

Designing Assessment Tools in a Service Oriented Architecture

Will M Davies and Hugh C Davis

Learning Technologies Group, ECS, University of Southampton, Southampton, UK
{wmd04r|hcd}@ecs.soton.ac.uk

Abstract

Assessment is an important component of formal learning, and Computer Assisted Assessment (CAA) is a well established component of most online learning. However, technical issues such as interoperability and security, and pedagogic reservations as to its effectiveness still remain barriers to the uptake of CAA. In this paper we examine a number of current assessment projects, predominantly emanating from the UK, to consider how a service oriented architecture can facilitate the implementation of tailored assessment environments, providing improved assessments within an interoperable and secure framework.

1. Introduction

The ELeGI [8] project is concerned with building an infrastructure to support all forms of learning, both formal and informal. Assessment is concerned with measuring progress in learning, and may be used by institutions, such as Schools, Universities and Professional Bodies, to inform themselves of the progress of a learner (we call this summative assessment), or assessment may be used by the learners themselves in order to measure and confirm their own progress (we call this formative assessment). Either way, the assessment is generally of concern to the formal learning domain, where a learner has clearly defined learning targets against which they are being measured or measuring themselves.

Because of the importance of assessment within formal learning environments, and the difficulty that teachers experience in finding time to mark and give feedback on all of a learner's work, Computer Aided Assessment has been an active research area within learning technology for many years, dating back to the use of punch cards. More recently web based educational systems have provided the channel to allow easy access, any time, any where to Computer Aided Assessment (CAA) servers.

Such CAA systems have generally been stand-alone, often bespoke developed and lock content inside proprietary systems. Until recently institutions had a choice between the one size fits all, monolithic commercial offerings, developing their own or customizing someone else's bespoke application. This approach has constrained the ability for new systems to leverage developments in previous ones.

Probably the most significant commercial systems in general use are Question Mark Perception [18] and the assessment tools within learning management systems such as WebCT™ and Blackboard™; these are capable of delivering summative assessments to large classes but are constrained in both their functionality and interoperability. Whilst the capability of these systems has increased over time, for commercial reasons they tend to follow rather than lead development in CAA.

The educational establishments in which CAAs are deployed are likely to contain a large number of innovative thinkers, it tends to be these people that drive development to better serve their institution's educational needs. It is likely that the nature of these, technology early adopters, has led to them developing custom systems as they have found this more productive than requesting features from commercial suppliers. Because of the diverse requirements of systems and limited funding, these developments have usually resulted in systems that have poor interoperability and do not fit the requirements of a broad enough user base to ensure widespread uptake and continued development.

One criticism of CAA is that it is difficult to assess higher order skills (evaluative skills, design skills, synthesis skills etc.) using objectives tests, such as multiple choice and true/false questions. This issue has been addressed by Duke-Williams and King [6] who have successfully used objective tests to achieve this goal, though they do stress the high level of care that must be taken when creating such tests.

Two alternative approaches to assessing higher order skills are free text marking systems and TRIADS. TRIADS [15] is an assessment creation and delivery system developed at the University of Derby, UK, which allows the author to design

complex scenarios, simulations and activities for learners, and to intelligently sequence users through sets of questions, allowing for a very wide and flexible range of question types. With the more recent development of the IMS Global Learning Consortium's Question and Test Interoperability (QTI) [9] specification, this flexibility has become a trade-off against the widespread interoperability that QTI is expected to bring.

When it comes to marking free text and essays, systems such as Automark [16] and E-Rater [17] are making good progress at marking both writing style and content. Such systems have been demonstrated to be just as reliable as human markers, but have yet to gain acceptance within the educational community.

As the mechanisms for building loosely coupled systems from components become more sophisticated the ability to assemble a complete assessment system from services and integrate this inside a Virtual Learning Environment (VLE) is becoming possible. A Service Orientated Architecture (SOA) will facilitate the rapid development of highly customizable systems that can be optimized towards a specific goal or pedagogical requirement. This framework will also make it easy to plug in extra components or combine services in novel ways to evaluate their effectiveness.

2. Standards and Interoperability

The development of questions and exercises suitable for computer aided assessment is a time consuming business and teachers who have invested time in this endeavour have often been frustrated to find that questions prepared for one system cannot be transferred to another. For this reason the IMS consortium have directed much effort towards the production of an interoperable specification for questions and tests.

QTI describes a data model to represent questions, responses, marking, results and aggregation of assessment items to form tests, and thus allows the sharing of objective tests between different organizations and software environments. The specification is implemented in an XML schema that allows for the exchange of items and tests between heterogeneous assessment systems.

Application developers are becoming increasingly aware that they cannot rely on a homogeneous environment in which to deploy their application. In a SOA this concept is taken one step further by combining a number of heterogeneous subsystems. In this environment the protocols used to

communicate and the interfaces between the systems become increasingly important and must be clearly defined. Clearly any SOA that handles objective assessments must have a common method of describing these assessments.

The main benefits of using QTI are interoperability, integration, banking and, potentially, wide pedagogical support. The ability to include QTI described questions seamlessly alongside other learning resources inside a single integrated environment can avoid the need for students to log in and out of disparate systems and allow formative assessment to be presented in conjunction with other learning resources. The use of QTI in a SOA increases the options for embedding assessments alongside other learning components. Using searchable test banks to store and share questions simplifies the re-use of assessments and tests, though this still requires a degree of cooperation on how the question metadata should be used.

There are potential problems associated with insisting on adherence to QTI; doing this may cause situations where the technology, rather than the pedagogical requirement, determines the way in which students are assessed. The QTI specification is still relatively young, even though version two has undergone an extensive public draft consultation phase it would be surprising if there were not still some inconsistencies in it. The QTI specification is large; to be QTI *compliant* a system must be able to interpret and render correctly a number of complex question types that may be unnecessary for the environment in which the player is used. The problem is mitigated inside a SOA where an individual service will be able to state its level of *conformance* to the specification.

Adherence of assessment systems to the QTI specification is so far limited. Commercial systems such as Question Mark and Blackboard have made attempts to track the evolution of the specification, at least as an import/export option. TOIA [23] is a complete assessment creation and delivery system, and is significant because it uses QTI as its native format. TOIA, which is free to use within the UK educational system, was written as a closed interface system but development is taking place to expose it as a set of web services [24].

3. Elements of CAA

Ease of question authoring is a significant proportion of the work in being able to deliver CAA. By authoring, we mean both the assembling of an individual question, possible responses and allocated marks to form an item and the aggregation of these items into the assessment to be delivered to a student. Authors of objective tests must not only ensure the soundness of their tests they must also get the assessment into the assessment system. It is clearly in the best interests of assessment software to make this process as painless as possible. However, it may be found that once inside the assessment system even standard QTI question items may not be exported unaltered to other systems. One cause of this is ambiguities in the QTI version one specification; with version two much care has been taken to minimise this problem. At the current time, there are no QTI version two authoring tools available for evaluation; this is unsurprising, as the specification was not finalised until January this year. Stand-alone tools for authoring QTI question items include Canvas Learning Author [4] and xDLSofts QTI Ready Designer [26]. Received wisdom suggests that by using one of these tools to author assessments before uploading to the delivery system vendor lock-in may be avoided. Because of the richness of QTI, authoring tools can become overly complicated by giving users more options than they are likely to need. An associated problem is encountered when the authoring tool does not allow the richness of logic required by the user, in this situation they may need to modify the XML source by hand. These problems could be reduced by providing users with a choice of authoring services, possibly dynamically switching between them as needed.

To illustrate the potential complexity of authoring objective tests, listed below are most of the question types that QTI can describe; these can be combined with the ability to deliver hints, reduced marks and template based questions with random dynamically generated numeric values. QTI can be highly expressive.

- multiple choice - Choose the one correct response.
- multiple response - Choose all the correct responses.
- true/false, yes/no - A simple binary choice.
- image hot spot - Identify the required area(s) on an image (by clicking).
- fill in the blank - Insert the missing words
- text short answer - Free form text field.

- essay text - Long response text field, likely to require human marking.
- numeric entry - Enter the correct number.
- slider - Move slide bar pointer to correct value.
- drag and drop - Place objects into the correct locations.
- order objects - Rank objects according to the given criteria.
- match item - Connect the objects in pairs.
- connect the points - Create an ordered connection of a set of points.
- Macromedia Flash Object - QTI player runs the Flash program it is given. The QTI player has a very limited interface to the Flash program, though there is no limit on the complexity of the logic that can be expressed inside the Flash object.

4. Assessment Lifecycle

Storage of assessment and question items is commonly referred to as *item banking*. The IBIS report [20] describes in detail the functionality required of an item bank; it was informed by the study of a number of existing item banks including e3an [7] and Scottish Colleges Open Learning Exchange Group's COLA project. COLA is now a highly successful live system providing assessments across the curriculum to Further Education colleges across Scotland, where they are presented inside several different VLEs. Interestingly both e3an and COLA acquired their content by asking authors to fill out MS Word templates to describe their questions; these were then parsed automatically for storage.

The requirements of an item bank include facilities to store, search for and retrieve individual items. Extended functions include the ability to assemble items into assessments and to be able to deliver alternative but equivalent items. The IBIS report recognizes the need to ensure that the items in the bank have been properly peer reviewed as part of a quality assurance process, it also recommends that the usage data concerning an item be fed back into the system. With feedback on usage the effectiveness of individual question items and possible bias can be monitored, this may result in items being removed from further use. As with any data repository use of item banks arouses questions concerning user identity, security and availability.

Scater and Howie [21] describe the ultimate online assessment engine; the paper describes 21 user roles in an assessment system and gives detailed use cases and user requirements of the system. In 2001

they evaluated two commercial products against their requirements and found nearly all of their requirements satisfied. From the technical perspective most of the challenges involved in delivering CAA have been met. Candidates can be authenticated and remote presentation, response gathering and feedback can be delivered using web browsers. Delivery of CAA is now a fairly well understood area but is often not completely integrated into students' other educational experiences. Augmenting the assessment delivery system with open interfaces would allow much finer granularity of control in integrating this with the rest of the virtual learning environment.

As well as showing the user roles Figure 1 illustrates the sets of data that must be coordinated to ensure that the assessment runs correctly, and demonstrates the point that there is much more to the assessment task than the assessment delivery engine alone.

There are a number of tasks that must be carried out before an assessment can be delivered, including authoring of questions, quality control of questions, selection of suitable questions to create a test, and

selection of delivery conditions, such as the group to take the assessment, on what platform and when? Once a student has finished an assessment there are still a number of tasks that the system has to perform relating to the student and to the assessment as a whole. Marks may need to be stored permanently in a grade book, feedback may need to be given to the student, and the results may need some form of moderation. As described already feedback from assessments can also be fed back into the item bank they were obtained from to augment quality assurance procedures. Analysis of assessment marks across a class may be used both to analyse the effectiveness of the assessment and to identify problem areas for the entire class with the subject material. Analysis of individual's marks across a number of assessments can identify anomalies that may indicate students with specific difficulties and also situations where some form of cheating may have occurred.

The work on defining the Ultimate Assessment Engine is now forming the basis of ongoing work within the E-Learning Framework [25] (described in the next section), to build a reference model for assessment services.

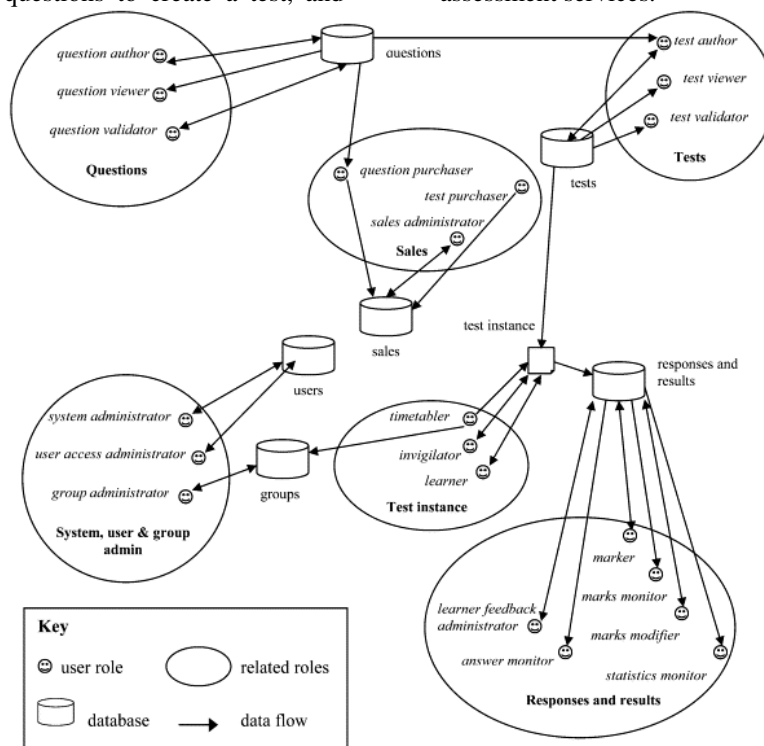


Figure 1: The Ultimate Assessment Engine (from Sclater and Howie 2003).

5. Reasons for Using Services

A SOA is becoming recognized as a highly flexible way of building a large application from components [2],[3]. From an institutional point of view this enables collaboration between universities, faster deployment of new functionality, and support for pedagogic diversity, and avoids lock in to single vendor solutions with the possible attendant costs.

From a technical point of view the open interfaces of the components make it relatively simple to connect components in novel and custom ways, encourage interoperability, and facilitate replacing one service with another to provide the same functionality in different ways. Here we describe three different learning environment projects and comment on their methodology.

The E-Learning Framework (ELF) [25] is an initiative by the U.K's Joint Information Systems Committee (JISC), Australia's Department of Education, Science and Training (DEST), and the Carnegie Mellon Learning Services Architecture Lab (LSAL). The ELF does not set out to build a learning management system but a framework which is a road map of functions that could be used in planning institutional e-learning systems. So far over 40 separate component functions have been identified which might be needed inside a comprehensive MLE. This approach allows production of architectures based on standards which enable interoperability, providing a common understanding for future developments. It also allows the community to more easily identify gaps or major barriers to progress on which to focus funding for development activity

Discrete packages of funding have been allocated to investigate the requirements of individual components and also to build exemplar services to satisfy these requirements. By steering the project in this way it is hoped that a critical mass will be achieved where institutions can spontaneously develop component services to augment an existing working system.

The service oriented approach was adopted in the ELF because it separates out the contract between the providers and consumers from the application itself. This approach is also neutral in terms of platform and language. The upshot of both of these things is that ELF can fit with commercial systems such as Blackboard and Web CT, and open source systems such as Moodle, regardless of the technology they're built on. The challenge now for ELF is the design of the "fabric" services of workflow, security, and management. The question is whether to build

services that are "fat" and complex or "thin" and simple.

The work on ELF has a large overlap with the Grid based work of the e-Science community, particularly in the area of Virtual Research Environments. At a recent ELF Conference the two communities produced a useful report [14] on the requirements overlap.

The Sakai Project [19] is being run by a consortium of universities, in the USA, to create an open source, extensible VLE. When asked to explain some of the benefits of the modular architecture of Sakai against a commercial VLE, Sakai Project Chief Architect, Chuck Severance said: [12]

"What we do at universities is teaching, learning and research. We can't outsource the software that supports that- you can't outsource your destiny. You shouldn't have to negotiate with an outside commercial provider about things that directly affect your core business. So, building your own MLE software allows an organization to take charge of its own destiny. But building a completely unique package for your own use is really a bit lonely and somewhat expensive. By working together in Sakai, we can control our own destinies and avoid the cost and risk involved in the solo path. Major research institutions can build the large components and the framework, smaller ones can customize the tools they need."

Whilst Sakai is not currently using a service based paradigm, its designers are aware of the value of service architecture and recognize that Sakai may have service interfaces inserted in the future. The pressure to deliver a complete running system led to the decision to use more mature technology based on Open Knowledge Initiative's Open Service Interface Definitions and extended Application Programming Interfaces..

The Department of Education Tasmania provides education for the 70,000 pupils in schools and colleges across Tasmania, Australia. They have undertaken a pioneering case study, LeAP [13], examining the building of a SOA Managed Learning Environment. LeAP was started in 2002 and was the first large project of its kind. The department describes itself as an early adopter of technology. They installed WebCT in 1999, but are clearly concerned about the implications of being locked into a proprietary system. The reports commissioned to start the ELF were written in collaboration with some of the key figures in LeAP.

6. Current Work

Here we describe a number of state of the art bodies of work which all have a significant contribution to make towards a SOA for CAA.

The European Learning Grid Infrastructure (ELeGI) describes one of its goals as to define and implement an advanced service-oriented Grid based software architecture for learning. In existing SOAs based on protocols such as web services, the issues of security, identity, access management and managing transaction persistence are not managed at the protocol level. Having to address these issues whenever a new service is created is a cumbersome overhead. By deploying its services over grid middleware ELeGI will remove much of the burden on individual service developers in addressing these problems. Grid middleware will be a valuable tool for the VLE service author.

Remote Query Protocol (RQP) [22] is being developed by a JISC funded project, Serving Maths, to create a protocol to support remote rendering and processing of question items in an SOA. This is a problem that has been a source of constant concern to those assessing the mathematically founded disciplines; very few available engines have the required sophistication to render mathematics. RQP is aware of the fragility of a distributed architecture and includes mechanisms for fail-over and load balancing by the project. Demonstration systems using RQP to deliver mathematics questions have been able to use a choice of question renderers for an assessment depending on the demands of the question. This is an excellent example of how the SOA puts power in the hands of the assessment authors.

ASSIS [1] is funded by JISC and is currently one of the partners being used to test RQP. The aim of ASSIS is to connect up an item bank with a QTI player and a service for running IMS Simple Sequencing [10] (a method of dynamically changing a user's path through learning material). QTIRun, the QTI player service used by ASSIS, is an extension of APIS the first publicly available system to play questions written in the latest QTI specification, version two. Item banking services will be provided both by an open interface to TOIA and also by an interface to Samigo, the Sakai Project assessment system. ASSIS is using CAA to give formative feedback to students as they work their way through learning packages. These packages are augmented with Simple Sequencing rules that allow authors to script a student's path through the learning material and adapt to different students' educational needs.

By combining different and also equivalent services ASSIS is providing a small-scale fully functional test-bed of a SOA.

JORUM+ [11] provides a repository of educational content for UK further and higher education. The main functions of JORUM are to offer institutions a secure, resilient location in which to store their educational content and also to facilitate sharing of content between institutions without the need for them to negotiate through each others network security policies. JORUM is currently in an extended testing phase and is being used by a number of early adopters. Importantly JORUM is addressing the legal considerations, including intellectual property rights, connected to the content it holds. Resources can be submitted to JORUM in a variety of formats including references to material held outside JORUM. To facilitate searching, resources are tagged with Learning Object Metadata. When searching for material users can search by keyword and also by using a graphical classification structure. To enable end users to select appropriate resources for use in their institution, all resources within JORUM can be previewed. The lessons learnt by JORUM will be useful to any future question item banking service.

D+ [5] is a service for conducting federated searches; it allows users to search for information hidden from simple searching mechanisms. D+ provides a service for searching the deep web hidden under database style indexing systems and other dynamic content generators. By being aware of a number of search protocols such as the Z39.50 protocol for searching libraries, D+ can present users with search results compiled from a number of sources. Federated search is important in the context of SOA CAA as it could be used to combine searches over a number of item banks that provide differing interfaces to their content.

Security in eLearning, and particularly assessment, is an important issue from a number of points of view, often summed up as "the three As"; authentication, authorization and administration. We must know that the student taking an assessment is authorized to take this test and has been authenticated as the person they claim to be, and when they have completed the test we must understand how to route the results. These functions are currently causing enormous problems within the world of global campuses working with heterogeneous software systems, and it is the promise of robust proven grid middleware that is one of the major attractions of this work.

7. Cooking Up an Assessment System

While developing all the services described, it is important not to forget that on their own they do not deliver an assessment system. To encourage uptake of the services discussed it will be necessary to develop exemplar or proof of concept consumers that demonstrate how the services may be integrated. Once these are in place then the research community will be able to experiment with innovative ways of using the services from inside a VLE. It is envisaged that the ability to modify the environment easily and leverage one service to develop another, will encourage the development of compound service aggregation. This then exposes a whole new issue of how to manage performance in the face of large quantities of service calls to services whose implementation is hidden. The ability to plug in modular services will also simplify comparison of the effectiveness of parallel services that provide subtly different environments.

We look forward to the day when the services for CAA are sufficiently sophisticated that a learner, working on their own, would be able to select appropriate questions to help them assess their own progress in their learning, and to get sensible feedback and help. This would be truly personalized assessment.

8. Acknowledgements

This work has been supported by the European Commission FP6-IST IP 002205 "European Learning GRID Infrastructure" (ELeGI), and also by JISC Exchange for Learning Strand B Project, "Technologies for Online Interoperable Assessment" (TOIA).

9. References

- [1] ASSIS; Assessment and Simple Sequencing Integration Services); <http://www.hull.ac.uk/esig/assis.html>
- [2] Blinco, K., Curtis, G., McLean, N. and Mason, J "LeAP Project Case Study: Implementing Web Services in an Education Environment," 2004, available at: <http://www.education.tas.gov.au/admin/ict/projects/imsdoe/casestudy/LeAPProjectCase.html>
- [3] Blinco, K., Mason, J., McLean, N. and Wilson, S. "Trends and Issues in E-learning Infrastructure Development," presented at: Altilab04, Redwood City, California, USA, 2004, available at: http://www.jisc.ac.uk/uploaded_documents/Altilab04-infrastructureV2.pdf
- [4] Canvas Learning; <http://www.canvaslearning.com/>
- [5] Discovery+; Brokerage for Deep and Distributed e-Learning Resources Discovery; <http://devil.lib.ed.ac.uk:8080/dplus/>.
- [6] Duke-Williams, E., and King, T.. "Using Computer-aided Assessment to test higher level learning outcomes," presented at 5th International CAA Conference, Loughborough University, Loughborough, 2001.
- [7] e3an; Electrical and Electronic Engineering Assessment Network; <http://www.e3an.ac.uk/>
- [8] ELeGI; European Learning Grid Infrastructure. EU IST Framework 6. <http://www.elegi.org/>
- [9] IMS Global Learning Consortium Question & Test Interoperability Specification (QTI); <http://www.imsglobal.org/question/>
- [10] IMS Global Learning Consortium Simple Sequencing Specification; <http://www.imsglobal.org/simplesequencing/>
- [11] JORUM; The JISC Online Repository for Learning and Teaching Materials. <http://www.jorum.ac.uk/>
- [12] Kraan, W. "Same area, different goals; Sakai and the JISC Framework Programme," [Online article] (2004, May 03) available at: <http://www.cetis.ac.uk/content2/20040503155445>
- [13] LEAP; Learning and Evaluation Action Program <http://www.bellanet.org/leap/index.cfm>
- [14] Liber, O. (Ed) A summary of the Personal Learning and Research Environments Domain 11-04 <http://www.elearning.ac.uk/resources/Personal%20Learnin g.ppt>
- [15] Mackenzie, D. Recent Developments in the Tripartite Interactive Assessment Delivery System (TRIADS). Updated from Proceedings of the Third International Computer Assisted Assessment Conference, University of Loughborough, Loughborough, UK Available at: <http://www.derby.ac.uk/ciad/lough99pr.html>
- [16] Mitchell, T, Russell, T, Broomhead, P & Aldridge, N. Towards Robust Computerised Marking of Free-Text Responses. International CAA Conference, Loughborough 2002
- [17] Powers, D. E., Burstein, J., Chodorow, M., Fowles, M. E., & Kukich, K.. Stumping e-rater:® Challenging the

validity of automated essay scoring, (GRE No. 98-08bP, ETS RR-01-03). Princeton, NJ: Educational Testing Service. ETS, 2001 available at:
http://www.ets.org/research/dload/powers_0103.pdf

[18] Questionmark Assessment Software for Web Based Training, Distance Learning, Needs Assessment, CBT, Corporate Training, Online Quiz, and Placement Test materials. <http://www.questionmark.com/>

[19] The Sakai Project; <http://www.sakaiproject.org/>

[20] Sclater, N. "Item Banks Infrastructure Study," JISC, 2004, available at: <http://www.toia.ac.uk/ibis/>.

[21] Sclater, N. and K. Howie, K. "User requirements of the "ultimate" online assessment engine," Computers & Education, Vol. 40, Issue 3, p. 285-306, 2003.

[22] Smith, A. "Remote Question Protocol Serving Items in a Distributed Learning Environment.", 2004, available at: http://maths.york.ac.uk/serving_maths/file.php/2/moddata/forum/15/679/rqp.pdf

[23] TOIA: Technologies for Online Interoperable Assessment. <http://www.toia.ac.uk/>.

[24] TIP; Tools Integration Project;
<http://www.jisc.ac.uk/index.cfm?name=delettip>

[25] Wilson, S., Olivier, B., Jeyes, S., Powell, A. and Franklin, T. "A Technical Framework to Support e-Learning." JISC, 2004, available at:
http://www.jisc.ac.uk/uploaded_documents/Technical%20Framework%20feb04.doc

[26] xDLSoft; <http://www.xdlsoft.com/>