

Web-enabled evaluation of e-Learning

Hassan Ahmed¹, Lampros K Stergioulas² and Costas S Xydeas³

Abstract $\frac{3}{4}$ The UNIVERSAL Exchange for Pan-European Higher Education aims to provide an open exchange/brokerage service for e-learning resources between institutions of higher education across Europe and elsewhere in the world. In order to ensure that the brokerage system offers Learning Resources of the highest quality, a characteristic that is crucial to its success, the system collects from users quality-related information, which can then be processed and displayed to users as need arises. This introduces a market-driven selection mechanism on learning resources and ensures that learning resource quality is continuously improved. This paper addresses the issues involved in the development of a generic “on-line” evaluation engine for “live”, “packaged”, as well as hybrid learning resources, and the use of this engine in providing useful, up-to-date information on available learning resources to users. Key to the development of such evaluation modality is the critical design trade-off between the contradictive requirements of complexity/completeness and the need for simplification in order to improve user satisfaction and user-friendliness. Results from real-life testing and user feedback are also reported and analyzed. The design/development process embraces a formative as well as a summative evaluation approach.

Index Terms $\frac{3}{4}$ Web-based learning resource assessment, Learning Resource Evaluation, Pedagogical Assessment.

1. INTRODUCTION

In recent years considerable emphasis has been placed on application of new technologies in the learning and teaching process. This led to the introduction of web-based learning within Higher Education curricula, which in turn generated significant interest in the effective evaluation of Learning Resources (LR). A number of web based tools and applications that tackle the task of evaluating web-based LR in specific areas have been presented in the literature [1][2][3]. This paper addresses the issues involved in the development of a generic “on-line” evaluation engine for ‘live’, ‘packaged’ as well as ‘hybrid’ eLearning resources, and the use of this engine to provide the user with useful, up-to-date information on available LRs.

The proposed engine aims to provide an online adaptive, centralized service for the evaluation of LRs in Higher Education for the benefit of LR consumers including

students, faculty staff and university administrators, as well as LR providers (e.g. individual academics, faculties, Higher Education Institutes etc.). The evaluation data is collected from users, and is then stored in the evaluation database. This data can be analyzed by the engine and transformed into meaningful LR evaluation information. The engine can then respond to requests from prospective LR users/consumers and LR providers with evaluation information on specific LRs, and by doing so users can judge the quality of LRs.

The main role of the eLearning resource evaluation engine (EE) within our LR exchange or LR delivery framework is to provide LR quality feedback to consumers and providers and by doing so, maintain a high quality LR catalogue/brokerage service. The use of this engine can also help in analysing the effectiveness of a particular teaching methodology, understanding of student learning process and producing a benchmark of teaching quality. Thus it was necessary to resolve strategic and tactical issues concerning the type of evaluation data to be collected, the amount of this data, the type of users providing this data, the time when evaluation should take place and the ways of analysis and presentation.

Since e-learning evaluation is a relatively new research area which is currently in its infancy, there exist no clear defining research and models. Most of the work is application-dependent and a limited number of evaluation methods exist, such as pre-evaluation, post evaluation, interviews, open/close ended questionnaires and likert scale etc[4][5]. In the proposed system the evaluation data has been collected by employing a range of intelligent quantitative questionnaires, open ended questionnaires, user log analysis, which provide data related to general acceptance, and usage of LR and learner assessment (examination) data, if relevant. These questionnaires cover LR content, presentation, interactivity, delivery and support and user satisfaction.

An intelligent questionnaire structure is proposed, which is responsive to variable evaluation needs, in the sense that second-level (detailed) feedback collection is invoked specifically for those attributes which are perceived by users as problematic. This type of flexibility provides a good compromise between the contradictive requirements of conciseness and detail in LR evaluation, which covers all

¹ Hassan Ahmed, Department of Communication Systems, Lancaster University, Lancaster, LA1 4YR h.ahmed@lancaster.ac.uk

² Lampros K Stergioulas, Department of Communication Systems, Lancaster University, Lancaster, LA1 4YR l.stergioulas@lancaster.ac.uk

³ Costa S Xydeas, Department of Communication Systems, Lancaster University, Lancaster, LA1 4YR c.xydeas@lancaster.ac.uk

aspects of LR quality. Such evaluation data is collected upon the completion of an LR delivery. In case of self-contained (independent) LRs of duration equivalent to 10 hours or longer, evaluation data is collected from both learners and course tutors, whereas for LRs of a shorter duration, evaluation data collection is applied only to tutors.

The LR evaluation results are presented in two modes, a *summary* mode and a *user defined queries* (UDQ) mode. A user can get general LR evaluation information via the summary mode. However if the user wants to examine specific LR quality aspects then he can define his own queries via the second mode of operation.

Section 2 describes the generic architecture of a learning resource exchange/brokerage platform. The LR evaluation engine as part of such a platform is discussed in detail in Section 3, which addresses in detail the issues of evaluation data collection, processing and presentation. User feedback results obtained from real trials conducting LR delivery and assessment through the proposed LR evaluation prototype engine are also presented in section 5. Conclusions are given in section 6.

2. DESIGN AND ARCHITECTURE

An LR exchange/delivery facility may consist of various functionality modalities, concerning delivery, evaluation, user interaction, system and user administration, metadata handling, contract/transaction authorisation, delivery management etc.[6].

Such a brokerage system may or may not store LR content. In the latter case, only an interface layer may be offered, which provides communication functionality between the brokerage system and various delivery systems. This is the responsibility of a delivery manager, which provides authentication and authorization services, delivery negotiation and delivery supervision. In such a system, the role of an *evaluation engine*, as proposed here, is to collect and store LR evaluation data and provides search, analysis and presentation facilities on the stored data. A suitable architecture for an evaluation engine is shown in Figure 1, comprising of the following four parts:

- Data collection tool
- Data analysis and presentation tool
- Questionnaire database
- Evaluation database

2.1 Data Collection Tool

The data collection tool facilitates the collection of LR evaluation data from LR providers and consumers as well as data related to other users activities (such as users' logs, background data, LR acceptance data). This data is collected at predefined times; such process can be initiated at pre-specified interval during delivery or following completion of delivery. LR providers may also supply evaluation-related

data, such as validation by independent experts (peer review), accreditation and previous evaluation information etc.

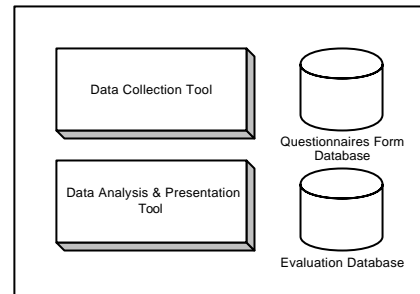


FIGURE 1: MAIN ARCHITECTURE OF LR ASSESSMENT ENGINE

The data collection from users takes place by means of on-line questionnaire forms. The evaluation questionnaire forms are adaptive in terms of “past evaluation history” and the current, specific evaluation requirements for each LR, in the sense that their structure and level of detail may vary for different LRs according to perceived user needs (two-level questionnaire structure). Furthermore, the selection/formulation of questionnaires depends on the type of user (learner, academic tutor or provider), since suitable types of data are required from different types of users. The selection of a suitable questionnaire form is also dependent on the type of LR (e.g. “live”, “packaged” or “hybrid”). More specifically, an evaluation questionnaire consists of:

1. *First-level attributes*, which are compulsory taking their values from a standard, *minimum length* (short) questionnaire.
2. *Second-level details*, are evoked by a *Questionnaire Formulation/Compilation Process* according to reconfigure-able specific criteria (the corresponding questionnaires are completely adaptable, initially set by the system to a *default* form).

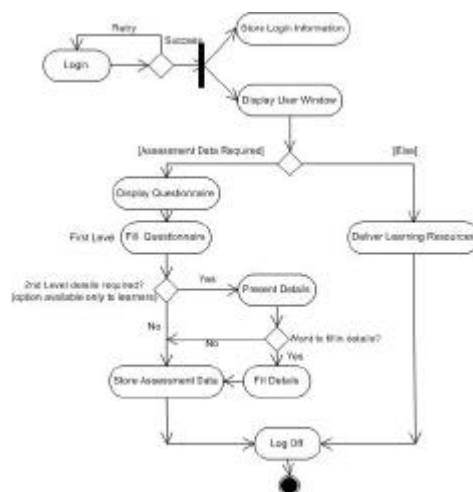


FIGURE 2: DATA COLLECTION FROM LR USERS/CONSUMER

The evaluation engine initially presents users with the short questionnaire (see figure 2). Depending on subsequent entries and following certain criteria/rules (pre-specified in terms of thresholds for perceived shortfalls/poor performance defined on each first-level attribute), the system might (or might not) activate the second-level *detail* structure in the questionnaire. In this way, questionnaires supported by such flexible, deployable structures are adapted to the changing sensed/perceived requirement and provide the necessary information for LR providers/tutors to focus on specific perceived LR weaknesses, with the view of introducing targeted improvement, replacement, or complementary support. This kind of questionnaire/adaptation provides a good reconciliation of the contradictive requirements of (a) short, user-friendly questionnaire/evaluation data size and (b) detailed but time consuming evaluation of all LR quality aspects. Furthermore, questions of the second level questionnaire can be user-defined, thus enhancing the overall feedback process. Data collection from tutor-users follows a similar scenario, with the notable exception that in this case second-level questionnaires are used as standard. The process of data collection from LR consumers is shown in Figure 2. The attributes (first-level) and default second level setting were defined in agreement with recent evaluation methodologies [7] [8]. Depending on the type of user, the following (first level) attributes are employed:

Student Questionnaire

1. Content
2. Interactivity Characteristics
3. Presentation
4. Delivery and Support
5. User Satisfaction

Tutor Questionnaire

1. General LR quality related issues
2. General Delivery issues
3. Quality of Student Assessment (if applicable)
4. Examination Data

. General user reviews/comments are also collected and stored in the evaluation database in the form of open text. Each first-level attribute is split into a sufficient number of second-level details to target more specialized areas of evaluation

Furthermore, a prior LR evaluation information can also be collected from LR providers. This collection process can be viewed as part of the LR provision process (figure 3).

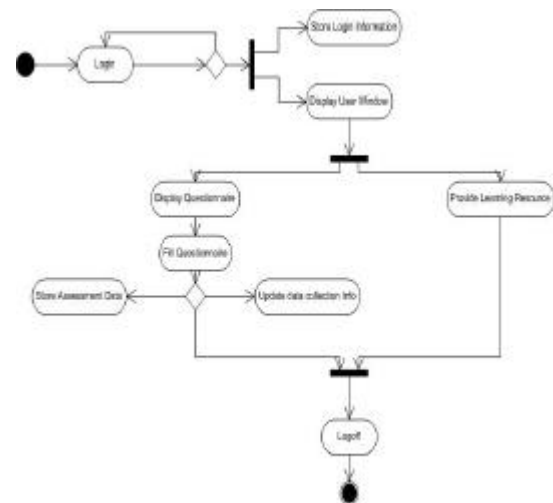


FIGURE 3: DATA COLLECTION FROM LR PROVIDERS

2.2 Questionnaires Database

Evaluation questionnaire forms (both standard first-level, and default or user-defined second level) are stored in suitable metadata structure within the questionnaire database.

2.3 Evaluation Database

All evaluation data (collected from users via the questionnaire defined in the questionnaire database) are stored in this database.

2.4 Data analysis and presentation

Users can extract information from the LR evaluation database via a versatile query interface. Two modes of interaction are provided: (i) a summary mode and (ii) a user-defined query (UDQ) mode; both of which are supported by *Help*, *Zoom* and *Scroll facilities*.

The summary mode provides instant access to aggregate evaluation result tables. Users are able to access important evaluation statistics (*indicators* of evaluation quality) using a simple and direct query environment. Summative Queries are initiated through the activation of *single function buttons* in a web page interface and the system responds with specific evaluation summary data. Figure 4 shows the user interface, which consists of 2 sections: a query selection area and a graphical/text display area. After the users selects a query from a fixed set of queries, the results are displayed in a pre-defined graphical mode. The results/graphs to be displayed correspond to first level questionnaire forms.

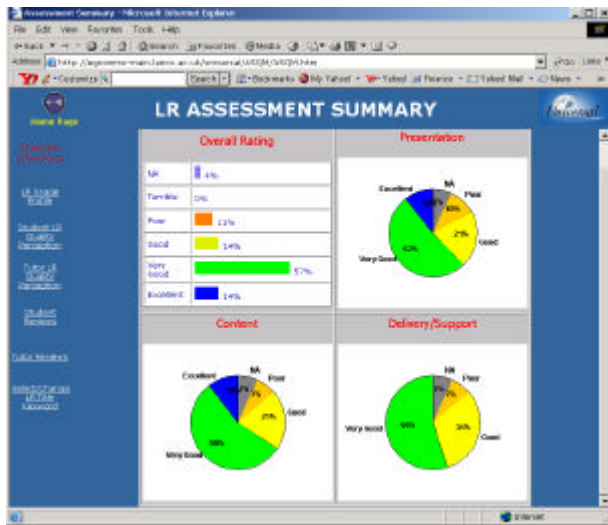


FIGURE 4: EVALUATION SUMMARY

The second mode, UDQM is more advanced, involving detailed nested, user-defined queries. The user is presented with various options for extracting, analyzing and displaying assessment data. UDQM thus enables users to *customize* queries to their specific needs. User-driven queries not only provide users with specific information, but also reduce the chances for data to be extracted erroneously or to be misinterpreted. UDQM also includes example Demos (based on previous popular queries from past history) for the benefit of the user. An example of such a query formulation screen is given in Figure 5.

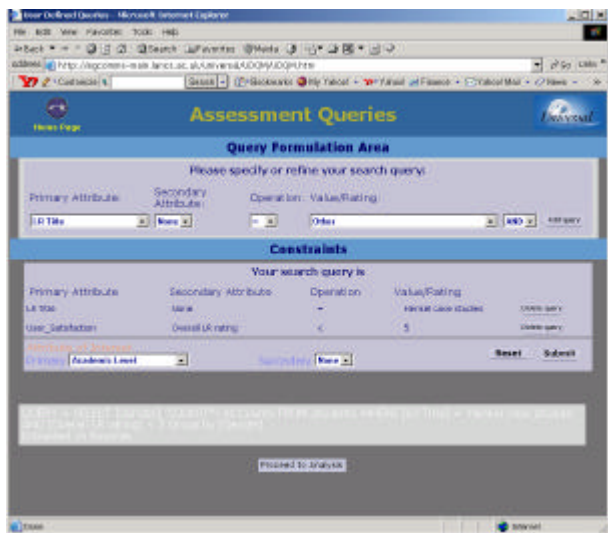


FIGURE 5: USER DEFINED QUERY MODE

A number of numerical/statistical functions are available for the task of processing the extracted dataset(s). This analysis stage should however be optional, since users may only want to view the unprocessed (raw) data and do the processing using their own tools. The data analysis and presentation facility is in the form of an interactive interface/display

window (Figure 6). The user has a variety of display options, including capabilities to process/display more than one dataset at the same time and to cross-examine data across different LRs.

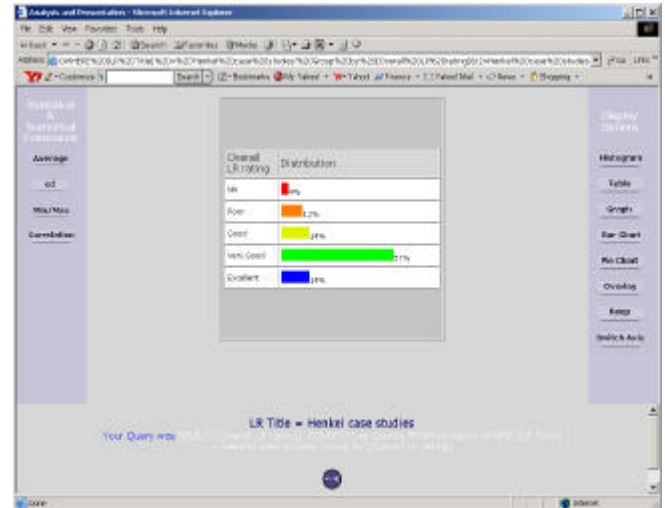


Figure 6: Data Analysis and Presentation

3. REAL LIFE IMPLEMENTATION AND RESULTS

Implementation of a prototype system, demonstrating the capabilities and benefits of the proposed design for an evaluation engine took place within the large-scale European project UNIVERSAL [9]. The UNIVERSAL project provides a Pan-European LR exchange service for higher education, through the EducaNext project.

The evaluation engine prototype was developed as part of the brokerage platform, which is based on a modern XML:RDF-bonding, metadata model, open-source philosophy and web-based technology.

3.1 Early testing phase: Role Playing

In the formative stages of the UNIVERSAL brokerage platform, real-life tests were carried out, asking participants to play roles of different types of users and evaluate the assessment engine from a user's perspective (see Figure 5). Users were asked to view/analyse LR assessment data from real life trials, pretending to be

1. A student and try to "simulate" what a student (who is also a UNIVERSAL user/perspective user) needs from an LR-AE.
2. A tutor who is using the UNIVERSAL platform and try to "satisfy" his requirements by using the LR-AE.
3. A provider who provides (or intends to provide) an LR to UNIVERSAL, and use the LR-AE to "simulate" what a provider would expect from an LR Assessment Engine.

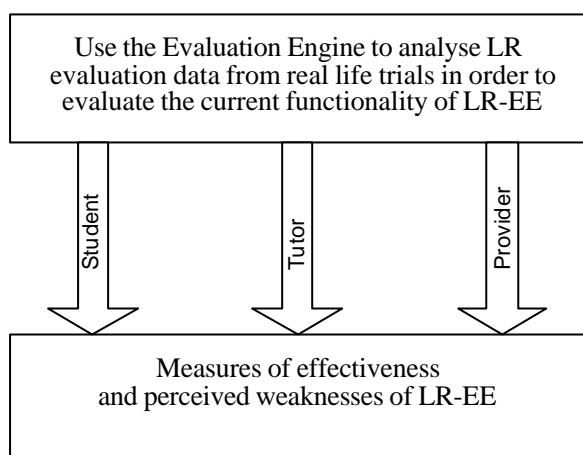


FIGURE 5: ROLEPLAYING IN REAL LIFE TESTS

3.2 Testing Phase

Following the completion of the service the evaluation engine prototype has been subjected to more rigorous testing, as part of large scale trials of the entire Brokerage platform, where the EducaNext public service is fully available on-line.

3.3 Results

Feedback from both testing phases suggested that :

1. The functionality of the adopted evaluation prototype is of a satisfactory level and
2. The two-level questionnaire structure (first level short questionnaire and second level lengthy/detailed questionnaire) is well accepted by all types of users.
3. Invoking only the long questionnaire for student-users was not acceptable, as it required considerable time to complete the questionnaire. There were no such complaints reported for the usage of short questionnaire. However, the option of using the long questionnaire is still be available for certain extra keen students.
4. Users were concerned with eLR evaluation data integrity issues. For example, a user could login more than once and give false or redundant (replicate) data. This is an important issue, to be addressed by introducing preventive measures in the form of security-enabling software agents, installed on the end-user computer, which inhibit the reuse of questionnaires. Also the system will store user login information and manage the evaluation activities in a way that prevents multiple responses of the same questionnaire by the same user.

4. CONCLUSION

An adaptive, flexible design is proposed for e-learning resource evaluation. The proposed design has considerable

merits, in-terms of deployable/adaptable data collection, extraction, analysis and presentation, and enables the user to define his own criteria for LR evaluation and control the entire process, including type of data collection, type of indicators, form analysis and presentation.

ACKNOWLEDGEMENT

The work described in this paper was undertaken as part of the UNIVERSAL IST European research project, funded by the European Commission.

REFERENCES

- [1] Britain, S. and Liber, O., 'A framework for pedagogical evaluation of virtual learning environments', *JISC JTAP final project report*, <http://www.jtap.ac.uk/reports/htm/jtap-041.html>. 1999.
- [2] Draper, S., Brown, M., Edgerton, E. Henderson, F.P., McAteer, E., Smith, E.D. and Watt, H.D., *Observing and measuring the performance of educational technology*, TILT: University of Glasgow: Glasgow., 1994.
- [3] Draper, S., Brown, M.I., Henderson, F.P. and McAteer, E., 'Integrative evaluation: an emerging role for classroom studies', *Computers and education CAL95 special edition*, 1996.
- [4] Jones, A., et al., 'Evaluating CAL at the Open University: 15 years on', *Computers and Education*, 1996, 26 (1-3), 5-15.
- [5] Oliver, M. & Conole, G., 'Evaluating Communication and Information Technologies: a toolkit for practitioners', *Active Learning*, 1998, 8, 3-8.
- [6] Stergioulas L.K., Bernd J. S., Ahmed H. and Xydeas C.S., 'Electronic Educational Markets from a Total Quality Management Prospective', In the same proceeding, 2002.
- [7] Doughty G.F., Magill J., and Turner I., 'Interactive Multimedia Computer Based Assessment', *Active Learning*, 1994, 1, 38.
- [8] Brown M.I., Doughty G.F., Draper S.W., Henderson F.P., and McAteer E. 'Measuring Learning Resource Use', *Computers & Education*, 1996, 27, 103-113.
- [9] UNIVERSAL project, The UNIVERSAL Exchange for Pan-European Higher Education, <http://universal.infonova.at/UNIVERSAL>