

# Development-oriented eLearning Tool Evaluation: the Edukalibre Approach<sup>1</sup>

Luca Botturi – University of Lugano

Vania Dimitrova, Julika Matravers, Chris Tebb, Drew Withworth – University of Leeds, UK

Jutta Geldermann, Isabelle Hubert – University of Karlsruhe, Germany

**Abstract.** The Edukalibre development approach described here offers libre educational tools based on the open development model. The paper presents a development-oriented evaluation approach comprising three phases. Phase 1 focused on rapid and distributed usability inspection producing development-oriented feedback for designers, and was built around a usability inspection framework. Phase 2 was conducted through student feedback, and gave some indication of the deployment of the system in real settings. Phase 3 was aimed at a broader holistic evaluation considering the views of developers and other stakeholders. This paper arose from the recognition that significant body of work was required to address the issue of suitable and complete evaluation for libre software development for education. The evaluation approach described here offers a significant step towards addressing these emerging needs.

## **Introduction: The Evaluation of eLearning Tools**

The increasing amount of Open Source eLearning solutions has recently provided the educational sector with access to flexible alternatives to the traditionally expensive commercial products which many institutions cannot afford. With the rapid invasion of these technologies into educational institutions, the evaluation of these systems has become increasingly important. Many of these technologies follow the Open Development Model which relies on the constant improvement of software products through the modification by a wide range of developers. The provision of constructive feedback is crucial to the success of such a novel development mentality that is *highly iterative* with notably *frequent software releases* and a *geographically distributed development team*. The development of a suitable framework for the comprehensive evaluation of eLearning tools that follow the open development model remains a challenge.

Typically eLearning tools have been evaluated on the basis of particular aspects of a system such as their costs, their technical specification, the features they provide and their usability (Britain & Liber, 2000). The importance of usability evaluation, in particular, is frequently stressed in the literature. Magoulas et al (2003), for example, propose the use of heuristic evaluation within a layered evaluation approach for usability testing within Adaptive Learning Environments. Research also seems to have given priority to the evaluation of system usability against the background that the evaluation of the effectiveness of a learning environment remains difficult (Strother 2002). The level of a system's usability may not only influence the user acceptance of a system but may also greatly influence the learning process (Balacheff & Kaput, 1996). However, it needs to be acknowledged that usability is a necessary, but not a sufficient, condition for an effective learning environment (Pulichino, 2004). A further important and difficult issue, for example, which is often not directly addressed by usability evaluation is the evaluation of functional requirements for an eLearning system from the range of system users (O'Droma et al 2003).

The evaluation approach presented in this paper was developed within the framework of the Edukalibre project, as a method for assessing the tools produced in the project and, at the same time, to develop refinement guidelines and indications for the developers. For this reason, the next section will introduce the project and will describe the tools that were evaluated. We will then describe the evaluation approach providing details for each phase. Finally, a summary will provide conclusions and outlooks for further work.

## **The Edukalibre System**

Edukalibre is an EU funded project that examines the connection between libre software development and creation of open content for education. The project involves developers and educationalists from several

---

<sup>1</sup> The work reported here was undertaken within the EU (Socrates/Minerva) funded Edukalibre project, see project web site at [www.edukalibre.org](http://www.edukalibre.org)

European countries (Spain, Portugal, Germany, UK, Switzerland, and Check Republic) working together to study both technological and pedagogical aspects of the successful deployment of the libre idea in university teaching. The Edukalibre system, composed of several key tools and based on the libre methodology, has been developed to support teachers and students construct educational resources collaboratively. The evaluation presented here focuses on the use of two Edukalibre tools - ConDOR and GISMO - see below.

**ConDOR: Construction of Dynamic Open Resources** is a bespoke groupware application which runs as a modular component of Moodle. ConDOR (see Figure 1) was designed as a tool to be used by students or teachers for the collaborative construction of learning resources. It allows “explorer style” access to the document repository, allowing intuitive navigation of the file/folder structures, and simple tools for uploading files and creating new folders.

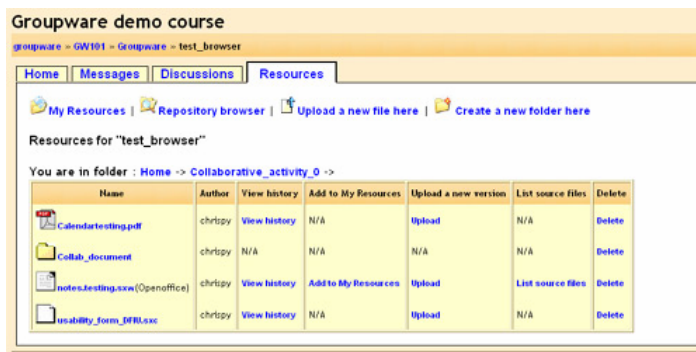


Figure 1 - Screenshot of ConDOR – tool for collaborative construction of dynamic open learning resources

**GISMO: Graphical Interactive Students Monitoring.** Student tracking data provided by Moodle is a valuable source of data that can be used by the instructor not only to check the students’ activities, but also to improve the quality of the materials. GISMO (see Figure 2) uses the students' tracking data as source data, and generates graphical representations that can be explored and manipulated by course instructors to examine social, cognitive, and behavioral aspects of both distance learning and conventionally taught students. Instructors can investigate the users' reading of course materials, which – at a glance - can give feedback on the quality of collaboratively constructed resources and their practical use by students.

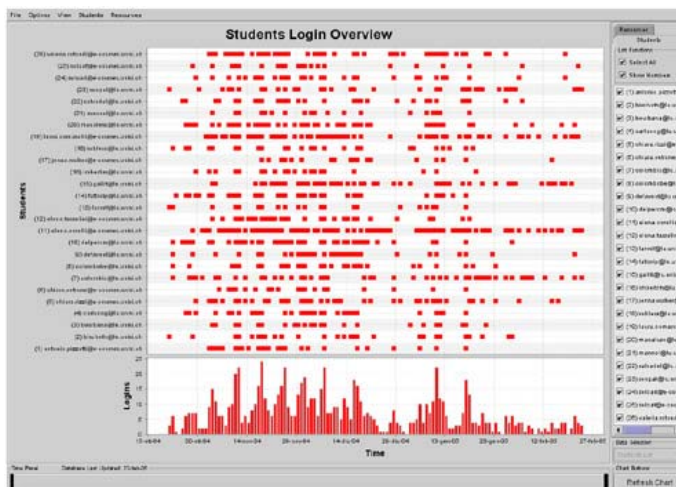


Figure 2 - Screenshot of GISMO – tool for monitoring the students’ work in both Moodle and ConDOR.

## The Edukalibre Evaluation Approach

### Challenges to the Edukalibre evaluation

The specifics of the open development model brought some challenges that had to be dealt with during the evaluation. Firstly, the whole design and development cycle was iterative, including frequent releases of software, sometimes not fully tested. Secondly, the development, as well as the evaluation, involved several geographically distributed teams. The advantage of this is the quick deployment in various settings and the wide dissemination. The downside is that the effort has to be distributed and the work properly synchronised. As a result, several small scale evaluative studies appear to be more appropriate than one large evaluation. Thirdly, the development was done asynchronously at times with several developer teams working independently in several countries (Spain, Leeds, Switzerland, and Portugal). This led to the need to be able to conduct evaluation in a flexible way to enable examining different tools within use context.

Another issue that challenged the definition of an evaluation approach for Edukalibre was the relationship

between software application design and development and the evaluation itself. A complete evaluation that looks for comprehensive data, requires real users in real setting and stable tools in their final version. Such an approach produces sound *ex-post* results which might be used to promote the tools or to improve the design in the future. Edukalibre, given its libre nature, had different needs: on the one hand, developers needed short term input and guidance relating to their development work in order to improve their tools; on the other, the project team felt the need of a wide evaluation of the impact of the tools in a real higher education setting in order to evaluate the usefulness and effectiveness of their product.

### Evaluation outline

Based on these goals and taking into account the specific needs of the libre development model, we developed a three-phase evaluation approach (see Figure 3).

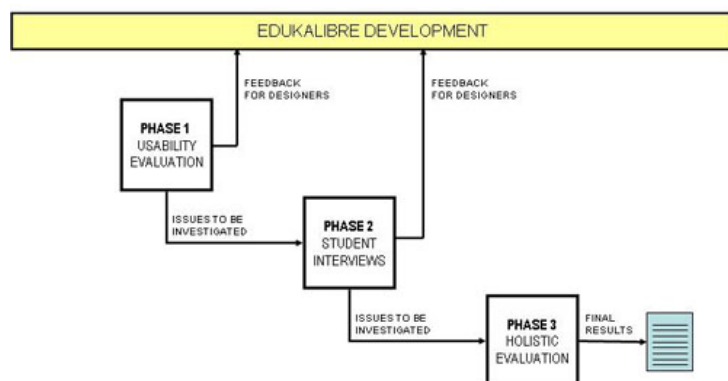


Figure 3 - Evaluation approach outline

Phase 1, conducted Oct-Nov 2004, included several studies involving experts and users from different partner institutions and focused on usability. Phase 2, conducted in Dec 2004, focused on student feedback, in order to verify the results of phase 1 and to investigate the reasons for the problems detected during the usability inspection. Phase 1 and 2 were then complemented with a broader evaluation in Phase 3 to match the expectations of both developers and users. This phase has commenced and will complete at the end of the project (end of 2005).

The three phases are discussed in detail in the following sections. For each phase, a general outline is followed by some examples of the kind of feedback produced, along with its impact on the project.

### Phase 1: Rapid Usability Inspection

**The MiLE Usability Method.** Usability evaluation is a narrow-scope assessment of how well the users can achieve specific goals with a web application. With the Edukalibre framework, usability evaluation was selected because (a) it could be conducted in a short time to keep the pace with the needs of the developers; (b) it could produce design-oriented feedback useful for the refinement of the application; and (c) it could be conducted in a uniform way by different teams in a libre fashion, if a specific methodology is followed. In order to meet the conditions expressed under (c), the MiLE method was selected. MiLE (Milano-Lugano Evaluation method) is an experience-based evaluation framework for web applications that strikes a healthy balance between heuristic evaluation and task-driven techniques (Triacca et al., 2004). MiLE can be used flexibly at different levels of granularity. It offers reusable tools and procedures to carry out both expert reviews (called *inspections*) and user testing within budget and time constraints.

MiLE uses scenario analysis as the driver for usability evaluation. Scenarios are then combined with selected usability heuristics in order to provide analytic feedback on the different aspects of the application design. The MiLE method moves through five stages. Stage 1 is analysis of the application requirements, in order to get information for constructing goals, scenarios and user types. Stage 2 defines macro scenarios (i.e., broad “stories about use” - cfr. Carrol, 2002) and user types (i.e., profiles of typical users of the application). Stage 3 defines scenarios (i.e., smaller-scope use cases), goals for each scenario, and tasks for each goal. The next two stages form the evaluation itself. Stage 4 is the expert inspection, guided by scenarios, goals, task and MiLE heuristics, which provides the identification of main issues. It can be combined with user testing, in order to confirm and further investigate the issues. Finally, Stage 5 produces inspection reports and draws guidelines for further development. The outcome of the steps in the analysis and planning phases make it possible to define a usability protocol that can be replicated by different teams, so that the work can be distributed and then the results collected and merged.

Usability involves a number of different issues. For this reason MiLE proposes to articulate usability judgments around the following dimensions: *Content*: is the information provided by the software complete, clear and accurate concerning the operations that the user is trying to accomplish? *Orientation*: does the user understand where s/he is, at what step in a given process or in what position in the application? *Navigation*: can the user easily reach the tool/page/content s/he is looking for? Can s/he identify the right path to it? *Predictability*: is the user able to know what is going to happen when s/he clicks on a link/button? Does s/he get surprises? *Layout*: is the layout consistent? Can users identify functional areas that support easy interaction? *Legibility/Graphics*: can the user read the content? Are the graphics useful for working with the application or are they confusing? The usability framework and the usability dimensions were put together in a single document, which was then distributed to all project partners, along with indications about how to conduct the inspection. GISMO was evaluated by 3 inspectors (a teacher scenario) and ConDOR by 6 inspectors (3 teacher and 3 student scenarios).

**Results.** In order to provide the flavour of the results obtained with MiLE in Edukalibre, and their impact on design, two sample results are reported below. Inspecting GISMO, an issue emerged concerning orientation and navigation for Task 02: See who logged in during last week. In order to see logins, the teacher should select *draw chart* and then *login overview*, which is not very straightforward. The identified solution is to organise menu semantically (i.e., following the teacher's perspective), and not operationally (i.e., following the programmer's perspective). For ConDOR, an issue was identified concerning Task 08: Identify the least and most productive contributors of Group 1 (a group in the scenario), concerning the content dimension. Actually, the ConDOR interface is document-oriented, and in order to track the activity of single students, the teacher had to browse and count messages/posts manually. This was indeed a common task among teachers, which was not supported by the tool. It was therefore included in the plan for next development to provide summary information about the activities of each group member (including the number of posts and file contributions, the date of the last contribution, etc.).

Usability evaluation provided useful indications for further development – yet it is by nature not complete, as it only focuses on a specific feature of the system, and it does not consider its educational impact and its effectiveness. This is the reason why Edukalibre moved forward to the next two phases.

### *Phase 2: Student Feedback*

In order to verify the results of the rapid usability inspection (phase 1), student feedback was gathered in phase 2. At the University of Karlsruhe, a seminar was organized with 14 German students in the advanced study period of Industrial Engineering. During four months, the students elaborated an environmental business game, which comprised computer supported case studies for location planning and technique assessment. Four groups of three students worked on a specified topic, representing fictitious companies faced with strategic questions and the subsequent decisions. In addition, two individual topics were related to business games and open source software (Geldermann et al., 2004). During the preparation of the business game, the students were invited to use the Edukalibre platform that was specifically set up for them. The platform comprised of the general Moodle platform (with news forum, calendar, documents etc.) plus the developments of Edukalibre up to that date, including ConDOR and Gizmo.

**Data collection.** For producing development-oriented feedback, a questionnaire was used with altogether 15 main questions, developed according to the issues identified in Phase 1.

**Results.** Among the 14 users of the platform, 8 logged in weekly and 6 hardly ever. All the users seldom used the groupware and rarely downloaded a file. The groupware was mostly used to download documents, sometimes to upload files, to try to communicate within their groups or for the chat. However, apparently the students used the features of Moodle (calendar, news forum, download files) more than the groupware, largely due to the Moodle based organisation of the class. The system was not too difficult to use for most of the students. The uploading time was not judged too long either. The best features of the system according to the students are the calendar, the news session, the repository browser (however extremely hidden – a finding that emerged also in Phase 1) and the multiple download formats offered by ConDOR. The worst features cited by the students are the messaging that apparently did not work all the time. Moreover, a lack of feedback by the platform was found faulty. Accordingly, the missing feature was an e-

mail notification of the messages sent or received or when getting an answer to their own posts.

It can be seen from the features that were not discovered by the user group that they apparently had little initiative to search for what they needed. The students did not seem very keen about having to use an Internet platform for the purpose of the assignment they had to make. Maybe the type of course used to test the platform was not an ideal case, but the Edukalibre platform is being designed for such a kind of seminar. Indeed, the groups are made of only three students and they prefer to meet regularly instead of using a platform. In a way, the students seemed to be conservative. Although they are eager to try different kinds of learning such as a “business game seminar”, they seem unsettled when the learning experience differs from their routine (e.g. an Internet platform and Open source against MS Office). In addition, the terms resources or shared resources were probably understood as “documents provided by the teachers” more than “my resources” signifying “our group documents” in the resource browser of the groupware. While most students agree about the importance of the shared resources, they are more cautious about the idea of collaborative authoring, especially the fact that other people can modify what they write. Although they like to see what others have written, they are not really enthusiastic about people reading and modifying their work.

In conclusion, the students of this first test group were not totally convinced of the benefits of any Elearning platform mainly because they can meet regularly and attend “traditional lectures”. They did not use the platform regularly, which explains most of the lukewarm answers found in this first questionnaire. More documentation such as a tutorial for the Edukalibre-platform would be beneficial. Yet, the idea is growing and the acceptance and extensive use of such a platform by the students could arise soon.

### *Phase 3: Matching the expectations of developers and users*

The main idea for Phase 3 is to investigate the developers and find out if and how their assumptions match with those of the users of their applications – students and teachers. Why is it necessary to investigate the developers? Although powerful for delivering results during development, we believe that standard HCI recommendations for usability testing are inherently limited in terms of what they can reveal. Usability problems are, at heart, conflicts between the assumptions of the developers and the needs of the users. These may emerge because the needs of users have not been adequately researched or because of inherent differences in how different groups perceive a system, and this can apply both to the specifics of a system (functionality), and generalities (such as why the system was created in the first place). Indeed, many usability evaluations do not take place in the full learning context (Lim et al., 1996) and often struggle to test unexpected and creative use of the system. Nor do these methods consider technologies as being dynamic, existing in a constant state of evolution (Star 1999).

In a survey as a first step in Phase 3, developers were asked what, in their opinion, would constitute "success" for eLearning generally and Edukalibre specifically. E-learning was seen by the developers as complementary to face-to-face teaching that would further enhance learning and enable better engagement of students, as well as would give opportunities for immediate feedback from the teacher. The developers would consider Edukalibre successful if it was deployed in real courses (note that the educational impact was not explicitly mentioned). The priorities were to create a system that was easy to use, simple to explain to the students, answering a direct need, stable and having support. Success in Elearning projects was defined when the users of the system (teachers and students basically) could use the system easily and take all the advantages of the system without any extra effort to learn the normal use of the system. There was considerable variety in these answers: referring to pedagogical issues, general issues ("enhances learning", "improvement in quality of teaching"), or not pedagogical at all ("adoption of the system"). Taking this latter point as symptomatic of the issue, a desire to have a new technology adopted by a number of institutions is a very different motivation for creating that technology than enhancing teaching. It may also require developers to write the technology in particular ways. Certain technological elements may have been included in Edukalibre because of the desire of developers to "play" with certain technologies and innovate in order to create something substantially different from the competition. In their own cognitive environment, this is understandable - it is through engagement with new technologies that developers acquire status, job satisfaction, and so on. But simply including new technologies is not directly addressing the educational needs of students (or teachers). Any teaching which uses the system may be forced to change to fit the system, instead of the reverse - pedagogical needs driving technological development.

This is not meant to tar Edukalibre with the brush of commercialism: it is merely an example of the potential ways in which Elearning technologies, being complex and significant to a wide variety of stakeholders (Benson, 2002), can become an arena which different interests and perspectives compete to define. The "social shaping of technology" thesis (Mackenzie & Wajcman, 1985) suggests this. The ideal is first that assumptions are revealed by self-reflection; second, that they can if necessary be challenged and changed. This has been termed "double-loop learning" (Argyris, 1999). Ideally, this takes place in a continuous cycle of evaluation then subsequently, implementation. The more user needs can be included and underlying assumptions revealed, the more participatory the construction of the environment will become.

## Conclusion

Open Source eLearning applications have become increasingly common in today's educational environment. However, the lack of suitable evaluation approaches - that extend beyond the assessment of a system's usability - to provide developers of eLearning applications with constructive feedback remains an issue. This issue now extends further into a world where the Open Development Model has become a dominant development mentality. It was the objective of this paper to propose an evaluation approach for eLearning systems that follow the Open Development Model. It was argued that such an evaluation approach would have to be holistic in the sense that it offers feedback beyond issues of usability. Furthermore, it should be applicable within a development environment that was highly iterative with frequent releases, and that potentially involved a high number of geographically dispersed developers of various backgrounds.

It was not the intention of this paper to provide a recipe for the complete evaluation of eLearning systems. The evaluation approach proposed here supports a holistic evaluation of Open Source Software that accounts for these factors. The approach was used with fairly stable versions of a learning system that was developed following the Open Development Model, and we were able to trial our approach within real higher education teaching situations. Both students and teachers were highly motivated to use and test the software, because they were conscious of the significance and immediate relevance of their feedback. This paper arose from the recognition that a lot more work is required to address the issue of suitable and complete evaluation for Libre Software Development. The evaluation approach proposed here offers a significant step towards addressing these emerging needs.

## References

- Argyris, C. (1999). *On Organizational Learning*. Oxford, UK: Blackwell.
- Balacheff, N., & Kaput, J. (1996). Computer-based learning environments in Mathematics. In A. J. Bishop et Al. (eds), *International Handbook of Mathematics Education*, Kluwer, Amsterdam, 469-501.
- Benson, A. (2002). Using Online Learning to Meet Workforce Demand: A Case Study of Stakeholder Influence. *Quarterly Review of Distance Education*, 3(4), 443-452.
- Britain, S., & Liber, O. (2000). *A Framework for the Pedagogical Evaluation of Virtual Learning Environments*. JTAP Report No. 041.
- Carroll, J. (2002). *Making Use-Scenario-based design of Human-Computer Interactions*. Cambridge, MA: MIT Press.
- Geldermann, J., Treitz, M., & Rentz, O. (2004). Teaching Integrated Technique Assessment based on Open-Source Software. *European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, Aachen, Germany, 10.-12.06.2004*, PP 503-508.
- Gonzalez-Barahona, J. M., Dimitrova, V., Chaparro, D., & Mazza, R. (2005). Producing educational resources in the "libre" way: The Edukalibre project. Edukalibre technical report.
- Lim, K. H., Benbasat, I. & Todd, P. A. (1996). An experimental investigation of the interactive effects of interface style, instructions and task familiarity on user performance. *ACM Trans. Computer-Human Interaction*, 3(1), 1-37.
- Mackenzie, D., & Wajcman, J. (Eds.) (1985). *The Social Shaping of Technology*. Buckingham, UK: Open University Press.
- Magoulas, G. D., Chen, S. Y., & Papanikolaou, K. A. (2003). Integrating Layered and Heuristic Evaluation for Adaptive Learning Environments. In S. Weibelzahl and A. Paramythis (eds.), *Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems*, held at the 9th International Conference on User Modeling UM2003, Pittsburgh, 5-14.
- Mazza, R., & Milani, C. (2004). GISMO: a Graphical Interactive Student Monitoring System for Moodle. Paper presented at the *UK and International Moodle User Conference - MoodleMoot 04*, Oxford, UK.
- O'Droma, M. S., Ganchev, I., & McDonnell, F. (2003). Architectural and functional design and evaluation of Elearning VUIS based on the proposed IEEE LTSA reference model. *The Internet and Higher Education*, 6, 263-276.
- Pulichino, J. (2004). *Usability and eLearning*. eLearningGuild Report.
- Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, 43(3), 377-391.
- Strother, J. (2002). An Assessment of the Effectiveness of Elearning in Corporate Training Programs. *International Review of Research in Open and Distance Learning*, 3(1).
- Triacca, L., Bolchini, D., Botturi, L., & Inversini, A. (2004). MiLE: Systematic Usability Evaluation for Elearning Web Applications. *EDMEDIA 2004, Lugano, Switzerland*, 4398-4405.