.2, etc. This would allow the advantages of interoperability to become manifest.

A significant view is that compliance may fall further. Although, at the outset, vendors were a key participant in the drive for standards development, this has declined of late, possibly because of a lack of customer demand for QTI Version 2 functionality This could result in a divergence between technical specifications and actual products.

A more optimistic view is that the interoperability will be driven by efficiency demands. It is widely accepted that the development of high quality items is resource intensive. In making this comment one is distinguishing between the trivially authored items that demand only recall and are frequently and appropriately used to secure attention and check for knowledge acquisition, and the more cognitively demanding items that involve the student in a formative process. By ensuring that such advanced items are standards compliant the academic and technical investment made in their production is protected, so long as there is a market for the item. Within this view, the market opportunity is sufficient to force the adoption of standards. This is controversial.

There are two new areas of item functionality that are of increasing interest and potential significance. The overwhelming majority of current electronic assessment tasks involve closed responses to circumscribed questions, e.g. multiple choice questions. Such approaches are entirely appropriate in some contexts but have limited authenticity and relevance outside the broad sciences. The issue then arises as to whether a technology can be developed for marking and providing feedback on more open responses. Significant progress has been made at the level of both short answer and full essay responses by both academic and commercial developers, such as Intelligent Assessment and Educational Testing Services, and systems are now being used for high stakes assessment and for progress checking in self directed learning. However, these developments are not matched by standard and specification development.

The second area concerns skills acquisition. There is a need to develop assessment procedures, including selective and diagnostic testing, that optimize skill acquisition in complex task performances. This area has been neglected to date and further analysis is required to establish whether existing interoperability standards provide adequate functionality in this area.

A further range of concerns centre on the core question of ensuring that we fully understand the meaning of a score achieved on an assessment that is drawn from an interoperable source. This will involve expanding the concept of validity to incorporate systematic influences on educational systems.

It is broadly acknowledged that, within a scenario of broad usage of items and or tests, work will be required to establish procedures for item production, to maintain a valid bank of items for test construction with the appropriate statistics, and to govern item exposure control. In each case the appropriate level of security must be controlled and recorded.

It is also argued that research is needed to establish the appropriate strategies for automated scoring of the complex item types that are or will be typical of performance assessments, and that improved multidimensional psychometric models will be required to accommodate the needs of performance assessment tasks administered in computer-based environments.

In the longer term, a range of new challenges will emerge as we seek to categorise performance in ways that feed into useful prognoses. Issues that should be addressed include: limits on the generalization of our approaches, their value in indicating broader constructs, the development of new psychometrically based cognitive diagnostic models, the addressing of new equity issues, investigation of the problem of large error in ability estimates due to fewer items, and overcoming problems with current differential item functioning DIF analysis models to include constructed response items. Issues such as these are highly relevant to the interoperability requirements of assessment result reporting.

Within the formative assessment domain, the development of multidimensional latent trait models that capture performance modifiability will provide a means of increasing the impact of assessment results on the topic of assessment. Use of such models may introduce interoperability complexities.

7. Interoperability Scenarios

Eassessment interoperability will only be implemented comprehensively if there is a market for it. The evidence for such a market is limited at present. Demand will depend on the perception of the value of interoperability which in turn will depend on the linkage to the wider learning process and therefore on the wider issue of interoperability in learning objects.

If this chain of causality is accepted, it follows that a demonstration of an implemented interoperable assessment system drawing on diverse sources of content and with tangible benefit would be of immense value in creating market momentum. Where might progress be made in creating such a system? The drivers and obstacles in deploying electronic summative and formative assessment across three education/training sectors is analysed below.

Some clarification of what is meant be formative assessment might be helpful here. It should be noted that assessment tasks can have both formative and summative functions. For the present analysis, formative assessment is considered to that which has no substantive role in the summative evaluation of performance and therefore is not affected by issues of confidentiality etc. Of course, summative processes have strong formative influence but this is not relevant to the pragmatic market sector definition we are adopting. In general there are greater opportunities for interoperable systems in assessment as part of a learning process rather than in summative processes. This is due to the added complexity of summative assessment which needs a more careful stakeholder analysis in order to establish interoperability needs more precisely.

School Sector - formative assessment

Drivers.

> There is a lack of practitioner time and limited authorship skills in the devising of CBA/CAA items and tests and the consequent need to focus authorship on a core group and share resources.
> Even at single nation level, the school sector involves a large volume activity with some consistency of relevant task. This enhances market potential.
> The cost of generating high quality items is high and therefore there is incentive to acquire such items from centralised sources.
> Formative electronic assessment resources can allow the teacher to define indvidualised learning paths within the overall constraints of any externally imposed curriculum.

Obstacles.

> Quality assured questions are expensive to generate and therefore are likely to be of high cost.
> There tends to be limited resource purchasing power at local level as available funds are focused on staff.
> National teaching and assessment practice varies and therefore international usage would require customisation to deal with language and cultural issues.
> The tendency to 'teach to the test' tends to drive formative assessment into resembling summative assessment for which practice varies between nations and, in some cases, regions. It follows that formative items and tests will be subject to similar contextualisation demands as summative assessment. Such non-standardisation limits the advantages conferred by interoperability.

School sector - summative assessment

Many of the formative assessment drivers and obstacles noted above are relevant. In addition there are the following.

Drivers

> There is a political imperative for shared understanding of achievement.

Obstacles

> The metrological calibration depends on the population and therefore there are practical limitations on interoperability. However, it might be argued that the possibility of adopting and making use of standards might encourage accreditation bodies to adopt a wider perspective in their processes.

University sector

Several of the drivers mentioned above apply to this sector, particularly the significant potential for formative electronic assessment. In addition, there are the following sector-specific factors.

A strong driver is provided by the trend towards self-directed learning with a consequent need for independent access to learning resources and the means to check on one's own progress. An obstacle is

created by the tendency for universities to favour bespoke locally-generated solutions, even though this limits the feasibility of using complex approaches and may require a large investment of academic time. The strong local ownership of the accreditation process militates against interoperable summative assessment. There is a lack of a credible business model supporting interoperable university level summative assessment systems.

Professional training
Specific issues here include; the large well defined markets, the limited local authorship base, and the strong emphasis on summative processes. In this sector, the presence of influential and often monopolistic stakeholders, who often link summative assessment outcomes with freedom to practice, erodes interest in interoperability.

Conclusion. The above analyses, though incomplete, suggest that the most promising area in which interoperable assessment systems might demonstrate value is formative assessment for all sectors but most particularly, given the size of the potential market, the school sector.

Further analysis of the school sector and formative assessment suggest a possible model for an item bank based system that has the potential to win sector support.

A possible scenario. Primary authorship of an item might be separated from 'item dressing' which locates the item within local learning pathways. For example, an initial or primary author, guided by a specification might produce a 'bare item' that includes all high cost production elements and captures the pedagogical aspiration in devising a complex assessment task that address higher order outcomes. To be useful, the item should carry metadata linking it to local curricula and to summative tests. This may be easier in some subjects, e.g. computing, maths, science, than others involving more subjective discourse. Subsequent authors could take the 'bare item' and refine it in the light of the local context, e.g. by translating the text, linking to the local curriculum, providing appropriate feedback etc. This process of refinement should be simple and would be facilitated by authoring tools. This sequential authorship model would pose new issues in interoperability.

There are several options within this model. For example, there would be a significant difference between a small volume high production-value bank and a massive item bank with inbuilt item redundancy. Qualitative feedback such as might be carried through an annotation category in metadata or Amazon/Trip-Advisor style ratings would add value to the collection.

Central coordination would be required alongside appropriate and authoritative quality assurance mechanisms. One authorship mechanism that would help in generating widespread ownership would involve the pooling of items from a network of authors. Content would be vital and there is a strong market-development argument for the use of public funding in generating such content within a demonstrator.

Looking beyond the specific scenario offered above, it is clear that interoperability would be promoted by the development and support of free open source tools and content. The free availability of such resources would encourage users to invest the time needed to draw together assessment from multiple sources rather than working within the present framework that tends to favour monolithic solutions in which interoperability is of limited relevance.

It is also clear that any plausible interoperability scenario would include a continuing need for the sharing of eassessment knowledge and expertise and (innovative) practice. This might be achieved through a European eAssessment network, as well as informal mentoring and support processes.

8. Relationships with the adoption life cycle

Within the description offered by the Centre for Educational Technology Interoperability Standards CETIS model of the adoption life cycle, significant progress has been noted in identifying and inputting assessment interoperability needs for conventionally framed CBA and in implementing relevant specifications, tools and systems. The main issues in adoption of assessment interoperability are cultural and are found in establishing good practice, raising awareness etc.

It is our view that continuing acceptance and promotion of interoperability by organisations that support and fund development work (i.e. the Joint Information Systems Committee JISC in the UK, the similar SURF Foundation in the Netherlands) is essential. As yet, ownership is insufficiently embedded for interoperability, even at the present level, to be self sustaining. Support is needed, from either the standards bodies themselves or related organisations such as CETIS. Necessary activities for users include developer fora, training, and conferences. Also, where appropriate, the development, dissemination and maintenance of application profiles and mappings/crosswalks will be necessary.

The following drivers and obstacles are seen as being key to the promotion of assessment interoperability.

Drivers. At a technical level, interoperability will be promoted by having an effective and robust infrastructure, with useable tools, particularly if these are free and open source tools which are adequately supported by their developers and/or a broad community of users and enhancers. Effective usage data and metadata will help users to identify items or tests of value. As already noted standards and specifications should be robust, stable, and extensive enough to encompass practice. The available standards will need further development to meet a growing range of needs, for example in dealing with advanced question types or the reporting of competencies for which the evidence is portfolio based. At a relatively trivial level, increases in communication bandwidth will allow a greater complexity of assessment task and more effective links into the 'community of learning'. This will make it more realistic to plan for a geographically dispersed student body such as already exist in the transnational universities such as the University of Phoenix and the Arab Open University.

Policy and regulation choices will also affect adoption. For example, Government or international policy could force major change in the market. An example of this is the requirement for UK school-level exam boards to introduce on-screen assessment. Government might also provide incentives, e.g. funding for development tied to the use of interoperability standards, and the sharing of open source outputs. These may be motivated by a wish to achieve demonstrable savings in resource or time. A factor that is already relevant is the need for vendors and customers to address the accessibility needs enshrined in existing legislation. Adoption is subject to factors internal to education. Training and support for teachers is essential. Knowledge of how to devise effective electronic assessment or how to integrate it within an effective learning experience is limited. This weakens demand for the import of relevant resources. The existence of useable and reusable content, perhaps within a Creative Commons or similar license arrangement might motivate staff development in this area. The economics of coping with demands for personalised learning with timetable flexibility are increasing. Imported solutions offer one means of coping with such demands. The concern about how to deal with plagiarism at all levels could also provide a driver for adoption in that the quality accreditation process may increasingly require conformity with agreed (inter)national standards of both assessment system design and plagiarism checking.

Finally, as always there are potential economic drivers for adoption. Notable among these is the fear of vendor lock-in that should encourage teaching institutions as well as awarding bodies to pursue interoperability.

Obstacles. Set against these drivers are several obstacles to the adoption of interoperability. It is acknowledged by a majority that the standards domain is opaque. The specifications are complex and not user-friendly. This would be of little consequence if such matters could be delegated to large vendors but this is an area of distributed authorship. It is also an area where competing commercial products have tended and may continue to add features rather than focus on compatibility with interoperability standards. In the professional training sector, powerful and 'monopolistic' vendors whose interest is in training that is narrowly based around their products/services will have limited incentive to create interoperable systems.

The already noted lack of expertise in the combination of technical skills and pedagogical awareness may be expected to lead to difficulty in recruiting appropriate personnel to develop interoperable systems or content. Obstacles to interoperability are connected to obstacles to the adoption of computer aided assessment in that scale of use creates the potential market for interoperability. The appropriate domain for use of computer based assessment is contested though it is commonly viewed as an element within a blended learning approach. There are limited levels of expertise in authoring questions that test higher level outcomes and, at

present, a range of interoperable question types that do not include, except at an experimental level, open ended student responses, e.g. essays. It follows that the interest in interoperability in these technically advanced areas cannot be developed.

Finally it is worth noting that there is a lack of a broadly shared epistemology of assessment. Without such a 'map' it is difficult to generate the momentum for the large reforms that would create a climate more conducive to the adoption of interoperability.

9. Recommendations

The recommendations that follow reflect the restricted scope of the discussions in the chapter, which has been concerned mainly with; computer based assessment, the adoption of relevant standards and the encouragement of effective practice. It reflects too the advanced level of development of the QTI specification and the view that adoption of interoperable resources should be placed in the foreground. There are many other aspects of assessment that pose interoperability issues. Some of these are discussed in other chapters. For example, in the chapter on Learner Information, there is a discussion of the recording and sharing of competencies, e.g. using the IMS Reusable Definition of Competency or Educational Objective, and related recommendations. The Learning Activity chapter raises issues about the design of learning, within which the driving role of assessment must be recognised, whilst the chapter on Learning Object Repositories includes analyses that are relevant to assessment task repositories or item-banks.

Recommendations 1 to 4 are aimed at promoting adoption of interoperable assessment resources and systems. The effects could be felt in the near future.

1. Interoperability will be most effectively promoted by the demonstration of its utility to users who shape the market for assessment products and systems. Therefore, projects should be undertaken that will yield content of demonstrable value across institutional and national boundaries and which will demonstrate that a market for interoperable resources and systems can be created. These projects should be focused on formative assessment resources for all educational sectors with priority given to the largest markets.
2. The resources developed in such projects should demonstrate the versatility of computer based assessment in evaluating achievement and extend the boundaries of the design of assessment. The authoring systems should allow two-step authorship. The first step would generate resources of sound pedagogical design and incorporate high production cost elements. The second step would allow refinement of the items for the local context.
3. EIN and other potential actors should be urged to develop a fuller specification for the above proposal and identify funding routes.
4. The above demonstrator projects may generate a demand for an ongoing service in which assessment resources generated via distributed input are delivered or disseminated. Further discussions should be facilitated (BY WHOM) on how such a service could be developed and sustained.

Recommendations 5 and 6 are aimed at influencing stakeholder perceptions of interoperable computer based assessment.

5. Staff development and training agencies should be urged to recognize that the effective use of computer based assessment is a professional skill that can enhance effectiveness and efficiency and should act to enhance both capacity and capability in item authorship and a broader understanding of item and test retrieval. The assumption underlying this recommendation is that increased expertise will develop the market though both push and pull mechanisms.
6. Governmental agencies, who have the benefit of funding continuity, should build awareness of the importance of compliance (via IMS and localised interoperability testing mechanisms).

Recommendations 7 to 9 concern the development of the QTI specification and associated tools.

7. In the immediate future, further development of the QTI specification should be limited. Development should be confined to areas of demonstrable need and user demand including; aggregation of qualitative and quantitative feedback on assessment items so as to enrich future usage, upgrade of QTI validation tool(s), the correction of errors or ambiguities that are identified by those implementing the specification, and

greater harmonisation with other associated IMS specifications such as ePortfolio, LIP, etc.

8. A high priority should be accorded to the provision of open source assessment related tools for developers and users. Such tools will facilitate their involvement and will lead to the availability of interoperable open source content, a key driver in market creation. There is a specific need to develop authorship tools that will assist (1) localisation of QTI implemented items and (2) the local interpretation or 'calibration' of the results of any assessment task.

9. Consideration should be given by IMS to the construction of a QTI 'Lite' Version 2 which is less complex, and easier to understand and implement. The existence of such a basic standard would facilitate interoperability at lower cost.

The final recommendation is general and is long term in its aspirations.

10. Specification and standard development aimed at supporting interoperability in the design of learning should recognize that assessment is a deeply embedded part of any learning process and that assessment functions should be included in the relevant specifications. Such issues are of particular relevance where assessment cannot be disaggregated easily from learning, e.g. game playing, simulated practice, and eportfolio construction.