User-centred Evaluation of an E-learning Repository

Giorgio Venturi, Nik Bessis
Department of Computing & Information Systems,
Institute for Research in Applicable Computing,
University of Bedfordshire, Park Square, Luton, LU1 3JU
United Kingdom
{giorgio.venturi, nik.bessis}@beds.ac.uk

ABSTRACT
The aim of this paper is to reflect on the evaluation of DELTA, a distributed learning resources repository. We employed scenarios and claims analysis techniques so as to derive assessable evaluation goals. These goals were related to three quality dimensions: quality of use, pedagogical effectiveness and acceptability for the end-users. We employed a set of six quantitative and qualitative research methods and triangulated the results obtained. The results demonstrate that DELTA met around half of the evaluation goals. This project was an opportunity to reflect on the degree of fitness for purpose of our methods. Two of the qualitative methods employed (user diaries and pedagogy workshop) were particularly effective in the triangulation process.

Author Keywords
User-centred, evaluation, e-learning technologies, scenarios, soft issues, triangulation.

ACM Classification Keywords
H.5.0. Information Interfaces and Presentation (e.g., HCI): General. D.2.1. Requirements/Specifications: Elicitation Methods.

INTRODUCTION
The aim of this paper is to reflect on the experience of evaluating the deployment of DELTA, a distributed repository for learning resources, in the context of the “East of England Resource Network” (EERN). DELTA is a Web Services implementation. In the context of EERN it has been extended to accommodate SSO (Single Sign On) via the use of a Shibboleth authentication component [1]. EERN is a regional pilot [2] involving two Higher Education (HE) and six Further Education (FE) institutions.

The primary aim of EERN is to enable teaching practitioners (lecturers, tutors and pedagogic material developers) in the region to author, share, deposit and annotate various types of learning resources. Practitioners can make their own resources available to the wider community by registering and classifying them. DELTA aims to add value to standard resource repositories, since it not only allows each resource to be tagged using standard metadata (such as title, subject, educational level etc.), but also in terms of its pedagogy. DELTA allows the practitioners to map a learning resource to its pedagogical context. This mapping process aims to make pedagogical choices more transparent to users. This could be useful for a novice teacher, or equally for those seeking to refresh their pedagogical understanding as part of their continuous professional development. Each resource in DELTA may be classified with respect to a resource type [3], learning outcomes (“Generic Learning Activity”) and stage of learning cycle (“Practice Activity”). The taxonomy of Generic Learning Activities is a derivation of the original Bloom’s taxonomy of learning outcomes – i.e. knowledge, comprehension, application, analysis, synthesis and evaluation [4]. The taxonomy of Practice Activities is taken from the conceptualisation of learning as a three-stage cycle [5]: (a) conceptualisation, (b) construction and (c) dialogue. These three stages are described in DELTA as (a) “exploring and presenting subject matter”, (b) “constructing and supporting learning” and (c) “reflecting on and discussing learning”. DELTA aims to facilitate teacher and learner-controlled material review and rating by the community. By making these reviews and ratings available to the community their value should increase and ‘grow in context’.

The evaluation of the EERN project started when DELTA had already completed one iteration of its development. During the first iteration, a usability walkthrough was conducted [6]. With respect to that, we intended to spend more effort and involve users. Three evaluation qualities were specifically identified: usability, effectiveness of pedagogy and acceptability of DELTA to its end-users. Our first goal was to make a breakdown of these qualities into assessable system evaluation goals. We had a set of nine
scenarios available, which had previously been elicited by the project manager in a workshop involving all the relevant stakeholders from the FE/HE partner institutions.

**METHOD**

The evaluation involved four main phases, namely:

A. We elicited the system goals by carrying out a claims analysis on the scenarios and deriving 24 claims. After a review with experts and stakeholders we synthesised these to nine claims;

B. We defined an evaluation plan including users, tasks and a set of qualitative and quantitative evaluation methods;

C. We collected the data in the actual context of use;

D. We analysed the results, cross-examining the results obtained by each technique.

**Scenarios and claims analysis**

Scenarios are stories about users and their activities that can "focus designers on the needs and concerns of people in the real world" [7]. They provide high-level, abstract representations of user-system interactions. Claims are hypotheses about the "effect of the features on the user activities" [8]. We elicited claims from each scenario. Here is the first of the nine scenarios we used:

Emily is the lecturer assigned to develop and deliver the new “Back to Study” course. The challenge is to provide an experience that enables individuals, who have been out of education for some time, to become skilled, confident and independent learners. Emily is keen to identify and employ, adapting where necessary, existing learning materials that can be used to support the development of study and research skills. These may be incorporated within face-to-face taught sessions or will be used as reference materials and self study exercises. She is told by a colleague that DELTA may reference some relevant resources. She decides to use it to find the relevant resources.

This particular scenario dealt with teaching practitioners searching and retrieving learning resources from DELTA. Users – new to DELTA – may need to refine the search results by subject (“study or research skills”) or pedagogical approaches, in this case “face-to-face” or “individual” learning. We elicited two claims from this scenario.

1. Simple and advanced search functionalities will be easy to use for people new to DELTA (novices).
2. Search results will be easy to interpret and filter by advanced search for novices;

Each claim is a natural language statement made up of three components:

- (f) the feature(s) under scrutiny (for example, “simple and advanced search functionalities”);
- (q) the quality(ies) assessed (for example, "easy to use");
- (u), the user profile(s) involved (for example, "novices").

Claims analysis identified the system goals being evaluated. In fact, while such an analysis is usually made to evaluate different design options against claims, we used the method to turn the scenarios’ rich description into items to be evaluated. It is easy to translate claims into Evaluation Goals (EG). For example, the Evaluation Goals derived from claims 1 and 2 are respectively:

**EG1.** Are (f) “simple and advanced search functionalities” (q) “easy to use” for (u) “people new to DELTA (novices)”?

**EG2.** Are (f1) “search results” and (f2) “filtering by advanced and pedagogical search” (q) “easy to understand” for (u) “novices”?

Not all claims have the same level of generality. For example, the first claim is more specific compared to the second that focuses on particular aspects of the search task (i.e. easiness to read and filter search results). However, both claims relate to usability – or better, “Quality of Use” [9]. We also discovered other types of issues which were not strictly related to the concept of quality of use. Let’s see another scenario we had:

Sandra is a tutor of a “Continuous Professional Development” (CPD) course. During the course of her teaching she finds an excellent case study of e-learning developed by University of Bedfordshire (UoB). She knows that her institution is using DELTA for sharing learning resources and case studies. After checking DELTA’s policy for including materials, she notes that she needs to write to the author to seek (copyright) permission to use the materials. Fortunately, there is a copyright statement on the UoB site saying the resources could be freely used for teaching educational purpose with a suitable acknowledgement. Sandra is not that familiar with DELTA pedagogical ontology and she has a very busy schedule this week, so she seeks help from a colleague to ‘tag’ the material so it could be retrieved. She was interested that during the tagging process the pedagogical ontology could be made visible. She and the ‘tagger’ had to think hard about some of the pedagogical tags. On more than one occasion they had to seek an explanation of why a particular tag was not made available by the DELTA system (i.e. was considered to be inconsistent with the underlying pedagogy represented by the ontology). She found using DELTA to create case study that she learnt a great deal through the pedagogical classification process.

This second scenario dealt with novice practitioners tagging learning resources with DELTA. Novices need guidance in using DELTA pedagogical classification, as the underlying...
ontology is quite complex. Novices should also feel that sharing learning resources is acceptable because it is part of their duties, it is required by their institution and it does not break the institution’s copyright policies. From the second scenario we elicited four additional claims:

3. Pedagogical search and classification shall be easy to use for practitioners with limited knowledge of pedagogy;

4. DELTA ontologies shall be understandable and valuable for practitioners, in order to understand the way they teach and think about ways to improve it;

5. Help on copyright policies shall be easy to access from the home page and understand for practitioners;

6. Retrieving and sharing learning resources with DELTA shall be acceptable for practitioners.

Claims (3) and (4) relate to the DELTA pedagogical classification and (5) relates to the copyright policies. In fact, (3) and (5) are linked to quality of use – like (1) and (2) - while (4) is about the quality of the DELTA pedagogic ontology. Claim (6) is more “ill-defined” as it relates to global acceptability and organisational issues. Hence, we had to face the issue of how to evaluate these aspects. EERN evaluation has a complex socio-technical nature involving both a technical (DELTA) and a social system (HE & FE organisations).

We obtained a list of 24 claims. We discussed the results of our analysis with other stakeholders: the project manager, the pedagogy expert and two of the six managers involved. We identified that within the twenty-four claims there was some redundancy and thus synthesised the initial twenty-four claims into 9 claims. We then translated claims into EG (See Table 1), in the same way. Finally, we associated them to three quality dimensions (QD):

- Quality of use (U);
- Pedagogical effectiveness (P);
- Acceptability (A).

### Defining the evaluation plan

It was felt to be crucial to identify whether users were likely to accept DELTA, actively contributing with their own experiences, or whether they would be relatively passive users. We felt that our evaluation should include qualitative methods (such as one-to-one interviews) in order to better understand the motivations behind users’ behaviour. On the other hand, we had the opportunity to employ quantitative techniques like usage logs and usability tests. We were also interested in seeing how much qualitative and quantitative results matched through triangulation. Triangulation is a process of cross-checking findings derived from both quantitative and qualitative research [10]. This term is often used in social sciences to indicate that more than one method is used in a study with a view to double checking results.

We set up an “evaluation toolkit”, comprising a selection of quantitative and qualitative research methods. We selected six methods: one quantitative, three qualitative and two that are both:

- Web questionnaire (quantitative and qualitative);
- Usability tests (quantitative and qualitative);

<table>
<thead>
<tr>
<th>Evaluation goals (EG)</th>
<th>QD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EG1. Are (f) simple and advanced search functionalities (q) easy to use for (u) practitioners new to DELTA (novices)?</td>
<td>U</td>
</tr>
<tr>
<td>EG2. Are (f1) search results and (f2) filtering by advanced and pedagogical search (q) easy to understand for (u) novices?</td>
<td>U</td>
</tr>
<tr>
<td>EG3. Is (f) pedagogical classification (q) easy to use for (u) practitioners with limited knowledge of pedagogy?</td>
<td>U</td>
</tr>
<tr>
<td>EG4. Are (f) DELTA ontologies (q) understandable and valuable for the (u) practitioners, in order to understand the way they teach and think about ways to improve it?</td>
<td>P</td>
</tr>
<tr>
<td>EG5. Is (f) help on copyright policies (q1) easy to access from the home page and (q2) understand for (u) novices?</td>
<td>U</td>
</tr>
<tr>
<td>EG6. Are the activities of (f1) retrieving and (f2) sharing learning resources with DELTA (q) acceptable for (u) practitioners?</td>
<td>A</td>
</tr>
<tr>
<td>EG7. Is the (f1) registration and (f2) log-in process (q) easy for (u) novices?</td>
<td>U</td>
</tr>
<tr>
<td>EG8. Is the (f) registration process (q) clear for (u) novices in explaining how to set up DELTA?</td>
<td>U/A</td>
</tr>
<tr>
<td>EG9. Is the (f) pedagogical summary (concluding the creation process) (q) clear for (u) practitioners with limited knowledge of pedagogy in explaining the pedagogical choices available?</td>
<td>U/P</td>
</tr>
</tbody>
</table>

**Table 1. Evaluation goals and qualities dimensions obtained from claims analysis**
• Logs analysis (quantitative);
• Interviews (qualitative);
• User diaries (qualitative);
• Pedagogy workshop (qualitative).

Our goal was to use the web questionnaire to profile people attending the training DELTA sessions and to use this data to select the participants for the other activities (usability tests, interviews, etc). Logs analysis was included because of its ability to provide general usage figures over the whole span of the project.

A total of 29 end users participated in the evaluation. Approximately half of them (15) belonged to FE, 8 to HE, while the remaining 6 belonged to both sectors. Fifteen of the subjects were direct users (lecturers, tutors, module/unit developers, library/learning resource officers) whilst the remaining fourteen were indirect users (IT managers and support officers, academic and course managers). Not all of the 29 people were involved in all evaluation methods. 14 out of 15 the direct users we profiled participated in the usability tests; 3 direct users participated in the pedagogy workshop; 4 indirect users were interviewed. The following subsections explain the methods employed.

**Web questionnaire**
We constructed and pre-tested a questionnaire and we handed it to the participants of EERN before their training with DELTA. The first section of the questionnaire sampled participants’ demographics, professional role and experience in e-learning repositories. We aimed to profile the participants to be employed by the other evaluation methods. The second section probed participants’ perception of positive and negative aspects of e-learning repositories and personal expectations towards them. We applied a qualitative, thematic analysis to the data of the second section and we categorised the issues found. Our aim was to gather their perceptions and expectations before being trained with DELTA.

**Usability tests**
Fourteen direct users were involved in the usability tests (N=14). We used scenarios to identify tasks, which were piloted with students prior to their deployment with actual (direct) users. Users carried out six types of tasks as follows:

• Logging in;
• Searching by full text;
• Searching by field;
• Searching by pedagogical approach;
• Reviewing a resource;
• Creating a resource.

The “think aloud” test protocol for usability tests [11] was carried out for each task. We recorded all relevant metrics for quality of use (completion rate, errors, task time). After all tasks were completed, we prompted users about the way they perceived the usability of the system (e.g. “how did you find using …?”). This provided us with qualitative feedback. After that, we asked them to fill in a psychometric questionnaire - SUS [12] - to generate an assessment of the perceived usability. We analysed the results, identifying errors from the notes taken during the observation and ranking errors by impact. We calculated their impact by multiplying their frequency with estimated severity. Severity was ranked in three categories: “showstopper” (leading the user to the failure of task), “major” (giving to the user serious problems in completing the task) and “minor”.

**Usage logs analysis**
Logs of user activities were recorded over a period of five months. Each entry in logs included:

• Date and time of request;
• IP address of user client;
• Specific actions taken, namely:
  - User sessions;
  - Default searches;
  - Advanced searches;
  - Search by Subjects
  - Resources created.

Figure 1 (reproduced below) shows an extract from a typical log. The raw data was imported into a spreadsheet. We filtered out (by IP address) all activities made by DELTA support/training staff, managers and evaluators. We computed the frequency of each action, by month and also the average amount of time taken by each user session.

**Interviews**
We interviewed two support staff and two managers in their own work contexts. We identified in advance themes to discuss and we probed for an understanding of the indirect

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>IP Address</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/12/2005</td>
<td>11:59:57</td>
<td>212.219.90.105</td>
<td>register, email:*</td>
</tr>
<tr>
<td>09/12/2005</td>
<td>11:59:57</td>
<td>212.219.90.105</td>
<td>login, member:*</td>
</tr>
<tr>
<td>09/12/2005</td>
<td>12:04:04</td>
<td>212.219.90.105</td>
<td>Search my resources</td>
</tr>
<tr>
<td>09/12/2005</td>
<td>12:04:06</td>
<td>212.219.90.105</td>
<td>Search my favorites,</td>
</tr>
<tr>
<td>09/12/2005</td>
<td>12:05:00</td>
<td>212.219.90.105</td>
<td>Default Search</td>
</tr>
<tr>
<td>09/12/2005</td>
<td>12:29:57</td>
<td>212.219.90.105</td>
<td>Search By Subjects:*</td>
</tr>
</tbody>
</table>

Figure 1. An example from logs file

1 The questionnaire and other evaluation materials – logs, detailed description of tasks and transcripts - are not attached due to space limitations, but they are available on request from the authors.
user’s own perception of DELTA. We focused on organisational issues (i.e. curriculum, ownership of learning materials). We transcribed the interviews and conducted a qualitative thematic analysis of the transcript.

User diaries
The direct users were prompted during DELTA training to write an e-diary including their concerns or impressions of it. We provided a blank template with two columns (date and comments). We explicitly encouraged users to report both positive and negative issues and we promised to award two small incentives to the two best diaries. We asked the users to return diaries at the end of the project and we received 5 out of the 16 templates we sent out (return rate: 31%). Four diaries were written on the template and one on plain paper. Qualitative, thematic analysis was made with diaries’ content.

Pedagogy workshop
The aim of this workshop was to reflect on and discuss the way DELTA classifies learning resources, with a specific attention to the underlying pedagogical ontology. The workshop included an individual and a group session on the classification of resources. In the individual session users classified a learning resource2 being watched by an observer, who occasionally prompted the user to reflect or give an explanation of his/her behaviour. In the group session, two subject matter experts carried out a similar task discussing step-by-step their choices with the practitioners. We transcribed the whole event and we carried out a qualitative thematic analysis.

We communicated the event not as an evaluation activity, but as an occasion for professional development. Our intention at this stage was to follow the “co-operative inquiry” approach:

In co-operative inquiry these [“evaluator” and “subject”] exclusive roles are replaced by a co-operative relationship, so that all those involved work together as co-researchers and as co-subjects. Everyone is involved in the design and management of the inquiry; everyone gets into the experience and action that is being explored; everyone is involved in making sense and drawing conclusions; thus everyone involved can take initiative and exert influence on the process [13].

Out of sixteen people invited, three participated to the workshop event. However, outcomes were very positive; they were especially enthusiast of their direct involvement in the project, not as subjects but as co-designers; part of the transcription follows:

Facilitator: We need to revisit the pedagogical ontology and then to involve the users again.

User A: This is really an example of User Centred Design, because the educational users are involved in the design process.

User B: It is even better to have fewer people, than many people, you can go deeper.

Facilitator: Yes… to my mind this [event] is very successful, to have a dialogue… this is co-operative design.

RESULTS
DELTA fulfilled four of the nine goals we defined (Table 1; EG1, 2, 5, 7) and fell short on other four. We didn’t reach an agreement on goal (EG6) within our evaluation team as we felt that it was too early to make a valid conclusion about the acceptability of the system. We communicated the results to the DELTA design team and we explained how the system could have been improved in the successive development iteration. Issues were mapped to the three evaluation goals categories (quality of use, pedagogical effectiveness and acceptability) and coupled with design recommendations where appropriate.

Quality of use
We found out that simple and advanced search, and filtering tasks were easy to carry out (EG 1 and 2, Table 1). The results from the usability tests showed that users performed significantly better the second time that they performed a search with respect to task time (N=14, p<.05). The same did not apply to pedagogical search, which took significantly longer than other types of advanced search even after one attempt. Classification of learning resources was a demanding task (EG3), with 6 people failing or asking for assistance and 3 people refusing to do the task at all (N=14). Overall, 44% could not classify learning resources without assistance during the usability tests. Copyright policies were included in the help section, but few people (two in total) consulted it (EG5) during the usability tests. DELTA lacked of context help: the term “copyright” was ambiguous because it could refer both to the holder and the type of copyright. Results from the usability tests showed that logging-in (EG7) was straightforward (all people could complete the task). However, we discovered from two user diaries that the registration process was not quick and did not explain clearly (EG8) that, before using DELTA, it was necessary to configure the proxy server of the institution. When users tried to access to it, they got a misleading error message. The following is an extract from a user diary:

I felt frustrated because it was my first try after the training and I wanted to exercise what I learnt at the training session. I tried to log in at the usual URL […] but I couldn’t because the system told me that the credentials ‘were invalid’.

---

Even if in this case the problem was solved by mailing to the support team, we feel that these issues could have been foreseen and explained during the registration process. Insufficient help and generic type of error messages were a frequent source of problems with the system, but it was not the only one. Qualitative analysis of the observations of usability tests revealed a total of 29 usability issues. Half of them were minor (15). However, 3 of them were ranked as “showstoppers”:

1. Lack of contextual help explaining the vocabulary employed by DELTA;
2. Generic and misleading error messages;
3. Difficulties in recovering from errors; sometimes the back button of the browser didn’t work.

We recommended that all usability concerns identified should be addressed, giving a higher priority to those with the highest impact.

Pedagogical effectiveness

Results from the pedagogy workshop showed that users of DELTA found it very difficult to comprehend the vocabulary of the ontologies (G4), especially “Learning Resource Type”, “Generic Learning Activity” and “Practice Activity”. Help was provided on paper and online. Actually, many practitioners preferred to go ahead without looking for an explanation of the key terms (“I wouldn’t go to the guide to look for these things, I would just go to the application and tag it… if I see the options, they might just ring a bell”). The information contained in the pedagogical summary was confusing and ambiguous (G9). It did not show the difference between the choices made by the user and the choices made by the system. We also found out that practitioners wanted to associate to the resource more than one resource type and generic learning activity, but this was not possible due to the constraints embedded in the ontology. One lecturer stated that this bias could lead to people not finding the resources that they are looking for (“we are all different and we could wish to classify a learning resource in a different way”). For some direct users (e.g. library/learning resource officers) the advanced features of DELTA in terms of pedagogical classification could actually be a deterrent. For example one of them stated in her diary:

When demonstrating how to add [a] resource, system didn’t act as I expected/remembered […]. As I am [a] support staff rather than a tutor, my knowledge is lacking in this area, so although I think I may be able to add some useful resources, when working out how to categorise resources (with respect to pedagogy) I am at a disadvantage.

We recommended to review the vocabulary employed by the ontology and to make the pedagogical classification optional.

Acceptability

With respect to end-user acceptability, user diaries revealed that some users were frustrated because after being registered they couldn’t access DELTA. The system failed to provide clear instructions on how to set up the proxy server (G8). We couldn’t assess whether retrieving and sharing resources with DELTA was acceptable (G6). During the five months of the trial we collected conflicting evidence. Some people were positively impressed by the quality of resources:

I found a really useful idea on how to run and assess group discussions, which I printed off, and also an interesting link about pros and cons of lecturing […].

Other found them useful but not for their everyday teaching activities:

The resources will be useful for my future wish to undertake doctoral studies in e-learning pedagogies but at the current time I do not need to search for such materials on a regular basis.

Log analysis showed that the search and retrieval of resources with DELTA was discontinuous. We found out that this was due to the division of the teaching calendar in two semesters. One learning technology officer stated in one of the interviews that tutors and lecturers had to review their teaching material only at the beginning of each semester.

EERN expects learning resources to be tagged by their authors. However, the ten resources tagged by practitioners should be considered a few, compared to those tagged by DELTA staff (about 190). At the end of the project DELTA was still dependant on work made by the DELTA staff. A learning officer also pointed out that DELTA did not store learning resources by itself, which could lead to maintainability and sustainability concerns (“who will fix the broken links?”). Quantity and quality of contents are perceived by practitioners as a benefit delivered by repositories (See Table 2). A thematic analysis from the qualitative data of the web questionnaire identified the three most important benefits from users’ perspective ranked in an order of occurrence. Results show that users do value efficiency, high quality content and a higher awareness of relevant pedagogy.
Some organisational and cultural constraints should be removed in order to get full acceptance. Users were burdened by heavy schedules (“since January, [I am] trying to catch up with workload, and [I] cannot afford to get side-tracked on this website”) and no time slot was specifically allocated to them for learning how to use DELTA (“[the] colleagues I encouraged to participate simply had very little time to commit and no formal time allocation in which to do so”). Practitioners have not adopted a mindset of sharing their resources yet. An academic manager stated during her interview:

At the moment we are finding very difficult to get the users to share their materials even around the department […] I think that our tutors are quite far away from the idea of sharing learning resources outside of the college.

The acceptability of DELTA cannot be taken for granted in the future, unless all these issues are tackled by the project management.

DISCUSSION

The evaluation of EERN was a unique opportunity to reflect on the fitness for purpose of our methods. We employed a range of six research methods for studying the usage of a distributed system in its actual context. At the planning stage we spent a considerable amount of time in weighting pros and cons of each evaluation method. For example:

- Web questionnaire was primarily aimed to profile EERN participants (i.e. gender, age, position, etc). Its use allowed evaluators to gather the participants’ perceptions and expectations before being introduced to DELTA. This was useful to understand which issues at a high-level had the priority (i.e. usability, efficiency, content, etc).

- Usage logs analysis was chosen to provide quantitative data on the usage of DELTA within the project lifespan. Unfortunately, EERN evaluation phase lasted 6 months only. In addition to that, we couldn’t use this technique from the beginning, due to technical issues. This limited the usefulness of this technique because we couldn’t gather data for long enough. Overall, log analysis didn’t provide enough evidence to the final acceptability of the system.

- We thought to include focus groups with managers as a proper way to elicit organisational issues. After some discussion, we discarded this idea because we foresaw that managers would not feel free to speak about their internal affairs with possible competitors listening. We employed personal interviews to gather such type of feedback while respecting their privacy. Interviews were extremely useful to have a detailed account of the organisational issues.

- With respect to pedagogy, we planned to have an expert of pedagogy assessing how well learning resources had been classified by practitioners. We modified our plan after analysing the first set of data coming out from the usage logs and usability tests. Data clearly showed severe issues arising in the mechanics of classification process itself; practitioners couldn’t complete the process without assistance from an expert. However, logs and usability tests results were less clear about the reasons why the practitioners failed or even refused to carry out this task, or how to improve it. At that stage we were still in time to modify our plan and organise a workshop focused only on the pedagogical classification. The workshop provided some useful answers and helped us to formulate various design solutions.

- User diaries were taken into account after the project manager advised us to do it. Actually, we were not confident of the added value of user diaries, as we had already several qualitative methods in our toolkit. We didn’t see the reason of adding another evaluation method. At the end of the evaluation we had to reconsider. People compiled detailed and insightful diaries. By examining these diaries usability issues not covered by the tasks were selected for usability tests.

### Table 2. Perceptions and expectations towards e-learning repositories (e-LR).

<table>
<thead>
<tr>
<th>Question</th>
<th>Ranking in Order of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are the benefits that you believe e-LR can bring to the work of practitioners or to their students?</td>
<td>Efficiency</td>
</tr>
<tr>
<td>2. What are the negative issues that you find annoying in using e-LR?</td>
<td>Inadequacy of content</td>
</tr>
<tr>
<td>3. What are your personal expectations in using e-LR?</td>
<td>Efficiency</td>
</tr>
</tbody>
</table>
We also discovered some acceptability issues. Deciding which methods to select was also difficult for another reason; we couldn’t anticipate how and to what extent we could triangulate the results. Only at the end of the evaluation we could analyse the way in which the six methods interacted (see Table 3):

1. Web questionnaires were essential for profiling the participants of usability tests and the pedagogy workshop. These offered insights into the issues behind user acceptability.

2. Usability tests were extremely effective in discovering usability issues and in assessing their impact. The usability issues found were confirmed by the pedagogy workshop and user diaries.

3. Usage logs were not very effective for triangulation purposes. They provided an overview of usage trends, but only for a limited period of time.

4. The pedagogy workshop was useful for gaining a better insight into issues found by usability tests and for giving design advice on how to enhance users’ understanding of pedagogy.

5. User diaries were returned at the end of the evaluation process, while we were carrying out the interviews. User diaries were very effective in cross-examining results coming from other methods. These diaries confirmed evidence revealed by the usability tests and also the pedagogy workshop. Moreover, user diaries revealed the difficulties met by practitioners in setting DELTA up. We would not have been able to discover this issue by ourselves. In our opinion, the positive aspect of this technique was to allow the practitioners to make an account of their experience by themselves, without being filtered.

6. Interviews with managers were useful for gaining an insight into organisational (the division of teaching activities into two semesters) and cultural issues (sharing not part of the mindset of people). Interviews explained – to some extent – the behaviours recorded by usage logs.

In our view the three quality dimensions (quality of use, pedagogical effectiveness and acceptability) exhibit interdependency within this socio-technical system. Quality of use is the fundamental quality; if users do not find the system usable the other two dimensions are compromised. If the system is usable, pedagogical effectiveness is considered; if the system supports dialogue and reflectivity upon current teaching practices, then it will be employed by the user organisation. Acceptability is a higher-level issue; it is influenced by the two former qualities and by other organisational and cultural aspects.

**CONCLUSIONS**

An evaluation of quality of use, pedagogical effectiveness and acceptability of DELTA was completed in our study. Scenarios and claims analysis were used to define a set of evaluation goals. We employed qualitative and quantitative research methods in the evaluation. Triangulation of the results obtained from each was carried out. Teaching practitioners (lecturers, tutors and learning materials developers) and other indirect users were involved in the evaluation process. The results show that DELTA fulfilled four out of the nine goals that we defined and fell short on other four. Searching and filtering tasks were fairly easy even for novices. Classification of a learning resource was rather more demanding. The main usability issues were lack of contextual help, generic error messages and failures in recovering from errors. Practitioners also found it difficult to understand the key terms of the pedagogical ontology. We could not find enough evidence to make a conclusion on final acceptability. However, the results show that there are several constraints related to organisational and cultural aspects.

<table>
<thead>
<tr>
<th>Evaluation Method</th>
<th>Type of method</th>
<th>Quality Dimensions</th>
<th>Triangulating with</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Qualitative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Web questionnaire</td>
<td>X</td>
<td>profiling A (EG. 6)</td>
<td>2, 4, 5, 6</td>
</tr>
<tr>
<td>2. Usability tests</td>
<td>X</td>
<td>U (EG. 1, 2, 3, 5, 7)</td>
<td>1, 4, 5</td>
</tr>
<tr>
<td>3. Usage logs</td>
<td>X</td>
<td>A (EG. 6)</td>
<td>6</td>
</tr>
<tr>
<td>4. Pedagogy workshop</td>
<td>X</td>
<td>U (EG. 3)</td>
<td>1, 2, 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P (EG. 4, 9)</td>
<td></td>
</tr>
<tr>
<td>5. User diaries</td>
<td>X</td>
<td>U (EG. 1, 2) A (EG. 6, 8)</td>
<td>1, 2, 4, 6</td>
</tr>
<tr>
<td>6. Interviews</td>
<td>X</td>
<td>A (EG. 6)</td>
<td>1, 3, 5</td>
</tr>
</tbody>
</table>

Table 3. The six evaluation methods classified with respect to type (qualitative/quantitative), level of detail, the quality dimensions supported and the triangulation with the other methods (1 - 6).
factors. We found EERN an interesting case to reflect on the fitness for purpose of our evaluation toolkit. We were pleased to note the effectiveness of user diaries and pedagogical workshops. User diaries pointed to usability and acceptability issues that we simply didn’t discover using other methods. The pedagogical workshop proved useful in explaining the pedagogical issues and providing recommendations about the ontology and the classification process.

ACKNOWLEDGEMENTS
This research has been funded by JISC (Joint Information Systems Committee) under the Distributed e-Learning Strand (regional pilots). Special thanks are due to Terry Mayes (Caledonian University), Chris Fowler, John Scott, Michael Gardner and Catherine Fung (all from Chimera Institute, University of Essex). Chris Fowler and Terry Mayes used their experience to help us in refining the methodology of the pedagogical workshop. John Scott helped us several times in the first phase of the evaluation. Michael Gardner and Catherine Fung developed the DELTA module that provided the raw data of the usage logs. Special thanks are also due to the Faculty Project Director, Andrew Callard (University of Bedfordshire). We would like also to thank Des Stephens, Tim French, Asli Adisen, Peter Norrington and all the EERN participants: Valerie Schicker, Laura Locke, Clare Avery, Carol Howard, Aaron Burrell, Beth Boller, Emma Rozmus, Ian Hunter, Jo Mallett, Kav Dev, Lynn Whitmore, Sharon Stone, Sue Blything-Smith, Tina Hohmann and all the others.

REFERENCES
1. DELTA project site. Retrieved the 3rd of April at: http://www.jisc.ac.uk/deletdelta.html