DEVELOPING A NATIONAL ITEM BANK

Niall Sclater and Mary MacDonald

Developing a National Item Bank

Niall Sclater
Learning Services
University of Strathclyde
155 George St
Glasgow G1 1RD
www.sclater.com

Mary MacDonald
COLEG
Argyll Court
Castle Business Park
Stirling FK9 4TY
mary.macdonald@coleg.org.uk

Abstract

The COLA project has been developing a large bank of assessment items for units across the Scottish further education curriculum since May 2003. These will be made available to learners mainly via colleges' virtual learning environments. Many people have been involved in the development of the COLA item bank. Processes have included deciding on appropriate item types and subject areas, training authors, peer-reviewing and quality assuring the items and assessments, and ensuring they are interoperable and tagged with appropriate metadata.

Keywords

computer assisted assessment, item banks, interoperability, metadata, quality assurance

Introduction

Various people have attempted to provide a definition of an item bank. These range from a simple:

> collection of text items that may be easily accessed for use in preparing exams (Ward & Murray-Ward, 1994)

to the more detailed but less generic:

> collection of test items that can be readily accessed for use in preparing examinations… normally computerized for ease of item storage and to facilitate the generation of new tests. Each item … is coded according to competency area and instructional objective, as well as empirically derived data such as measures of item difficulty and discrimination (McCallon & Schumacker, 2002)

Some of these definitions incorporate the concept of different but equivalent assessments being produced dynamically and automatically for each learner

from a bank of items. Others imply that the database will be used to store data about the usage of the items by learners. All the definitions suggest that items should be classified by descriptive data (metadata) of some sort to enable them to be located. The decisions taken over what type of metadata to use as well as the content and type of items will differentiate one item bank from another.

While item banks have been around for many years and in a range of contexts, various factors are now coming together which suggest that their use is set to increase considerably. Firstly the software is now available and the hardware ubiquitous enough to deliver assessments to learners either through virtual learning environments or bespoke online assessment systems. Secondly there is an internationally-recognised format for the transfer of items between these systems (IMS, 2002). This format can also be used to store items in a database separately from any proprietary assessment delivery system. Thirdly there are now pressing economic and political imperatives for the development of national and international item banks.

Developing items and assessments across a subject area or sector can bring economies of scale in the development process and a considerable reduction in duplication of effort in different colleges and universities, funded largely by public money. The quality of items which are peer reviewed and validated centrally is likely to be higher than those developed on an ad-hoc basis in an individual institution. An increased adherence to technical standards should mean that the life of items is prolonged and that items are more likely to be deliverable through a variety of assessment systems and virtual learning environments.

There are already some successful examples of item banks under development. These tend to be either

1. *assessment-specific* eg the Cambridge English for Speakers of Other Languages item bank for the University of Cambridge Local Examination Syndicate

or

2. *subject-specific* eg the Electronics and Electrical Engineering Assessment Network (e3an), Helping Engineers Learn Mathematics (HELM) and various initiatives taken by the Learning and Teaching Support Networks which are developing items in economics and computing.

However a third type of item bank is now emerging: sectoral. The COLA (COLEG OnLine Assessment) project is developing a large bank of items across the entire Scottish further education (FE) curriculum. A simple definition of an item bank which incorporates these three types might be:

a collection of items for a particular assessment, subject or educational sector, classified by metadata which facilitates searching or automated test creation

Management of the COLA project

The COLA project was established with funding from the Scottish Further Education Funding Council (SFEFC) which identified online assessment as a strategic priority. Most Scottish colleges had already deployed virtual learning environments (VLEs), providing new opportunities for online learning. Feedback from the sector showed that the lack of a national database of assessment instruments was proving a barrier to widespread use of the VLEs for assessment purposes. In addition the Scottish Qualifications Agency (SQA) had recently produced a set of guidelines for the use of online assessment (SQA, 2003) and was developing a strategy in this area. The Funding Council believed that the use of online assessment could reduce the burden on academic staff and encourage more of them to engage with information and communications technology for learning and teaching.

COLA's aim was to develop a bank of high quality assessment instruments capable of being delivered through the four main VLEs in use in Scottish colleges in a wide range of courses at all levels within further education (FE). The project is managed by the Colleges Open Learning Exchange Group (COLEG), – a partnership of 42 Scottish colleges which undertakes collaborative projects to develop, exchange and promote open, flexible and online learning materials. COLEG manages each project while college staff write, produce and peer review the materials and quality assurance staff check them through a rigorous quality assurance process before dissemination to the sector. COLEG has used the same approach for the COLA project.

A steering group was formed to oversee the project, which includes representation from the various agencies, senior FE managers and FE practitioners with expertise and experience in online assessment, VLEs, interoperability issues and staff development. The project team includes a project manager, an administrator, a technical consultant with expertise in online assessment and interoperability, a technical advisor experienced in online assessment and a staff developer. A technical advisory group was also appointed from college staff with substantial experience in on-line assessment and expertise in VLEs and interoperability issues. This group has strong links with the CETIS Assessment Special Interest Group and IMS, the international body responsible for assessment interoperability specifications.

Selecting areas for assessment

The prime aim of COLA is to provide a bank of assessment instruments to encourage more widespread use of VLEs by college staff across the curriculum. Awareness of the project was raised through local subject networks and staff were encouraged to put forward their suggestions for areas of the curriculum that would be appropriate for online assessment. As a starting point for selecting areas of the curriculum, staff were asked to focus on learning outcomes within SQA units that would be appropriate for objective

testing. In practice subject specialists created assessments to meet the formative assessment requirements of complete outcomes, parts of outcomes (performance criteria) or a combination of topics (performance criteria) from several outcomes.

It was recommended that an assessment should contain a maximum of twenty items in total. There was a general view among academic staff that twenty multiple choice items would normally cover the formative assessment requirements of one SQA outcome.

Choosing item types

The project has concentrated on developing pedagogically sound objective tests, using a limited number of item types. After consulting e3an on the item types they had selected for their item bank in engineering, the types chosen were true/false, multiple choice, multiple response, fill in the blank and matching. There were a number of reasons for selecting a limited range of item types. The wide range and spread of assessments would be limited by specifying a small number of item types, allowing the assessment of a variety of skills and cognitive levels. A focused programme of staff development could be provided for writers. The assessments had to work in a range of VLEs and it was expected that the VLEs would accept these item types if they were marked up using the IMS Question and Test Interoperability (QTI) specification (IMS, 2002).

Development of the templates

To simplify the process of item creation, standard Word templates were developed for the college writers. This approach had already been used successfully by the e3an project and it was expected that staff familiar with using Word would be able to input content to the templates relatively easily. A template was created for each item type. Item templates allow authors to specify the stem of an item, the options and the correct answer, to incorporate graphics in the stem and the options and to provide feedback for each option. They also include a section for additional information such as the expected time to be taken, a description of the item, keywords and the subject topic. In addition an assessment template was developed to contain metadata about the assessment itself and to specify which items were contained in the assessment.

Metadata

As the COLA item bank grows it will become increasingly important to provide an adequate means of identifying items and assessments. The provision of appropriate and accurate metadata makes this possible. There is now a Worldwide standard for learning object metadata (LOM) published by the IEEE (IEEE, 2003) which was chosen as the format in which to store COLA metadata. If a COLA item or assessment is uploaded to a VLE or content

repository, the metadata should be instantly recognised and allow users to search for the material on the metadata fields.

IEEE LOM had never before been used to classify items and assessments. However a group of UK experts has got together to produce an *application profile* (a kind of subset specifying mandatory and optional elements) of the LOM for use within UK further and higher education. This is known as the UK LOM Core. It seemed appropriate to utilise this application profile for COLA in order to maximise the chances of its metadata being understood by other systems. COLA worked with experts in metadata and assessment to produce further application profiles of the UK LOM Core for items and assessments. Work done for the COLA project on metadata and content packaging has fed directly into v2.0 of the IMS Question and Test Interoperability specification.

The COLA templates allow authors to enter items and assessments and also to complete most of the metadata used to classify them. A template conversion tool which was built for the project ensures that metadata fields are transferred accurately and consistently from the templates to the LOM format, while automatically completing some of the more esoteric fields which authors might have found difficult to understand. This is a much better solution than giving authors access to a tool which requires them to understand the LOM format itself. It ensures that metadata across the entire collection of COLA assessments and items has high levels of quality and consistency without creating an excessive burden on authors.

Each item is classified by the Scottish Credit and Qualifications Framework (SCQF) level, a number from 1 to 8. Assessment-level classification metadata is defined in a similar way to that of items. In addition to the level there are entries for the SQA Outcome Number, the Performance Criteria, the Unit Number and the Unit Title.

Identifying, training and supporting assessment writers

COLEG used its standard approach to recruit writers, working through its network of contacts in the colleges to disseminate information about the project to staff and to invite them to commit to the project. The project was launched with an awareness-raising event for college staff – curricular, technical and management – to explain the aims of the project, timescales and funding arrangements and to listen to their views on implementation.

Following the event, colleges were asked to confirm the services that they could provide to the project. Standard levels of payment were set for writing and inputting of the assessments into the templates, for peer reviewing, for quality assurance and for project management. Writers confirmed on a proforma the curriculum areas/topics and peer reviewers of their assessments. At the same time technical staff with relevant experience of the different VLEs were invited to join the technical advisory group for the project and to advise the steering group on technical issues.

Thereafter a series of two workshops was organised for writers and peer reviewers. The workshops provided information about the project and clarified its focus on objective tests. The various item and assessment templates and the item types chosen were explained. A set of guidelines was created, including a writer's/peer reviewer's quality checklist for each item type and for the assessment information and a guide to completion of the templates. Evaluation forms showed that the workshops were well-received by participants. Technical and pedagogical quality of the items is likely to be higher than if they not been carried out. Certainly there would have been confusion about the use of the templates. The writers also confirmed that the workshops helped their understanding of the pedagogy of objective tests not just their understanding of the templates.

A timescale of six weeks between May and June 2003 was set between the first writers' workshop and the deadline for submission of the assessments by the writers. Following the workshops, one to one guidance on pedagogy related issues was available. Email and telephone support was also available for both pedagogical and technical issues. A further series of one day workshops was held for a second phase of development work between July and September 2003. In total 66 writers delivered 165 assessments (approximately 3000 items) in the first two phases of the project. Only three writers withdrew from the project.

Quality assurance

COLEG implemented its standard quality assurance procedures in the project, including checking the quality of the items (from the subject specialist's and the learner's perspective), checking the quality of the production (grammar, typos) and checking the technical aspects (eg completion of template fields, use of standard file names).

First, the writers were required to check their assessments and the collection of items within the assessment against the writer's/peer reviewer's checklist. Before submission to COLEG, the assessments were peer reviewed by other subject specialists against the same checklist. It was proposed that a subject specialist from a different college would carry out the peer review. This was not always practical because of the short timescale, the availability of subject specialists in relevant curricular areas and the practicalities of peer reviewing at a distance from the writer. In many cases another subject specialist from the same college as the writer carried out the peer review. Completion of a checklist for each item and submission of these to COLEG would also have become an administrative burden, so a "summary checklist" was created, which asked the writer to confirm that the assessment had been created in accordance with the checklist and peer reviewed against the checklist.

Once the assessments were submitted to COLEG, a quality check was carried out on each assessment to check the quality of the assessment from the learner's perspective, the quality of production and the technical aspects.

Recruitment of quality assurance staff for this stage of the process was difficult. There was a very limited response from colleges to the invitation to provide this service. Instead a small team of three experienced staff undertook this demanding task. To support this stage, quality assurance checklists were drawn up for each item type and for the assessment information template. The staff were briefed on the three elements of the process – pedagogy, production and technical issues.

In checking the quality of the items from the learner's perspective the quality assurance staff identified several key issues:

- what is to be assessed?
- why has a particular item type been selected?
- is the item or instruction (stem) clear?
- is contextualisation necessary and appropriate?

75% of the items created were considered to be of good quality. After further development work, it has been possible, with the exception of five assessments, to validate all the assessments in the first development phase. Where there were questions over the quality of the assessments, the robustness of the peer review process was questioned, particularly where the wording of an item was inappropriate or the item type used was not suitable.

It was felt that the quality of the feedback to the learner was important. The online context and the VLE technology sometimes limited the feedback that could be provided, however. For some item types it was stated in the guidance that only standard (No, this is incorrect or Yes, this is correct) feedback could be provided because it would be impossible to predict the learner's responses to the items. In practice some writers proved to be extremely creative with the additional general feedback that they provided.

It was not possible to clarify some of the technical issues related to the templates at the time of the workshops. In addition further issues were identified at the later stage of testing of the exemplar assessments. In both cases these were addressed at the quality assurance stage.

In the main it was felt that the writers had made a reasonable attempt to complete the fields in the template. The general view was that it was important for the writers to gain skills in data input and that this would give them a better understanding of how the VLEs would handle the assessments. It was also established that it would be possible to standardise more of the content in the template such as feedback. This would reduce the potential for error.

The quality assurance staff were asked to "fix" minor discrepancies in the assessments and to "fix" the issues created by the technology. This allowed many of the assessments to be "validated" without having to be returned to the writers. In other cases more significant amendments were required but the writers were asked only to confirm that the changes made were acceptable to

them. In some cases writers were asked to provide further content or to change the content or item types that had been used. In the main the process has worked well though it has been time consuming. In one or two cases writers have wanted to provide their own amendments.

Version control and file management has been an important issue during the quality assurance process. A file management system has been developed for the project that classifies the assessments into three categories:

- initial version: received from the writer following peer review
- part-validated version: quality assured but checking or amendment required by the writer
- validated version: approved by the quality assurance staff

A spreadsheet has been developed to record details of writer, peer reviewer, subject area and level, quality assurance process and administrative details. Overall this system has worked well, though management and maintenance of the files has been time-consuming and requires a great deal of care and attention to detail. In a small number of cases writers changed items that had been validated and these needed to be rechecked. Wherever practical, writers have been asked to notify the quality assurance staff of the amendments that they want to make rather than changing the templates themselves.

A protocol for referencing and saving files, based on the writer's name, was standardised in the templates. This provided a simple referencing system but it caused two problems. Items are numbered consecutively in an assessment to provide a unique identifier for each item. If a writer creates more than one assessment, then the numbering of items in the second and third assessment follows consecutively. If changes are made to the number of items in one assessment, changes to the numbering of items in subsequent assessments are required. Item files are saved according to this same protocol, so the files also need to be saved under the new number.

Transfer to QTI and VLE Formats

One of the primary aims of COLA was to encourage colleges to use their VLEs by providing online assessments which could be run from the VLEs. In order to do this, the assessments had to be in a format which the VLEs would understand. The only international specification (not yet a standard) for the exchange of items and assessments is the IMS Question and Test Interoperability (QTI) specification. Many vendors pay lip service to their products' compliance with this specification but do not properly implement it. The four VLEs in use in Scottish colleges all claim some level of compliance with the specification. COLA took the decision to store all content in this platform independent format which is undoubtedly increasing in uptake Worldwide.

It was necessary to develop a program to convert the items and assessments from the Word templates to the QTI format. This task was carried out by the JISC-funded Technologies for Online Interoperable Assessment (TOIA) project which had the necessary expertise in QTI in collaboration with an expert group representing the four main VLEs. There were many complications due to the different ways in which the VLEs interpreted the QTI specification and their limited implementations of some of the item types. Using a third party product called Respondus which accepts QTI it is now possible to transfer COLA content into WebCT and Blackboard. Teknical now accepts COLA content directly and it is still hoped that a solution can be found to put COLA items into Granada Learnwise.

Having produced the items and assessments to the correct standard, they can also be uploaded and stored in some of the emerging learning content repositories with ease, allowing items to be searched for on their metadata. In order for teachers to be able to search for items with ease the conversion tool creates two indexes which can be read using Microsoft Excel – one for items, the other for assessments. On each line of the spreadsheet is one item/assessment and all the metadata associated with that item such as author name and SQA Unit Number.

Distribution

Much discussion took place in the technical advisory group meetings as to how to distribute the items and assessments. While the distribution of CD-ROMs would have provided a further opportunity to disseminate the project to colleges, the technology is a backward one and it was considered to be simpler for colleges to download the latest versions of the item bank and install them in their VLEs directly from a central website. The COLEG named contact in each college would be authenticated to do so.

Separate indexes of all items and assessments will be provided on the website, both searchable on any item of the metadata. Staff will then be able to download the items and assessments required in IMS QTI format so that they can import them into their VLE.

Conclusions

The templates were developed in Word for ease of use and overall writers have coped reasonably well with them. The development process has however highlighted a number of issues. There are limitations in the type of data input that the template will allow. It would be possible to standardise the feedback in some cases, thus reducing the potential for error. The filing protocol is cumbersome and a more simple referencing arrangement should be devised. A web-based development system, while requiring authors to be online, would remove the problems encountered with authors misnaming and misplacing the various item and assessment templates and graphic files.

The workshops were well received by writers and the same format will be used in the future. The guidance and checklists will be reviewed and improved in light of feedback during the quality assurance process. This process has clarified where writers are likely to make errors. Also, it will be possible to demonstrate real examples of the different item types and the creativity of writers using the items generated in the first phase of the project.

The quality assurance process itself has worked well and reduced the burden on writers. The process has been resource intensive however and there has been a limited pool of staff to undertake the work. The process has also highlighted the importance of recruiting experienced quality assurance staff. Care and attention to detail is crucial. Improvements to the templates and to guidance and staff development for writers and peer reviewers should reduce the quality assurance work required in the future.

The file management systems and version control again have worked well but are resource intensive. Changes to the file protocol for the template should reduce the administrative burden.

The development of a conversion tool which produces items which will render correctly in a range of VLEs has been a difficult process because the VLEs do not interpret the QTI specification in the same way. In addition there has been no commonly agreed format for item metadata. However the TOIA-COLA Assessment Metadata Application profile provided a mechanism for storing the metadata required by the COLA project in a standards-compliant format readable by a number of existing repositories.

For the development of future assessments, COLA is considering a web-based system which would remove some of the logistical problems currently being faced such as authors having to give items and assessments the correct name and having to zip up assessments and email them for validation and central collation. It would help considerably with version control of the templates, the conversion tool and the content itself and remove the dependence of the conversion tool on a particular version of Word.

A web-based system could act as a repository for the assessments themselves and a place where college staff could try out the assessments before loading them into a VLE. It would allow instant viewing of the items and assessments by the authors themselves and provide much better control over the various administrative processes. It could also allow automated notification to validators by email when an assessment is ready to be validated.

The next stage of the project will be to evaluate how easy colleges find it to import the COLA items to their VLEs and to analyse the uptake of the materials by staff and students nationally.

References

CETIS Assessment Special Interest Group, assessment.cetis.ac.uk

COLEG, www.coleg.org.uk

e3an project, www.e3an.ac.uk

HELM project, www.lboro.ac.uk/research/helm/

IEEE (2002) *Standard for Learning Object Metadata*, IEEE 1484, 12.1-2002

IMS (2002) *Question and Test Interoperability Specification v1.2*, IMS Global Learning Consortium, www.imsglobal.org

McCallon, E.L. & Schumacker, R.E. (2002) *Developing and Maintaining an Item Bank*, ELM Metrics Inc, www.elmmetrics.com

McAlpine, M (2002) *Design Requirements of a Databank*, CAA Centre, www.caacentre.ac.uk

Sclater, N (2003) *TOIA-COLA Metadata Application Profile v1.2*, www.cetis.ac.uk/profiles/uklomcore/

SQA (2003) *Guidelines on Online Assessment for Further Education*, www.sqa.org.uk

Ward, A.W & Murray-Ward, M (1994) Guidelines for the Development of Item Banks, *Educational Measurement: Issues and Practice*, National Council on Measurement in Education