Collaborative Open Environment for Project-Centred Learning

European Commission Sixth Framework Project (IST-027073)

Deliverable **D 5.2: Evaluation definition: method and metrics**

This deliverable presents how evaluation will be done inside the Cooper project. The document presents an evaluation overview and an evaluation methodology. The evaluation in Cooper is performed at the conceptual, technical, and impact levels and this document analyzes how the evaluation will be done from all these points of view. The evaluation considers the aspects described in D 5.1 "Scenario and Requirements Analysis" and it will be the basis for the T5.5 which is "Perform the evaluation, analyse the results and elaborate the evaluation report"

Editor	*Valentin Cristea, Vlad Posea, Stefan Trausan-Matu*
Work Package	*WP5*
Status	*Final*
Date	*31 July 2006*

The COOPER Consortium

1. Learning Lab Lower Saxony (L3S), University of Hannover, Germany
2. Politecnico di Milano, Italy
3. Open Universiteit Nederland, Netherlands
4. Web Models s.r.l., Italy
5. Abbeynet S.p.A., Italy
6. LT Design Software GmbH, Germany
7. Università della Svizzera italiana (ALaRI), Switzerland
8. University Politehnica of Bucharest**,** Romania

Document Control

Title: **D 5.2: Evaluation definition: method and metrics**

Author/Editor: **Valentin Cristea, Vlad Posea, Stefan Trausan-Matu**

E-mail: valentin@cs.pub.ro, vladposea@yahoo.com, trausan@cs.pub.ro

Amendment History

Version	Date	Author/Editor	Description/Comments
1	23/02/2006	8 - UPB	Second draft
1.1	13/03/2006	8 - UPB	Third draft
1.3	30/05/2006	7 - AL	Requirements
		5 - AN	Collaborative tools (VoIP)
		2 - PM	Structured and unstructured processes
		1 - L3S	Recommendation system
		3 - OUNL	Contribution to several sections of the document
1.4		8 - UPB	Fourth draft
1.5	25/07/2006	2 - PM	Concept evaluation
		1 – L3S	General review
		7 - AL	General review

Legal Notices

Contents:

Executive Summary

Work Package 5 includes three case studies and the overall evaluation in the Cooper project. The evaluation aims to identify whether the models, scenarios, and tools developed in Cooper enhance the collaborative project-centered learning in higher education and industry. More specific, the enhancement refers to an improved management of virtual teams of persons with heterogeneous backgrounds and competencies, a better collaboration sustained by the use of knowledge repository and recommendation service, an adequate pedagogical scenario and competency assessment, and a good distant cooperation in a distributed virtual team through communication tools.

E-Learning evaluation is the subject of many previous works that were concerned with determining whether learning has been enhanced through the use of technology, how the learners interact with new tools and services, how usable were these tools, what was the content quality, etc. Despite the large number of evaluation studies, very few works refer to the evaluation of technologies used to support collaborative and cooperative learning in distributed environments. Also, there is a lack of evaluations developed for collaborative and cooperative project-centered learning.

The evaluation in Cooper is performed at the conceptual, technical, and impact levels. Concept evaluation addresses the question whether the models used in the Cooper methodology (supporting collaborative teamwork, building pedagogical scenarios and learner assessment involved in the teamwork and supporting the teamwork with recommendation services) are sufficient for a complete and accurate description of the domain of use cases. The technical evaluation will measure the capacity of the technology to generate and install flexible and customizable learning environments that implement didactical scenarios and assessment strategies. The impact evaluation addresses the changes that are brought about by implementing the Cooper methodology in terms of process, learning, cooperation, management support and effective use of technology.

The evaluation will be guided by the "Scenario and Requirements Analysis" document [Cooper 06a], which summarizes the set of requirements by focusing on the core problem considered by Cooper. Also, the evaluation will take into account the use cases in order to choose some indicators that need to be watched and the values obtained by the focus groups to be compared.

The present deliverable is concerned with the evaluation method and metrics. The deliverable organization is as follows.

Chapter 1 presents the evaluation issues and the ways they were considered in previous works.

Chapter 2 presents previous work in e-learning evaluation, organized in four sections: benchmarks (Section 2.1), criteria (Section 2.2), toolkits (Section 2.3), and standards (Section 2.4).

Chapter 3 is the core part of the document and presents the evaluation in Cooper. Section 3.1 specifies the purpose and scope of the evaluation, which capture the extent to which the results delivered by the project will contribute to forming a protected, shared COOPER environment, that will be easily deployed over any University's or Company's Intranet.

Section 3.2 describes the main principles guiding the evaluation.

Since the evaluation takes into account different user categories (roles) and tailors the metrics and instruments accordingly, the roles used in evaluation are presented in Section 3.3.

Section 3.4 discusses the evaluation objects, stakeholders and methods. The section is organized along the three levels of the evaluation process, namely: concept evaluation, technical evaluation and impact evaluation. The section also introduces the metrics in conjunction with some evaluation instruments: site activity logs, questionnaires, and social networks.

Section 3.5 describes the evaluation procedure.

Chapter 4 includes references.

Appendix 1 concludes the deliverable with the presentation of a demo questionnaire for students.

1 Evaluation Overview

E-Learning evaluation is the subject of many studies, projects, and papers. Several organizations, programmes and companies offer specialized services to maintain the e-learning platform, support the educators in using them, provide consulting services for technological aspects, and evaluate the Learning Management Systems / Learning Content Management Systems and various e-learning software applications (see for example *http://www.edutech.ch/*).

Evaluation studies were concerned with e-learning developments and aimed to determine whether learning has been enhanced through the use of technology, how the users interact among themselves through the technical platform, how the learners interact with different tools and services, how usable were these tools, what was the content quality, etc. Other studies and papers refer to the evaluation standards and benchmarks [BEEP 04, Frydenberg 02], evaluation measures [Bonk 02], evaluation methodologies and strategies [JISC 04, Pfister 99], and to the results of many evaluation projects [Cristea 05, Stone 03, An 04].

Despite the large number of evaluation studies, few of them refer to the evaluation of technologies used to support collaborative and cooperative learning in distributed environments. Lack in the evaluation techniques available is due to the difficulty of identifying the effect of a particular design feature on learning in such large and complex learning environments [Pfister 99].

Some important evaluation **issues** have been identified by Baumgartner [Baumgartner 96] for virtual collaborative learning. The first issue is the evaluation's **object**, which should be the inherently interactive learning process itself, and not the learner progress assessed by pre- and post- tests of learner knowledge. The second issue is the adoption of a **model** of social learning to describe the collaborative learning process. This model shall permit finding evaluation metrics for the technology that supports the collaborative learning. Thirdly, the **effectiveness** of the new technology should be considered at different levels such as learning design for cooperation, the fit of virtual learning technology with the cultural, organisational, economic and political context in which it is used, and others. The evaluation may lead to **recommendations** for improving the design, training the learners and tutors in effective strategies for cooperation, or for adopting new pedagogical strategies.

Pfister and Wessner [Pfister 99] define four basic steps of their evaluation approach: (1) defining the goal of the evaluation; (2) defining criteria by which success of the learning technology will be assessed; (3) designing an appropriate study; (4) assessing whether criteria have been met and distributing the results.

Defining the goals

The authors identify several possible **motivations** for the evaluation, such as quality assurance of a distant learning platform, the requirement of research and industrial partners, purely scientific interest, finding the best solution to support a specific feature of an e-learning environment, and testing the developer's assumption on how the environment should be used. Other reasons could be the validation of training as a business tool, justifying the costs incurred in training, help improve the design of training in order to provide better value and increased benefits for an organisation, help in selecting training methods (classroom, on-job and self-study methods), etc.

Cullen et al. [Cullen et al. 1993] introduce three concepts by which they define the **scope and purpose** of evaluation, namely the *object of evaluation* (what do you want to evaluate), *stakeholders* (who are the stakeholders and what are their interests), and *stakeholder learning* (what should each stakeholder learn as a result of the evaluation).

The object of evaluation can be multiple, such as the one defined for the DELTA project [Cullen 93], which included: the learning technology, the life-long learning enabled by the technology, a process of innovation in technology enabled life-long learning, and a process of innovation in technology enabled life-long learning through which programme participants pursue their interests. This explicit highlight of the compound evaluation object permits identification of the stakeholders (technology developers, educators, organisations, funding bodies). The identification of the stakeholder groups is important since they should be included in the evaluation process. Also, this serves the identification of stakeholders' learning.

Defining the criteria

This step offers the criteria by which the new models, methods, techniques under evaluation will be judged. In a first approach, the concerns of the stakeholders are considered. For example, a **research computer scientist** who wants to realise a new technical solution for tightly coupled synchronous work across the web could be interested in the smooth functioning of the technology, its usability, and supporting exchange and negotiation of ideas among learners. In a second step, more fine-grain criteria are developed to respond to the stakeholders' interests. In particular, for collaborative learning environments, which intend to support the collaborative learning processes among students, the criteria should permit to observe if, how, and when they really collaborate in order to validate the usage of such systems, and identify the collaboration level among the members of working groups. As stated in [Holst 00], the development of observable, reliable indicators of effective learning is difficult, and the outcomes of collaborative learning are inherently difficult to assess.

Some approaches [Henri 92] are based on the message content examination. Henri makes a cognitive analysis of the learning process. She defines five dimensions along which the content of messages can be analyzed (participation, interaction, social, cognitive and meta cognitive) and finds concrete, observable indicators for each of these dimensions in terms of actual message content, such as the number of messages that are subject specific and who replies to whom (in particular, whether learners reply to each other as well as to prompting by the tutor).

Designing an appropriate study

Designing the evaluation study is firstly concerned with establishing the **participants** and dividing the tasks among stakeholders [Holst 00]. Usually, the evaluation design should consider the expertise of the interested parties who are specialists in their field.

Another issue is the **schedule** of the evaluation. For example, in a *formative evaluation*, the results should feed back into design, which requires that developers and evaluators work closely together, and evaluation process produces results just in time to be considered in re-design decisions. Holst reports a time consuming experimental evaluation made on the prototype DOLPHIN whose late results were insufficiently used in the future design.

The third issue is the evaluation **approach** itself, which should fit the goal and scope of the evaluation. The *experimental comparison* isolates some particular factors, such as a new collaborative and cooperative tool, and studies its influence on learning. They are repeatable and objective, but more difficult to apply in virtual collaborative and cooperative learning due to the large number of factors that influence the learning process such as subject domain, the size

of the learning group, the nature of the learning material, the effectiveness of the tutor.. The alternative is the *ethnography*, which is based on the examination of how virtual learning technology complements or conflicts with existing social, working and learning practise. The ethnographer becomes part of the situation she is evaluating. The ethnographic study should be implemented in a way that avoids getting lost in overwhelming detail, and presents the results in a useful summarized form, which is not very simple to implement.

Assessing and distributing the results

The results of the study should be presented as **quantitative** descriptions of the evaluation criteria, but also as **qualitative** descriptions of those events and situations that were either particularly problematic or very successful. These can form the basis of suggestions for improvement identified by the evaluators, which should be negotiated with designers.

2 Existing Tools for Evaluating e-learning Systems

In this section, we present some tools proposed by previous works for evaluating e-learning system. They include some benchmarks (Section 2.1), criteria (Section 2.2), toolkits (Section 2.3), and standards (Section 2.4). These are a source of inspiration for our evaluation methodology. The criteria and standards have been considered for our impact evaluation in proposing the questionnaires and the log analysis. We consider benchmarks to be very important and we want our results to refer to existing benchmarks and also to set some benchmarks in our domain. Finally the toolkits were considered as a model of how the evaluation section of the platform will be developed.

2.1 Benchmarking

Benchmarking is "the process of identifying the best practice in relation to products and processes, both within an industry and outside it, with the object of using this as a guide and reference point for improving the practice of one's own organization." [Oxford 02]

Benchmarking is a method of self-evaluation that is based on two different process ideas :

- referencing and comparing one thing with another;
- searching for and creating reference points or benchmarks and understanding the reasons why they are reference points. [Jackson 01]

Benchmarking has been used in e-learning in several projects like [Benchmarking 05], [UKHE 05] and documented in papers like [IHEP 00], [Jackson01] by institutions like the Institute for Higher Education Policy in the U.S.A. or The Academy for Higher Education in the U.K.

From the point of view of our project and considering the classification of benchmarking proposed by Jackson [Jackson 98] our benchmarking activities are **explicit** because they involve a deliberate and systematic analysis. Also there is a **collaborative and cooperative internal** process because there are many partners that are benchmarking a single internal product – the Cooper platform. The benchmarking is focused on **process** and on **inputs** analyzing the differences in the learning process according to the data collected from the actors. Finally the benchmarking process is both quantitative and qualitative because it's based both on **metrics** (site logs) and on **bureaucratic information** (questionnaires).

Many benchmarks have been identified for the e-learning domain. We consider that the following are the most important from the point of view of our project [IHEP 00]:

- Course materials promote collaboration among students – the importance of collaboration in technology enhanced learning was promoted by the IHEP benchmark since 2000.
- Courses are designed to require students to work in groups utilizing problem-solving activities in order to develop topic understanding – also it is considered that the students should work together using problem solving activities - one step to project oriented learning described in Cooper.
- Data on enrollment, costs, and successful/ innovative uses of technology are used to evaluate program effectiveness – this benchmark addresses the issue of evaluation using data from the learning platform in order to improve the process.

2.2 Evaluation criteria

Several types of evaluation have been identified according to [Broadbent 03] that can be applied to COOPER evaluation:

- Objective-oriented evaluation: did we accomplish our goals doing this product *in this way*?
    - We have to create evaluation forms based on the objectives of the project. For each objective, we must find indicators to measure the degree in which it was attained.
- Management-oriented evaluation: where do we go next?
- Customer-oriented evaluation: What did the program do?
    - We have to create forms for the focus groups and compare the difference between the ones that used our product and the ones that didn't
- Feedback from the partners. There are quality management systems that focus on the companies and evaluate the opinions of the persons involved about the way things are going. And this is also an indicator of quality

There are also some criteria used for evaluation of e-learning considered by Brandon Hall [Hall 04]. The commented criteria are presented next.

- **Content**

It is important that the program has to include all the relevant information that the users need. This feature describes the quality of the information also.

- **Instructional Design**

Is the course designed in such a way that users will actually learn?

- **Interactivity**

The program has to engage the user to interact with others or with the program.

- **Navigation**

This feature describes the navigation of the users through the application and its interface. It describes how accessible it is, how easy-to-learn is the application and if there are enough documentation, but not excessive to determine program options.

- **Motivational Components**

There are many solutions in the area of the e-learning, but the program has to bring some new features or something interesting like engaging the user through different elements to motivating him like humor, testing, adventure, unique content, surprise elements, game elements. The user must be captivated and always discover something new or interesting in the program.

- **Use of Media**

The new and modern solutions of e-learning have integrated solutions of sound, video, animation, etc. There are two aspects: the price of these solutions and excessive use of them has to be not annoying.

- **Evaluation**

This part is very important in the process of the e-learning and there are many methods to do the evaluation of the users: final exam, quizzes, etc.

- **Aesthetics**

The program (the way it looks) has to be attractive, customizing and not to disturb the user in the process of e-learning, but to help him.

- **Record Keeping**

A program for e-learning should store the data about users (personal data, social data, etc.), about their progress and about their knowledge. A course manager has to receive the data automatically and use it in the process of e-learning, for example to customize an online test or the interface or to provide statistics data about users.

2.3 Evaluation "Toolkits":

The Evaluation "Toolkits" are resources that help to evaluate a learning process through guides and examples

- LTDI Evaluation Cookbook:

  http://www.icbl.hw.ac.uk/ltdi/cookbook/contents.html

- Evaluating Learning Technology Guide

  http://www.elt.ac.uk/ELT%20documents/materials/evalguide.pdf

- Online Evaluation Toolkit

  http://www.ltss.bris.ac.uk/jcalt/

2.4 Standards / References

Some other e-learning standards and principles that might be considered are listed below [BEEP04]:

- *Five Pillars of Quality Online Education*. Sloan Consortium Report to the Nation by George Lorenzo and Janet Moore, November 2002. The most important criteria that were identified are "learning effectiveness, student satisfaction, faculty satisfaction, cost effectiveness, and completely barrier-free access" [BEEP04]. http://www.sloan-c.org/effective/pillarreport1.pdf
- Standards developed by the *Open and Distance Learning Quality* and divided into 6 main categories: outcomes, resources, support, selling, providers, collaborative provision - http://www.odlqc.org.uk/odlqc/standard.htm

- *Course Peer Review Process*. Weber State University (UT) – process that aims to improve the quality of courses through multiple layers of reviewing http://wsuonline.weber.edu/faculty/PeerReview.htm
- *Seven Principles of Effective Teaching*. Michigan Virtual University – this paper emphasizes the seven principles of face-to-face learning and how they apply in online learning. One of the first principles is "cooperation between students". http://technologysource.org/article/seven_principles_of_effective_teaching/

3 Evaluation in Cooper

3.1 Purpose and scope of the evaluation

COOPER evaluation aims to identify whether COOPER project fulfils its promises to "*develop a model-driven, extensible environment that supports individual and collective competency building in virtual teams, whose members are geographically dispersed, have different backgrounds and competencies, working together in projects to solve complex problems*." [Cooper 05] In this way, COOPER is intended to satisfy the requirements of supporting long-distance cooperation of teams of students working on complex projects in two learning environments: graduate (or post-graduate) **university studies** and **Company universities and company training**. Both suppose participation in focused projects (e.g., projects for masters or specialization courses or projects for launching of new products or technologies) developed by learners coming from different institutions and backgrounds, which are world-wide dispersed.

The evaluation takes into account that COOPER is focused upon specific functions of project centered learning, namely:

- Managing virtual teams of persons with heterogeneous backgrounds and competencies;
- Facilitating collaborative and cooperative teamwork processes by knowledge sharing and use of recommendation services;
- Supporting competency building and competency assessment through adequate pedagogical tools;
- Supporting the distant cooperation in a distributed virtual team through communication tools.

In this context, the evaluation poses three vital questions regarding how well adopters of the Cooper methodology will perform their work. The essential questions to be asked are concerned with:

(1) the ability of the teamwork process model to capture the requirements of the "learning by doing" paradigm,

(2) the ability of customizing processes for given users, capturing specific requirements, and especially

(3) what impact will be globally achieved by the Cooper approach in terms of overall improvement of the learning process.

The **scope** of the evaluation consists of the following results provided by COOPER:

- the **reference model** and the generic processes that cover pre-project, project, and post-project activities, which are essential to the success of teamwork experiences;
- interoperable and validated **pedagogical scenarios** (such as the Virtual Company) and assessment strategies that make use of the IMS Learning Design, IMS Question and Test Interoperability, and the OUNL-Cito model for assessment;
- tools to support knowledge co-construction, sharing and re-use, with an emphasis on **recommendation services** for collaborative and cooperative teamwork;
- synchronous communication tools (such as audio/video conference, application sharing, co-browsing, presence services and chat) and asynchronous tools (such as discussion forum and logs);
- the **software platform** in which these models, scenarios, strategies and tools are integrated.

The evaluation will check the extent to which the results delivered by the project will contribute to forming a protected, shared **COOPER environment,** which will be easily deployed over any University's or Company's Intranet.

3.2 Evaluation principles

Comply with specifications. The evaluation will be guided by the "Scenario and Requirements Analysis" document [Cooper 06a], which summarizes the set of requirements by focusing on the core problem considered by Cooper, namely supporting teams of students who learn by doing projects. This document is considered the starting point for conceptual and technical design, and for the implementation of the platform. Consequently, it is the reference document for the establishment of the evaluation objects and evaluation criteria in Cooper, as well.

Consider the use cases. The evaluation methodology takes into account the use cases in order to choose some indicators that reflect all the requirements described in these use cases

Use of benchmarks. The methodology of evaluation makes sure that some benchmarks are respected (for example the ones used for analyzing collaborative group performance as part of the educational online knowledge management systems [Semar 06]).

Use terms of reference. The evaluation is based on the comparison of the results obtained from groups of people who used other learning platforms and the ones that used Cooper by evaluating them according to established criteria. The questions/indicators should show the difference brought by the Cooper model.

Use successful models. The evaluation methodology is based on the thorough analysis of models and tools used in other projects. Anyway, for a correct and accurate treatment of the environment that Cooper aims to realize, only issues related to collaborative activities and project-based learning will be considered.

3.3 User roles

The user represents not only a core concept in the design of the COOPER environment, but also a key participant to the evaluation process. Obviously, the evaluation should take into account different user categories (roles), and properly tailor the evaluation metrics and instruments. User categories are specifically defined for the three case studies. Thus, for ASP, the user roles are: Student, Teacher, Guest, and Administrator. In ALaRI the users can be faculty, students, industrial and academic collaborators, sponsors. In CoWare, the user can be *development engineer* who can be used for the course organization, *training developers* that are in charge of training, and *application engineers* and *customers* that are the recipient of the course.

A comparative analysis of the roles in the three organizations (ASP, ALaRI and CoWare) has been made, in order to identify possible similarities and define a single set of roles for the purpose of COOPER evaluation. The main characteristics of the roles (access rights and teamwork processes in which users are involved) have been considered. As a result, we restricted the user roles to the following ones, which will be used in evaluation:

- student - are the training recipients
- teacher - are the training providers
- administrator - of the learning environment
- guest - interested to access public information.

These names are generic as every institution has its hierarchy of training recipients and training providers. The evaluation is applied to these four generic categories and can be further particularized according to each institution's hierarchy.

3.4 Evaluation objects, stakeholders and methods

The evaluation targets **three aspects** regarding the Cooper methodology and its adoption by users in performing their work, namely **concept evaluation, technical and implementation evaluation**, and **impact evaluation.**

3.4.1 Concept evaluation

Concept evaluation addresses the question whether the models used in the Cooper methodology are sufficient for a complete and accurate description of the domain of use cases. This would mean that:

a. the workflow model, the data model, the hypertext model with extensions, and the associated conceptual tools (defined in Copper WP1) **completely** support the design of the collaborative teamwork processes as required in D5.1, thus covering pre-project, project, and post-project activities that are common to the three case studies;

b. the models are **flexible** and allow extensions of the Cooper framework (in other words, the models easily support the addition of a process that was not included in the framework);

c. the methodology used in Cooper (which grounds the framework development on the use of correlated workflow, data, and hypertext models) **effectively** supports the propagation of the collaborative teamwork processes structure into the structure of the Cooper platform, making it more easy to understand and use;

d. the pedagogical model **includes all** the ingredients (team building, project assignment, project planning, student assessment, etc.) that can be combined in different scenarios to support learning by collaborative project development.

Stakeholders in the evaluation

The following stakeholders will be involved in concept evaluation:

- evaluators of the UPB team will conduct the evaluation, will process the results, and will suggest recommendations for possible improvements
- users (student, teacher, administrator) from ASP, ALaRI and COWare will participate to the evaluation
- developers will help identifying the possible improvements based on the evaluation results.

Concept evaluation criteria and methods

In the following, criteria and techniques used to respond to the above mentioned concerns (**a** to **d**) are presented. For each criterion, the corresponding metrics and the methods used in evaluation are specified.

- **Completeness**

Completeness evaluation responds to concerns **a** and **d**. It will consist of a qualitative, thorough analysis of the models (described in the deliverables D1.2 and D2.1) against the requirements (gathered in the deliverable D5.1), which will be performed by UPB (supported by developers), based on complete scenarios or on samples from the domain of use cases.

Completeness will also be evaluated with respect to new functional requirements and user needs that could come up during the application development and/or during the first experimental sessions.

- **Flexibility/extensibility**

In order to evaluate framework's extensibility (concern **b**), users will identify possible extensions to the actual teamwork processes, which could address, for example, other didactical scenarios and assessments. Developers will make the necessary extensions to the initial framework and will demonstrate the easy to add new functionalities that were not included in the general framework. The extensibility evaluation will be part of the development of the second version of the platforms.

Questions related to these metrics will be included in *questionnaires* distributed to users of the platforms implemented at ASP, ALaRI and COWare, and will be filled up in the post- project phase. An example of such a questionnaire is presented in the appendix A.

3.4.2 Technical and implementation evaluation

The technical and implementation evaluation addresses the question whether the model-driven approach can be technically realized and implemented within time and cost constraints.

The evaluation will measure the capacity of the technology to generate and install **customizable** learning environments that implement didactical scenarios and assessment strategies. Also, the **technical performance** of collaborative tools included in Cooper will be evaluated.

Stakeholders in the evaluation

The following stakeholders will participate in this evaluation:

- developers will provide the necessary training and will assist in the performance of the experiment, then will help identifying the possible improvements based on the evaluation results.
- users (authorized student, teacher, administrator) from ASP, ALaRI and COWare will perform the evaluation
- evaluators of the UPB team will conduct the evaluation, will process the results, and will suggest recommendations for possible improvements

Technical evaluation criteria and methods

In the following, criteria and techniques used to respond to the concerns of technical evaluation are presented. For each criterion, the corresponding metrics and the methods used in evaluation are specified. The technical evaluation will mostly be done by experts in the domain. Some aspects like the perception of quality of the VoIP can also be evaluated by the actors through questionnaires.

- **Customizability**

For customizability, experiments will be performed to evaluate the support for flexibility and adaptation in the model-driven design of the COOPER environment. In this respect, the project members (both the administrators and authorised students within teams) from ASP, ALaRI and COWare will customize the COOPER environment to fit the specific needs of their projects and the existing local infrastructure. The developers will provide the necessary training and will assist in the performance of the experiment. After the experiments, the participants will respond to *questionnaires*.

- **Performance**

The metrics bellow will be used to evaluate the performance of some components of the platform. The tools they relate to are also mentioned.

accuracy (recommender system)

precision (search engine) - the number of good results from the number of results returned

quality of the sound (in the VOIP system)

quality of the image (in the VOIP system)

- **Models' effectiveness**

This evaluation aims to determine if using the BPMN and WebML models has produced results that offer advantages over other technologies. The effectiveness of the models means that users will understand the structure of the platform easier and also they will learn it easier. The metrics bellow will be used to evaluate the effectiveness of Cooper models (concern **c in section 3.4.1**).

easy to understand

- the platform is well structured and the interface is easy to understand

easy to navigate

- the organization of menus and links facilitate the navigation in the platform during the collaborative and cooperative project development

quality of the interface design (optional)

- according to: Semiotics (if the meanings of the messages proposed are understandable by the users), Cognitive (the cognitive effort the user do when reading an intranet page and the understanding of the information architecture), and Graphics (the graphic design and layout) aspects, in order to identify which kind of problem it may occur and how it is possible to correct it.

Questions related to these metrics will be included in questionnaires distributed to users of the platforms implemented at ASP, ALaRI and COWare, and will be filled up in the post- project phase.

3.4.3 Impact evaluation

The impact evaluation addresses the changes that are brought about by implementing the Cooper methodology in terms of process, learning, cooperation, management support and effective use of technology. These changes refer to several issues, such as:

A. **Process improvement:** Measuring how the "new" process guided by the Cooper framework improves over an unstructured approach to learning-by-doing in projects. Specifically, the process includes a number of steps that might be omitted, especially in the pre- and post-project phases, and their impact should be measured in terms of their benefits vs. their additional overhead.

B. **Learning improvement:** Measuring the improvement in the individual or team outcome, in terms of the competencies, according to competency-specific metrics. In this setting, the effectiveness of an adaptive recommendation systems taking into account individual needs will also be measured.

C. **Cooperation improvement:** The improvement in the "social" behaviour of the group, and the extent to which such social behaviour has been consolidated due to the use of the Cooper framework.

D. **Effective project management support:** The improvement in the quality, timeliness, and effective storage/retrieval of all the documents delivered throughout the project, and the ability to include them in a long-lasting repository.

E. **Effective use of the technology:** The actual accessibility of user interfaces, the ease of adaptation of the platform to user needs, the degree of acceptance of technology to the users and groups, depending e.g. on their age/role/qualification.

Stakeholders in the evaluation

The following stakeholders will participate in this evaluation:

- evaluators of the UPB team will conduct the evaluation, will process the results, and will suggest recommendations for possible improvements
- users (authorized student, teacher, administrator) from ASP, ALaRI and COWare will perform the evaluation
- developers will help identifying the possible improvements based on the evaluation results.

Instruments used for the impact evaluation

The main instruments and methods used for the evaluation are:

- site activity logs

- questionnaires for all the user categories mentioned above, distributed to all users at ASP, ALaRI and COWare

- social networks.

- results of evaluating the competences acquired by the learners

- expert opinions on the current state and the state after Cooper implementation.

These should allow obtaining information about all the features described in the use cases and in the process specifications. This is why, in the sequel, we organize the information about evaluation indicators according to the addressed feature. Also these tools will measure similar indicators in different ways in order to confirm the results obtained. For example the results obtained from questionnaires on the usefulness of a tool will be validated with the help of the site logs that will measure the usage of the tool. An example of a questionnaire for the students is presented in Appendix 1 while the requirements for log analysis are presented in Appendix 2.

Impact evaluation criteria and methods

In the following, criteria and techniques used to respond to the five concerns (A to E) mentioned above are presented. For each criterion, the corresponding metrics and the methods used in evaluation are specified. Generally the criteria given for the collaborative tools and also the ones for the social networks are used to evaluate the cooperation improvement. The evolution of these indicators in time and also the study of the networks that were formed together with the actors' answers to questionnaires will allow this concern to be evaluated. Effective use of technology is evaluated most through questionnaires but also with the help of the logs where actions performed by the users in every module are logged and analysed following the given criteria. The project management support will be evaluated based on the logs of the document repository as well as on the answers of the students and teachers in this matter. The process improvement will be considered after analyzing the results of the recommendation systems and

of the social networks at several key moments during the project as we think that these reflect best the improvements appeared in every step of the process. From the point of view of learning improvement this will mostly be evaluated by the teachers (tutors) and we'll obtain these evaluation through questionnaires.

I. Collaborative tools

- Chat **usage** and **quality**

  number of utterances in chats

  number of participants in chats,

  average number of participants/ conference

  user satisfaction regarding the chat system

The metrics a)-c) are extracted from logs while d) is extracted from questionnaires. These metrics measure the degree of use and the quality of the tool. The results obtained from metrics a) to c) should be compared with the ones obtained for the other communication tools. Also the results from a)-c) will be validated against the results in the questionnaires (example if 80% of the students declare in the questionnaire the chat is very useful but from the a,b indicators only 20% used it for more than 1 conversation then the indicator is not relevant and accurate). Also their evolution in time (along the learning sessions) should be analyzed.

- Forums **usage** and **quality**

  a) average posts per day,

  b) number of posts per day

  c) number of participants per thread

  d) number of posts per thread

  e) degree of use of the forum

  f) satisfaction regarding the forum usefulness

The first metrics (from a) to d)) are obtained from logs and they measure the usage of the forum while metrics e) and f) are extracted from questionnaires and they indicate both the degree of use and the satisfaction of the user. Also, an evaluation of the messages' content can be done: an expert evaluator or forum moderator could identify the "off topic" and "on topic" posts on the forum and make a qualitative evaluation.

- VoIP **usage** and **quality**

  a) number of users / session

  b) average time/session

  c) ease of use

  d) usefulness of the VoIP system

  e) quality of voice

  f) quality of video

The metrics a) and b) are extracted from logs while metrics c) to f) are extracted from questionnaires. The quality of voice (metric e) can also be measured on recorded meetings using the PESQ algorithm. The metrics reflect also the degree of use and quality of the tool. Comparisons between the values for the 3 collaborative tools will indicate the preference of the users for one of them. Also the average time spent on VoIP should be logged. If the evaluation is done on a longer period of time, it can be used to show the changes in the preferences of the actors.

- Application sharing **usage** and **quality**

    a) number of sessions

    b) number of participants /session

    c) satisfaction of the users

The degree of use for the application sharing is measured through logs (metrics a and b) and the quality through questionnaires.

- Scheduled / recorded meetings **usage** and **usefulness**

    percentage of scheduled / recorded meetings from the total number of meetings

    the communication tools used in scheduled / recorded meetings

    usefulness of scheduled meetings

    usefulness of saving and recording meetings

The first metrics (a) and b)) are extracted from logs while the others are calculated based on the questionnaires. The questionnaires aim to discover if the tools provided (scheduling meetings and saving meetings) have been useful for the actors. The log analysis will check if the data extracted from questionnaires are compliant with collected data. Also the data from the logs might show (through a) metric) the importance of the collaboration tools to the students, because we can consider that they record the most important meetings.

II. Project repository

- Virtual folders **usage** and **usefulness**

    a) percentage of teams that used them

    b) number of accesses to documents

    c) percentage of users that configured virtual folders

    d) usefulness of virtual folders

Metrics a) to c) are extracted from logs and metric d) is a questionnaire metric. These metrics will demonstrate the usefulness of this tool.

- Tagging documents **usage** and **usefulness**

    a) percentage of documents tagged

    b) usefulness of the tagged documents

The first metric is obtained from logs and the second from questionnaires. They both measure the usefulness of the tool and they are supposed to confirm one another.

- Rating documents **usage** and **usefulness**

    a) percentage of documents rated

    b) percentage of users that rated documents

    c) number of visits to documents with higher ratings vs. number of views of documents with lower ratings

    d) usefulness of rating documents

Metrics a) to c) are obtained from logs and metric d) is based on questionnaires. They should help evaluate if the tool has been used by students and also to establish (eventually with the help of the social networks) differences between people who used it and people who didn't use it. Also the usefulness can be seen from the number of visits to high-rate documents

- Search engine **usage**

    a) number of searches / user

    b) average number of searches per session

III. Recommender systems

- document recommendation **usage** and **quality**

    a) average number of clicks (on recommendation links) per student

    b) ratings of the recommendations

    c) satisfaction of the user on document recommendation

The metrics a) and b) are obtained from logs while c) is obtained from questionnaires. Metric a) shows the degree of use of the recommendation system and can be compared with the average number of clicks on non recommended documents, in order to see how much the ratings were used versus the usage of non-recommended documents. The rating of the recommendations shows the quality of the system. User satisfaction measured by questionnaires show how the user perceived the quality of recommendations.

- peer recommendation **usage** and **quality**

    a) number of recommendations of peers

    b) number of communications with the recommended peers

    c) satisfaction of the user on the peer recommendations

    d) confidence in the suggestions of the peer

The metrics a) and b) are extracted from logs and show the usage and quality of the system. A regular communication with a recommended peer means that the recommendation was good. Also recurrent requests for recommendations will show that the system is used and trusted. The metrics c) and d) are obtained from questionnaires and are complementary to the log metrics.

IV. Assessment and assignments

- Assessment tools **usage** and **accuracy**

a) Percentage of assessment tools used

b) Accuracy of the evaluation

The metric a) is obtained from logs while b) is obtained from questionnaires. Their purpose is to evaluate the way the assessment of the students has been done and by what means and how the students perceive the accuracy of their evaluation.

- Assignments **usage** and **quality**

a) number of questions asked on the posted projects
b) feedback offered
c) response rate
d) number of questions/users
e) how fast did the answers come.
f) The degree of satisfaction of the students with their project

The metrics a) to e) are obtained from logs while e) and f) are obtained from questionnaires. The purpose of these metrics is to evaluate the way the students dealt with the assignments and if they were satisfied with them and if their questions about their assignments were answered in a fast and useful way.

V. Social network metrics

Social networks represent a tool for the evaluation of the **characteristics of team collaboration** and communication. The metrics that are considered to be calculated for this type of evaluation are presented bellow, together with a short explanation of what we want to evaluate using them. The social network will be constructed based upon the collaborative activities done by the actors. For example if two actors perform an activity together (like a chat or a conference) a link will be constructed between them in the social network. The data about the activities done will be obtained from the database logs. From these logs we'll construct the social network which will actually be a graph. The nodes of the graph will be the actors of the learning platform while the edges will be the collaborative activities performed by the actors and obtained from the logs' analysis. On this graph several indicators will be calculated following the definitions that are given below.

- centrality

Centrality of a node is defined as the total number of ties to other actors of the network [Anklam 03]. The centrality measures the contribution of the teachers/tutors to the learning process and also can show the "key" (most active) students. The students who are "important" in the social network also have good results, which can prove the effectiveness of the collaborative tools

- centrality eigenvector

This is similar with centrality except the fact that it also considers the importance of the neighbors, similar with the page rank in Google. It assigns relative ranks to all nodes in the network based on the principle that connections to nodes having a high rank contribute more to the rank of the node in question [Wikipedia06]. This metric can refine the previous one since it

offers a more accurate method of finding "key" students or to identify important persons (like teachers for example).

- centralization

“A centralized network will have many links concentrated around one or a few nodes, while a decentralized network is one in which there is little variation between the numbers of links each node possesses” [Wikipedia06]. A centralized social network shows that project teams don't cooperate with each other, the links in the network being concentrated on teachers. So, this metric helps detecting communication patterns or problems in the use of tools.

- cohesion

Cohesion refers to the degree actors are connected directly to each other by cohesive bonds [Wikipedia06]. Groups are identified as ‘cliques’ which are maximum complete sub-graphs [Reffay 03]. This metric will help establish if the teams were well formed and if the team members worked well together

- radiality

Radiality is “the degree an individual’s network (formed by individuals he knows) reaches out into the network and provides novel information and influence” [Valente 98] This can be used to evaluate the degree of teacher involvement in the communication process

- network reach

This represents the set of actors that are in the neighbourhood of a node (2 or maximum 3 steps from the node) [Albert 02]. This metric can be used to measure if the students get information directly from key actors in the network (teachers, project managers) or they indirectly get information from peers.

- density

Density describes the general level of cohesion in a social network or the number of different people the actor interacts with [Anklam 03]. A big density will show that the communication tools were intensively used among team members, and also among members belonging to different teams. We can measure here the density for the team-level social networks, which should be maximum or else could indicate a communication problem. Also we can measure the density of the social network obtained from all the users of the platform. If the density drops then it suggests that people didn't communicate outside their teams.

- structural equivalence

This metric determines which nodes play similar roles in the network [Lorraine 71] – can be used to measure if persons with similar roles behaved similarly (if all teachers used the collaborative tools in the same way) Also this shows differences between same people using different communication tools or differences between communication strategies of different teams on similar projects.

3.5 The Evaluation Procedure

The evaluation will be performed during training session(s) in ASP, ALaRI and COWare, and will be based on comparing the activity of "test" groups of participants, who will use the new

Cooper platforms, with "control" groups of participants, who will use the old platforms in the institutions.

In each institution, the evaluation will follow the **reference model** and will use the knowledge and information gathered in the normal learning process. The profiling of individuals and projects, teams building, and planning of activities will correspond to a normal **pre-project** phase.

The only difference in this phase will consist in designating the "test" groups and the "control" groups. Learners of the "test" group will be eventually trained on using the Cooper platform.

The **project development** phase will also follow the normal "flow", based on pedagogical scenarios and assessment strategies appropriate for such project-centred learning, and using the generalized recommender based approaches for supporting knowledge co-construction, sharing and re-use**.** During the project phase, **site activity logs** such as those associated with the Project Management Repository, recommendation system, and the synchronous / asynchronous collaboration tools will be updated. Also, information about learners' progress for assessment according to the educational model for assessments will be registered.

At the end of the training, the standard activities involved in the **post-project** phase will take place.

In addition, for evaluation purposes, the **results of the assessments** for different project groups together with the users' feedback (**questionnaires**) concerning the use of Cooper platform will be collected by evaluators for further processing.

After a few months (3 - 6) learners will pass a new test for the evaluation of the acquired competences and of the changes in attitude towards e-learning.

Repeat the evaluation procedure for the second version of the platform.

Concerning **expert** evaluation, it will determine if changing is made in the direction that the partners strive for, and if this is attributable to Cooper.

Evaluating application dependant usability aspects

It is important to distinguish the evaluation analysis into application independent / and application dependant usability aspects.

The application independent usability aspects regard features such as navigation quality, content accuracy, consistency, interface cognitivity, accessibility, graphic and layout quality, etc.

Application dependent usability aspects will be done also through user testing with questionnaire and interviews and log analysis. In order to use these instruments to evaluate application dependant aspects scenarios of use will be created.

These scenarios should be selected on the basis of relevance to application’s mission and depending on the type of stakeholder, they should be the most frequently used, and concern also usability problems detected during the inspection phase.

These scenarios of use should be developed by the stakeholders and also the stakeholders will have the possibility to tune the instruments offered by the evaluation in order to evaluate these scenarios. For example according to a scenario some questions can be added to a standard questionnaire. These **customized questionnaires** can be prepared, according to specific processes, so they could be **more problem-driven**, in order to identify possible technical or communication inefficiencies or obstacles to the correct use of the learning platform, and consequently to the virtual interaction among the team members, and in general among the users themselves.

For instance, for the ALaRI user (student – teacher – industrial collaborator) it would be useful to prepare specific questionnaire concerning the actual e-learning platform (developed with WebRatio tool), and assigning the user specific tasks and activities to perform.

In this way, with problem-driven questionnaires, it could be possible also to identify with major accuracy the type of problem.

Gathering data

Quantitative data are gathered in logs and also in questions for which the respondents provide direct quotations along a predefined scale. .

Qualitative data can be gathered through open-ended questions. For some evaluation criteria, qualitative data can highlight some drawbacks of the site functionality (for example, very large delays in responding to user queries or failures observed during the platform use) or, on the contrary, some aspects the users found to be very useful and very high quality.

Analyzing data

The gathered data will be used to complete the evaluation. The approach may be based on a particular model, such as grounded theory, statistical analysis or narrative case study based on a combination of quantitative and qualitative methods taking into consideration what each can and cannot do.

The validity of the gathered data and conclusions should be verified by reviewing them with the designated target audience, with peers or other stakeholders and seeking examples that might contradict the conclusions. Also data will be validated through comparison between similar indicators obtained with different tools (site logs, questionnaires, social networks).

Integrate the evaluation in Cooper.

The selection of evaluation instruments for data analysis will be done as part of the task "T5.4: Select the evaluation areas and instruments."

Evaluation instruments for logs analysis, questionnaires processing, and social networks construction and analysis will be elaborated and integrated in the Cooper platform to be used by stakeholders according to the Cooper reference model.

4 References:

[Albert 02] Albert, R., Barabasi, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74 (2002) 47–97

[An 04] An, Luiza and Restrepo, L. G. “An experience in the evaluation of e-learning for IT Training and certification”, retrieved from the Web http://luisguillermo.com/CAITA2004.pdf on 20.01.2006

[Anklam 03] Anklam, P. KM and the social network. In Inside Knowledge, Vol. 6, Issue 8. (2003)

[Armstrong 00] Armstrong, M. "Strategic Human Resources management", Kogan Page 2000, ISBN: 0749433310

[Baumgartner 96] Baumgartner, P. "Evaluation vernetztes Lernens", http://www.avmz.unisiegen. de/extern/GMW/events/VC96/prg/abs/2_7.html

[BEEP 04] Best Educational e-Practices "E-Learning Evaluation: Standards and Procedures", Project Eagle, St. Petersburg College, Sept. 2004

[Benchmarking 05] 2005 E-learning Benchmarking Project – Survey report – http://e-learningindicators.flexiblelearning.net.au/case_studies/media/ACE_report_A.pdf

[Bonk 02] Bonk, C. “E-Learning Evaluation Measures”, Indiana University and courseShare.com April 17, 2002, retrieved from the Web on 15.01.2006, www.trainingshare.com/download/train2002/eval_methods.doc

[Broadbent 03] Broadbent, B. and Cotter, C. “Evaluating e-learning”, Appears in the 2003 Pfeiffer Annual: Training, retrieved from the Web on 15.01.2006, www.evaluate-europe.net/Members/jilam/Broadbent2003

[Cooper 05] Cooper Consortium Collaborative Open Environment for Project-Centred Learning. Annex I – “Description of Work”, Contract No. 027073, November 2005

[Cooper 06a] Cooper Consortium "Scenario and Requirements Analysis. Cooper Technical Report D5.1", March 2006.

[Cooper 06b] Cooper Consortium "Teamwork Processes Assessment and Methodology. Cooper Technical Report D1.1", May 2006.

[Cooper 06c] Cooper Consortium "Specification and design of process modelling concepts and of VOI units in WebRatio. Cooper Technical Report D4.1", May 2006.

[Cooper 06d] Cooper Consortium "State of the Art Report in Knowledge Sharing and Recommendation. Cooper Technical Report D3.1", May 2006.

[Cristea 05] Cristea, Alexandra et al "Evaluation of adaptive hypermedia systems' conversion", in Proceedings of the sixteenth ACM conference on Hypertext and hypermedia, Salzburg, Austria

[Cullen 93] Cullen, J., Kelleher, J. and Stern, E "Evaluation in DELTA", Journal of Computer Assisted Learning. 9: 115-126.

[Dillenbourg 99] Dillenbourg, P. "What do you mean by collaborative learning?", In P. Dillenbourg (Ed) Collaborative-learning: Cognitive and Computational Approaches. (pp.1-19). Oxford: Elsevier

[Frydenberg 02] Frydenberg, Jia "Quality Standards in eLearning: A Matrix of Analysis", Irvine Distance Learning Center, University of California, retrieved from the Web http://www.irrodl.org/content/v3.2/frydenberg.html on 18.01.2006

[Hall 04] Hall, B. "FAQs about E-Learning", retrieved from the Web http://www.brandon-hall.com on 20.11.2005

[Henri 92] Henri, F. "Computer conferencing and content analysis", In A.R. Kaye, ed. Collaborative learning through computer conferencing: the Najaden papers, 115-36. NY: Springer.

[Holst 00] Holst, Shirley "Evaluation of Collaborative Virtual Learning Environments: The State of the Art", http://www.ipsi.fraunhofer.de/~publications/concert/2000/Evaluation.pdf

[IHEP00] The Institute for Higher Education Policy "Quality on the line", retrieved from the Web http://www.ihep.com/Pubs/PDF/Quality.pdf on 11.01.2006

[JISC 04] The Joint Information Systems Committee "Evaluation Strategy for E-Learning Programme", retrieved from the Web on 20.01.2006, http://www.jisc.ac.uk/uploaded_documents/Apx E-learning-evaluation-strategy 2.doc

[Jackson 01] N. Jackson "Benchmarking in UK HE: an overview" in Quality Assurance in Education Journal, December 2001

[Jackson 98] N. Jackson 'Introduction to benchmarking assessment practice'. In *Pilot Studies in Benchmarking Assessment Practice in UK Higher Education*. Quality Assurance Agency for Higher Education.

[Lorraine 71] Lorraine, F., White, H.C.: Structural equivalence of individuals in social networks. Journal of Mathematical Sociology 1. (1971) 49-80

[Oxford 02] "benchmarking" *A dictionary of business*. Oxford: Oxford University Press. Oxford Reference Online. BLC Boston College Libraries. Retrieved June 8, 2005.

[Pfister 99] Pfister, H.R and Wessner, M. "Evaluation von CSCL-Umgebungen", To appear in: J. Wedekind (Ed.): Virtueller Campus '99, Medien in der Wissenschaft. Band 9, Münster u.a., Waxmann Verlag.

[Reffay 03] Reffay, C., Chanier, T.: How social network analysis can help to measure cohesion in collaborative distance-learning. Proceeding of the Computer Supported Collaborative Learning Conference (2003)

[Semar 06] Semar, W. Evaluation of a benchmark system for analyzing collaborative group performance as part of an educational online knowledge management system, in Arabnia, Hamid; et al. (Hg.): Proceedings of the International Conference on Information and Knowledge Engineering - IKE'06. Las Vegas: CSREA Press, 2006, retrieved from the web http://www.inf-wiss.uni-konstanz.de/People/WS/IKE06-cc.pdf

[Shepherd 99] Shepherd, C. "Evaluating online learning", http://www.fastrak-consulting.co.uk/tactix/Features/evaluate/evaluate.htm

[Stone 03] Deborah L. Stone, Steven W. Villachica, “Web-Based E-learning Evaluation, Levels One to Five and Beyond”, Presented at the 2003 VNU Training Conference, Atlanta, GA, DLS Group, Inc. Denver, CO, retrieved from the Web on 17.01.2006, www.dls.com/1123-99sec-slides.pdf

[UKHE 05] UK Higher Education Academy Benchmarking and Pathfinder programme http://www.heacademy.ac.uk/benchmarking.htm

[Valente 98] Valente, T.W., Foreman, R.K.: Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. Social Networks, 20(1). (1998) 89-105

[Warwick] Centre for Academic Practice, University of Warwick “Evaluating E-Learning”, retrieved from the Web http://www.warwick.ac.uk/go/cap/resources/eguides/ on 19.01.2006

Appendix 1

Questionnaire for students

This is a **draft** of the Questionnaire for students.

Questions are grouped according to several evaluation **criteria**.

The **final form** of the questionnaire will also include an introductory section (short guidelines for filling in the questionnaire) and a concluding section (what should be done after the completion of the questionnaire). Additional explanations will be available before each group of questions.

Students will answer the questionnaire after completing the project-based phase of learning.

The COOPER platform

Criteria: ease of use, learneativity

5 How attractive is the interface?

○ 1 (not at all attractive) ○ 2 ○ 3 ○ 4 (very attractive)

6 Is the interface easy to understand?

○ 1 (very difficult) ○ 2 ○ 3 ○ 4 (very easy)

7 Do you think additional training is needed for using the platform?

□Yes, couldn't use the system without additional training

□No, all the information needed for using it is available onsite

8 Was the training (if provided) sufficient?

○ 1 (not sufficient) ○ 2 ○ 3 ○ 4 (sufficient)

9 The help for using the platform was:

○ 1 (not useful) ○ 2 ○ 3 ○ 4 (very useful)

10 Does the organization of menus and links facilitate your navigation in the platform?

○ 1 (not at all – they are confusing) ○ 2 ○ 3 ○ 4 (yes – they are clear and helpful)

11 How would you rate the quality of the graphic interface design:

○ 1 (very low) ○ 2 ○ 3 ○ 4 (very high)

12 The search engine was a helpful tool for finding information on the Cooper platform:
○ 1 (not at all helpful) ○ 2 ○ 3 ○ 4 (very helpful)

13 How easy it was to organize your documents into virtual folders?
○ 1 (very difficult) ○ 2 ○ 3 ○ 4 (very easy)

14 What communication facility did you use more frequently during the project work (chose one):

□Chat

□Forum

□VoIP

□ Others

15 How easy to use were the communication tools?

a. VoIP
○ 1 (very difficult) ○ 2 ○ 3 ○ 4 (very easy)

b. Chat
○ 1 (very difficult) ○ 2 ○ 3 ○ 4 (very easy)

c. Forum
○ 1 (very difficult) ○ 2 ○ 3 ○ 4 (very easy)

16 How easy was the use of the tools for meetings scheduling?
○ 1 (very difficult) ○ 2 ○ 3 ○ 4 (very easy)

Recommendation system

Criteria: degree of use, usefulness, quality

17 How often did you use the document recommendation system?
○ 1 (never) ○ 2 ○ 3 ○ 4 (every time I needed some information for my work)

18 How accurate were the documents provided by the recommendation system ?

○ 1 (not at all accurate) ○ 2 ○ 3 ○ 4 (very accurate)

19 How useful was the information for your work?

○ 1 (not at all useful) ○ 2 ○ 3 ○ 4 (very useful)

20 How often did you use the peer recommendation system?

○ 1 (never) ○ 2 (seldom) ○ 3 (often) ○ 4 (every time I needed some new information for my work)

21 How accurate were the results provided by the peer recommendation system?

○ 1 (not at all accurate) ○ 2 ○ 3 ○ 4 (very accurate)

22 How useful were the peer recommendations for your work?

○ 1 (not at all useful) ○ 2 ○ 3 ○ 4 (very useful)

23 Did the peer provide relevant answers?

○ 1 (never) ○ 2 (in few cases) ○ 3 (in many cases) ○ 4 (always)

24 Did the peers' answers came in an acceptable amount of time?

○ 1 (never) ○ 2 (in few cases) ○ 3 (in many cases) ○ 4 (always)

Communication tools: chat, forum, VOIP

Criteria: degree of use, usefulness, quality

25 In order to communicate with your team members, how often did you use:

a. chat

○ 1 (seldom) ○ 2 ○ 3 ○ 4 (very often)

b. forum

○ 1 (seldom) ○ 2 ○ 3 ○ 4 (very often)

c. VoIP

○ 1 (seldom) ○ 2 ○ 3 ○ 4 (very often)

26 In order to communicate with your teachers, how often did you use:

d. chat

○ 1 (seldom) ○ 2 ○ 3 ○ 4 (very often)

e. forum

○ 1 (seldom) ○ 2 ○ 3 ○ 4 (very often)

f. VoIP

○ 1 (seldom) ○ 2 ○ 3 ○ 4 (very often)

27 Rate how well these communication tools support collaboration among learners:

a. chat

○ 1 (very badly) ○ 2 ○ 3 ○ 4 (very well)

b. forum

○ 1 (very badly) ○ 2 ○ 3 ○ 4 (very well)

c. VoIP

○ 1 (very badly) ○ 2 ○ 3 ○ 4 (very well)

28 Rate how well these communication tools support collaboration between learners and teachers:

a. chat

○ 1 (very badly) ○ 2 ○ 3 ○ 4 (very well)

b. forum

○ 1 (very badly) ○ 2 ○ 3 ○ 4 (very well)

c. VoIP

○ 1 (very badly) ○ 2 ○ 3 ○ 4 (very well)

29 The quality of the sound in the VOIP system was:

○ 1 (very bad) ○ 2 ○ 3 ○ 4 (very good)

30 The quality of the image in the VOIP system was:

○ 1 (very bad) ○ 2 ○ 3 ○ 4 (very good)

31 How useful for completing your project work were:

a. chat system

○ 1 (not at all useful) ○ 2 ○ 3 ○ 4 (very useful)

b. forum

○ 1 (not at all useful) ○ 2 ○ 3 ○ 4 (very useful)

c. VoIP

○ 1 (not at all useful) ○ 2 ○ 3 ○ 4 (very useful)

Scheduling tools

Criteria: degree of use, usefulness, quality

32 How many meetings have been scheduled?

○ 1 (few) ○ 2 ○ 3 ○ 4 (many)

33 How many of the scheduled meetings were cancelled?

○ 1 (none) ○ 2 (few) ○ 3 (some) ○ 4 (all of them)

34 Were the logs of the meetings recorded?

○ 1 (never) ○ 2 ○ 3 ○ 4 (every time)

35 Were scheduled meetings a good way to improve the team work?

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (definitely, yes)

Virtual folders, tagging documents

Criteria: degree of use, usefulness, efficiency, quality

36 Saving information in virtual folders was an effective way to share information with other team members:

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (very effective)

37 Saving information in the virtual folders an effective way to organize your work:

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (very effective)

38 Organizing your documents in virtual folders was:

○ 1 (a very difficult task) ○ 2 ○ 3 ○ 4 (a very easy task)

39 Tagging documents was a useful way for time saving while working on your project.

○ 1 (strongly disagree) ○ 2 ○ 3 ○ 4 (strongly agree)

40 Tagging documents was a useful way to quickly find relevant information needed in my work.

○ 1 (strongly disagree) ○ 2 ○ 3 ○ 4 (strongly agree)

41 Documents' ratings were a valuable clue in selecting relevant information for my project.

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (quite a lot)

42 Documents' ratings were an accurate assessment of the quality of the information they contain:

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (very accurate)

Applications sharing (co-browsing, co-editing...)

Criteria: usefulness, efficiency

43 In your opinion, were the application sharing tools useful in supporting collaboration with your team members?

○ 1 (not useful at all) ○ 2 ○ 3 ○ 4 (very useful)

44 How would you rate the contribution of the application sharing tools to the efficiency of your team work?

○ 1 (not important at all) ○ 2 ○ 3 ○ 4 (very important, essential)

Assignments and student evaluation (assessment)

Criteria: students' awareness of the projects offered, students' satisfaction with team work

45 The topics of the projects proposed were of interest to you.

○ 1 (strongly disagree) ○ 2 ○ 3 ○ 4 (strongly agree)

46 How satisfied were you with the project you've been assigned to?

○ 1 (not at all satisfied) ○ 2 ○ 3 ○ 4 (very satisfied)

47 How many project objectives were your team able to fulfilled?

○ 1 (few) ○ 2 ○ 3 ○ 4 (all)

48 How would you rate the collaboration between your team members?

○ 1 (really bad) ○ 2 ○ 3 ○ 4 (very good, excellent)

49 In general, the communication with your team members was:

○ 1 (very poor) ○ 2 ○ 3 ○ 4 (very good)

50 Was there a person who distributed the tasks within your team?

□Yes, there was a team leader

□No, we agreed together about the distribution of the tasks

51 Was there a person usually mediating the conflicts within your team?

□Yes

□No

52 Were the criteria for project work evaluation clearly stated from the begging?

○ 1 (not at all clear) ○ 2 ○ 3 ○ 4 (very clear)

53 In your opinion, how accurate was the evaluation of your team work?

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (very accurate)

54 How accurate was the evaluation of your own work within the team you've been assigned to?

○ 1 (not at all) ○ 2 ○ 3 ○ 4 (very accurate)

Appendix 2

Log implementation requirements

This list is based on the requirements for the evaluation that were discussed with the partners during the project meetings and on the evaluation requirements draft that were also commented by the partners

In order to obtain all the indicators described in the main document (including the ones for the social networks) the data mentioned bellow should be logged in the database.

- Recommender systems: on creation of a recommendation, the following data will be logged in the database
    - id of the user
    - id of the page where the user is
    - id of the session
    - id of the recommended page / id of the recommended user
    - id of the recommendation
    - timestamp
    - the ratings of the recommendation.
- General platform logging – on loading every page, the following data will be inserted in the database
    - Id of the user
    - Id of the session
    - Id of the page/resource
    - Last performed action
    - Timestamp
- Chat
    - Chat conference (for each chat conference):
        - Id of the user, Id of the conference, Time when the user enters / leaves the chat
        - Id of the user, Id of the conference, Timestamp, mesasge text, for each sent message
    - Peer to peer chat
        - Id of the message sender
        - Id of the destination
        - Timestamp
        - Message
- Forum (for each thread):
    - Id of the thread
    - Id of the thread initiator
    - For every message posted in the thread
        - Id of the user who posted
        - Message
        - Rating of the post
        - Id of the message
        - Id of replied message (if the post is a reply to another message)
- VoIP (for each call):
    - Id of the call initator (when enters)
    - Id of the destination
    - Start Time of the call

    - End Time of the call
- Logs for other modules
    - Scheduled meeting (when the meeting is scheduled):
        - Id of the meeting
        - Id of user that schedules the meeting
        - Time of scheduling
        - Time of the planned meeting
    - Scheduled meeting (when the meeting takes place):
        - Id of the meeting
        - Ids of participants in the meeting
    - Virtual folders (for every action):
        - Id of the user
        - Action (create, edit, delete, etc.)
        - Id of resource viewed/modified
        - Timestamp
    - Tagging documents (for every action):
        - Id of the user
        - Id of the document
        - Tag
        - Timestamp
    - Rating documents (for every rating action):
        - Id of the user
        - Id of the document
        - Rating
        - Timestamp
    - Search engine for the document repository (for every query):
        - Id of the user
        - Search text
        - Id of the session
        - Timestamp