



HAL
open science

From Data Analysis to Design Patterns in a Collaborative Context

Vincent Barré, Christophe Choquet, Hassina El Kechaï

► **To cite this version:**

Vincent Barré, Christophe Choquet, Hassina El Kechaï. From Data Analysis to Design Patterns in a Collaborative Context. *Journal of Interacting Learning Research*, 2007, Special Issue Usage Analysis in Learning Systems: Existing Approaches and Scientific Issues, 18 (2), pp.285–308. hal-01434072

HAL Id: hal-01434072

<https://hal.science/hal-01434072>

Submitted on 22 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From data analysis to Design Patterns in a collaborative context

Vincent BARRE, Christophe CHOQUET, Hassina EL-KECHAÏ

LIUM / IUT de Laval

52, rue des docteurs Calmette et Guérin
53020 LAVAL Cedex 09 (France)

Abstract. The underlying aim of the work related in this paper, was to define Design Patterns for recording and analyzing usage in learning systems. The implied "bottom-up" approach when defining a Design Pattern brought us to examine data collected in our learning system through different lights: (1) the data type, (2) the human roles involved in the production of the data, or interested by their uses, and (3) the nature of the data analysis. This method has allowed us to have a global view on the data, which can be easily generalized and formalized.

Keywords. Re-engineering, Data analysis, Design patterns

1. Introduction

The desynchronization between the design and the uses in distance education, penalizes the iterative optimization of the system's quality by not taking into account uses with a reengineering objective. That's why in (Corbière, & Choquet, 2004a) we have proposed a meta-architecture model which explicitly integrates a step dealing with the observation and the compartment analysis of distance learning systems and learning process actors in an iterative process, guided by design intentions. We underline, in particular, the need for a formal description of the design point of view of the scenario, called the prescriptive scenario, as well as the assistance in uses analysis by allowing the comparison of descriptive scenarios (an a posteriori scenario that effectively describes the learning situation's sequence (Corbière, & Choquet, 2004b; Lejeune, & Pernin, 2004)) with the predictive scenario. This produces information, significant for designers from a pedagogical point of view, when they perform a retro-conception or a reengineering (Chikofsky, & Cross, 1990) of their systems. In the

framework of REDiM (French acronym for Reengineering Driven by Models of e-Learning Systems) project, we are particularly interested in supporting the implementation of designers two main roles: (i) to establish the predictive scenario of a given learning situation, and (ii) to anticipate descriptive scenario construction by defining situation observation needs allowing the effective evaluation of the learners' activity. Thus, describing in details and formalizing the data recorded during a learning session, in order to build significant indicators which qualify the activity, is a crucial step for us. Moreover, such an analysis expertise could be capitalized in a pattern, which could be used in the design of an another TEL system.

In this paper, we will focus on a particular collaborative e-learning system named *Symba*. More precisely, we will observe the effective use of a pedagogical scenario in the context of a collective activity supported by collaborative tools. Our experiment thus consists of a project management collective activity, and more specifically, of a web-project management activity (specification and implementation of a web site). From our pedagogical reengineering viewpoint, considerable interesting information can arise from this experiment. In particular, we are interested in comparing descriptive scenarios with predictive ones. Moreover, in a collaborative context, another interesting advisability is to compare roles emerging from the activity to those anticipated by designers. In our experiment, and in accordance with a normalization context, we have used pedagogical model arising from IMS consortium Learning Design (Koper, Olivier, & Anderson, 2003) in order to describe learning activities and to explicit pedagogical scenarios. Nevertheless, we only use IMS LD as a mean for designers to express their intentions, and not in an implementation perspective.

The underlying aim of the work related in this paper, was to define Design Patterns for recording and analyzing usage in learning systems, in the framework of the DPULS project (DPULS, 2005). The implied "bottom-up" approach when defining a Design Pattern brought us to examine our experience's data through different lights: (1) the data type, (2) the human roles involved in the production of the data, or interested by their uses, and (3) the nature of the data

analysis. This method has allowed us to have a global view on the data, which can be easily generalized and formalized.

We are thus within the framework depicted in Figure 1. We have set up an experiment (*Symba*) to produce data that will be described and analyzed so as to produce indicators (feature of a data highlighting its connection with an envisaged event having a pedagogical significance (DPULS, 2005)). Once formalized, the analyses will then lead to Design Patterns.

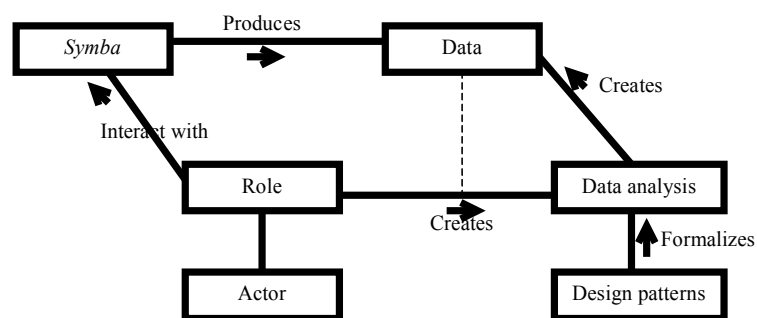


Figure 1. Focus on some aspects of our methodology

Thus, according to the framework depicted in Figure 1 and after a short presentation of *Symba* experiment, we will focus on the three main viewpoints leading towards design pattern elaboration. More precisely, we will first have to ask ourselves about the formal definition of roles endorsed by actors of the system and to specify the motivation that each role has towards data analysis. In (DPULS, 2005) we define a role as a set of tasks performed by one or more human or artificial agent according to specific needs and competencies, e.g. designer, learner, teacher... The second viewpoint will allow us to clarify what kind of data will be analyzed, that is, to formalize and classify each data that will be manipulated by actors in many roles in our experiment. The last viewpoint, which will lead to the formalization of design pattern, will focus on data analysis to distinguish who analyzes data, from who uses the analysis results.

Then, those three viewpoints will lead to the formalization of two kinds of design patterns. One kind will be used by analysts in order to produce new data (that is, will represent a formalization of their analysis knowledge), whereas the other kind will be used for “high level” purposes, that is for engineering, reengineering or regulation purposes, and those patterns will lead to the production of indicators.

2. Presentation of *Symba* experiment

We have used an experimental CSCL support environment called SYMBA (Betbeder, & Tchounikine, 2003). This environment is a Web-based system, developed by the LIUM laboratory in the framework of a Ph.D. study, in order to support Collective Activities in a Learning Context. It was designed following a double objective: (i) allowing students to explicitly work on their organization and (ii) providing tailorability (Morch, & Mehandjiev, 2000) features which let students decide which tools and resources they want to be accessible in order to achieve tasks they have defined.

With this system, students have to develop a dynamic web-site using previously taught web project management methodology. A predictive scenario of our experiment is depicted in Figure 2, we will now detail this scenario. According to our theoretical framework, students have first to work collectively (and agree) on the project organization (such as what to be done, who does what, when tasks have to be finished, which tools are necessary for a particular task...) before beginning the second part, which consists of collectively performing the tasks they have defined, according to their organization.

During all their activities, learners are self-managed and have to define (and collectively agree on) roles they will endorse and act in consequence. From this viewpoint, it will be interesting to take into account that roles have many meanings, and particularly a functional meaning (that is, related to an action, linked to people's status in an organization) and a socio-affective meaning (that is, related to the way people slip their personality into a functional role).

In concrete terms, the learners' activity was organized in five steps (see Figure 2), and for each one instructional designers set up a task model (see Figure 3) which was communicated to learners for the first three steps.

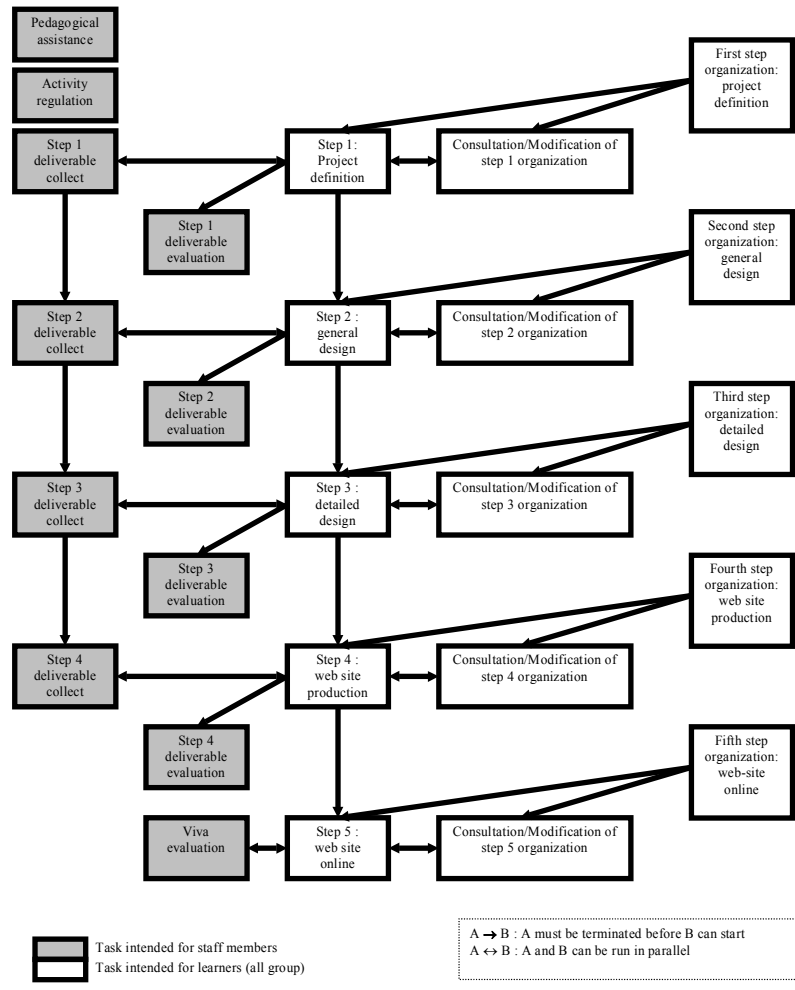


Figure 2. Predictive scenario of the web-project management

The formal description of what is a ‘correct’ project organization (that is, task model for each step) is formalized using IMS Learning Design (see data S-4.1 below). In order to explicit this task model, we present in Figure 3 what has to be done for the second step.

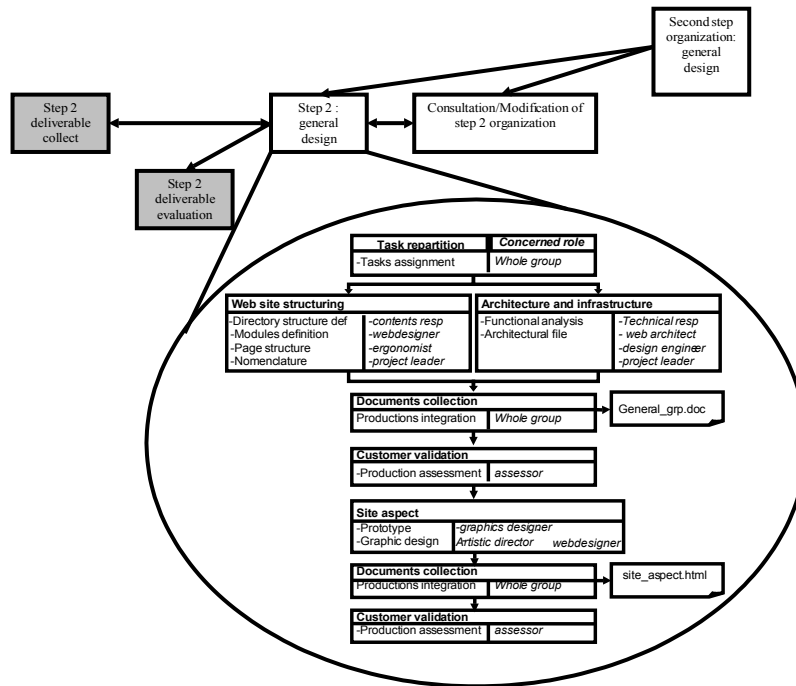


Figure 3. One step of students' predicted organization

3. Many roles, different motivations in data analysis

In our system, one actor can endorse many roles (for example, the teacher can be either 'instructional designer', 'assessor tutor' or 'observed uses analyst'). We think that it is very important to focus on role definition rather than on actor definition. We will, thus, began by presenting the different roles endorsed by actors of our system (see Figure 4), and, then, we will detail motivation that those roles have in data analysis.

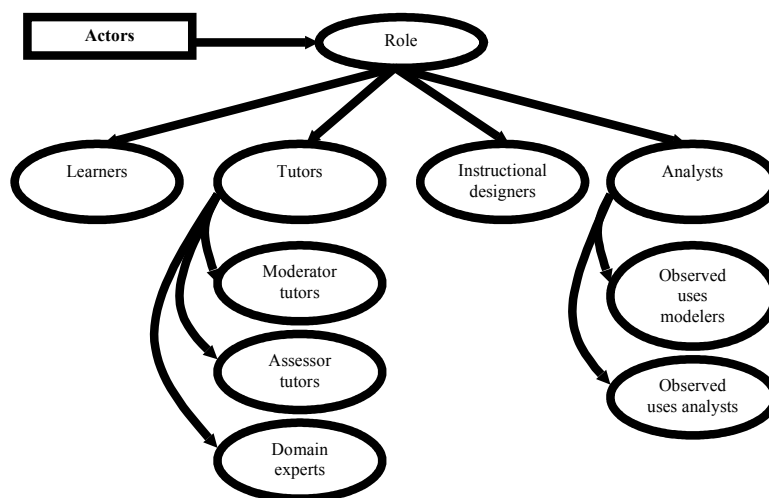


Figure 4. Four kinds of roles (with sub-roles) in our experiment

3.1 Presentation of roles

This experimental system is used by actors endorsing four distinct categories of roles (one actor can play more than one role in our experimental system). First category is made of fifty-six *learners* in higher education, from the Laval Institute of Technology, University of Maine (France). They were associated in small groups of five students and they worked either at the University or at home using tools offered by *Symba*. Those proposed tools are centered around the description, organization and perception of the activity, but learners must also use the environment in order to explicit the organization of their work, with a sharable plan and task editors. The activity proposed to the learners lasts for four weeks (35 working hours per week) and a predictive pedagogical scenario implying a collaborative learning was proposed, even if students are free to adopt or modify it. One can notice that this predictive scenario may involve concepts that have not yet been taught.

The second category is made up of three kinds of *tutors*. We have *moderator tutors* whose role is to monitor activity within the learning session and to fill in reports for evaluating tutors (i.e. *assessor tutors*) in charge of evaluating learners' activity. This measures knowledge they have acquired. Lastly, *domain experts* are in charge of assisting learners in their tasks by helping them to solve specific problems connected to their expertise domain.

A third category is made of *instructional designers*. They specify the predictive pedagogical scenario and the uses of the learning system to be observed, they also use results of the effective use of the Learning System analysis in order to improve it (reengineering process).

The last category is made of two kinds of *analysts*. *Observed uses modelers* are building tracks with collected raw data, either from the Learning system or not (that is, with collected observed uses), whereas *observed uses analysts* are analyzing the observed uses in order to synthesize information.

3.2 Different motivations in data analysis

In our experiment, people in many roles want to (and are interested in) analyze data. *Instructional designers* want to verify if the roles they have predicted are adopted by learners and to detect unforeseen new roles. They are also interested in understanding the effective progress of a session in order to discover inconsistencies in it, for reengineering purposes. *Observed uses modelers* are interested in finding new techniques to improve their analysis abilities, whereas *observed uses analysts* are interested in finding new patterns to improve their analysis abilities.

A part of a *moderator tutors* job is to make reports for assessor tutors on learners abilities to collaborate and to work in a group. *Assessor tutors* want to evaluate knowledge acquired by learners in Web project management by verifying if the produced organization is coherent with the method taught during web project management courses. Lastly, *domain experts* are also involved in analyzing data. Whilst they do not currently analyze data since this analysis cannot be done during the learning session (exploratory manual analysis in this first stage), they nevertheless would be interested in analyzing data so as to understand what learners have done previously when they ask them for help.

4. What kind of data is being analyzed?

In this section, we will distinguish primary data (data that have not been processed) from derived ones (data obtained from other data). In our experimental system, we have three kinds of primary data (see Figure 5) : raw data (that is, data recorded by the learning system), additional data (that is, data linked to an activity but not recorded by the system during sessions) and content data (that is, data produced by the system actors). We also have derived data, some of them being indicators (that is, they have to be interpreted, taking into account the learning activity, the profile and roles of the actors, as well as interaction context), others being intermediate data.

From a reengineering perspective, we will use some raw data (either recorded by the learning system or not) in order to derive some new data that will be useful for system actors. We will

also need some additional data, such as the predictive scenario for the activity, and content data, that is, outcomes produced by actors during their activities.

From a pedagogical perspective, learning assessment includes both cognitive achievement and collaboration capabilities aspects. Whereas the first aspect can be evaluated by learners production analysis (Jonassen, Davidson, Collins, Campbell, & Bannan Haag, 1995) and mainly concerns data labeled as S-3.1, S-3.2, S-5.1 and S-5.2 in Figure 5, the second one can be evaluated using comporment observation and perception analysis (Henri, & Lundgren-Cayrol, 2001).

In this paper, we will focus on this second aspect and more precisely on role emergence and we will now detail the most important data that helps us to formalize emerging roles arising from learners' activity.

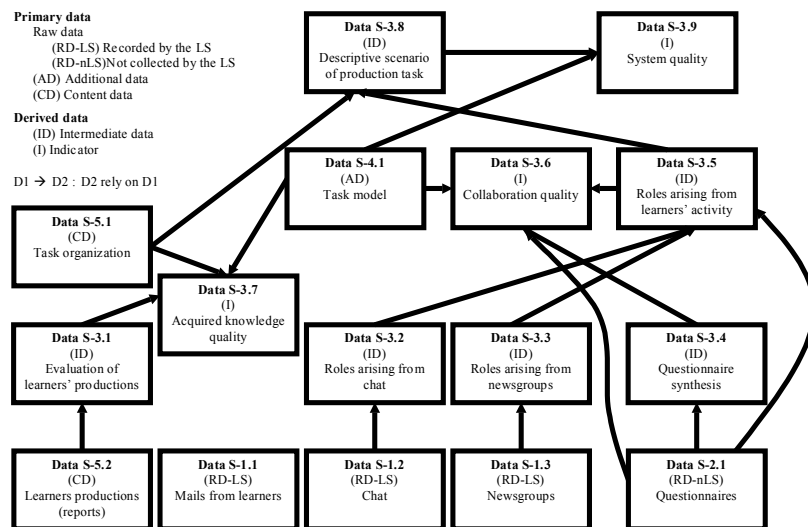


Figure 5. Dependencies between data

We will first detail the raw data, either recorded by the learning system or not. Please note that many of our raw data deal with communication tools' tracks and that, in the original tracks, messages are written in French. All those messages have been translated into English for insertion here.

Data S-1.2 (data arising from chat) corresponds to the transcription of all communications exchanged between learners via the chat service. A partial transcription of such messages can be found in Figure 6.

02/06/2004

10:45:33 *Arnaud* : I agree to start working on functional guidance, but I will need some help.

10:46:02 *Myriam* : Yes, everything's going well, I should have finished soon and I would help you after that.

11:10:59 *Arnaud* : Do you need some help for the database ?

Figure 6. Excerpt of messages exchanged on the chat service

Data S-1.3 (data arising from newsgroups) corresponds to the transcription of the entire set of messages posted on newsgroup services. An example of such a message can be found in Figure 7.

41
31/01/2005, 17H43
Myriam
Noemail

Re : IMPORTANT : WORK DISTRIBUTION

I agree to do the graphic design, but I rather would work with *Arnaud* since I have already begun this work with him and it will be simpler to continue together rather than with another people. Moreover, we are in the same class group and therefore it is easier to meet.

Figure 7. Excerpt of a message exchanged on the newsgroup tool

Data S-2.1 (data arising from questionnaires) consists of questionnaires, whose main goal is to evaluate the group functioning by measuring parameters such as participation, collaboration and organization. Student answers to questionnaires are measured with a Likert scale (Babbie, 1992), graduated from 1 to 5 (strictly disagree, disagree, neither agree nor disagree, agree, completely agree). Learners can also give some detailed explanations about their answers. An example of such a questionnaire can be found in Figure 8.

I- PARTICIPATION

1- You always felt yourself integrated in your team because you were widely sharing information.

Completely agree: from a functionality viewpoint, as well as from a graphical viewpoint, we always concert all together before taking a decision. Discussion was therefore always privileged.

2- You always felt yourself integrated in your team because all decisions were taken after considering all opinions.

Completely agree : we have always take into account everyone's positions in order to make project progress in the good direction, with a good collective spirit.

Figure 8. Excerpt of a completed questionnaire

We will now detail data obtained by combining it with other data (either primary data or already synthesized ones).

Data S-3.2 (data related to collaborative communication tools, i.e. role emergence from chat) is derived from the transcription of all communications exchanged with chat service. Emerging roles are extracted from this transcription using pragmatic markers (Cottier, & Schmidt, 2004) that need to be defined. Notice that a particular learner can assume many roles throughout a given learning session. All these roles will be reported. So, you need to define pragmatic markers associated to each role you want to find within the gathered sequences. In order to do that, you need to detect both the *lexical field* of pertinent terms linked to a particular role and the *dialog topics*. Detecting lexical fields mainly consists in finding words that can be termed as activity warning, such as: work, organization, design, to work, to organize, to design, etc. You must often base yourself on the fact that those terms are the most used throughout dialog in order to consider them as pertinent. To detect dialog topic, you can use their proximity with lexical fields previously detected. In order to correctly define pragmatic markers, you also need to identify who is the originator and who are recipients of the gathered dialog and to interpret exchange meaning (see Figure 9 for an example).

Pragmatic markers linked to our experiment have been manually extracted and organized into an ontology. We are currently working on an automated service that will extract emerging roles from communication tools tracks.

<p>Speaker 1 (<i>informant role</i>): "I have looked what is asked for on Symba, and, visibly, the work expected for this phase is the <u>project plan</u>"</p> <p>Speaker 2 (<i>supervisor role</i>): "I have checked your work you have done on <u>functional analysis</u>. I think you must improve it."</p> <p>Speaker 3 (<i>organizer or leader role</i>): "For this <u>general design step</u>, I propose to organize ourselves in the following way"</p> <p>Speaker 4 (<i>follower role</i>): "Ok, everyone agrees ?"</p> <p>Speaker 5 (<i>debate opener or negotiator role</i>): "We must <u>discuss that</u>, I don't agree. I think it is better to..."</p>

Figure 9, pragmatic makers example: lexical fields are **highlighted**, dialog topic are underlined

This data therefore consists of a list of roles arising from observed communications. This list is annotated with information about which student(s) takes which role(s) and consists in a structured file (see Figure 10). One can notice that these roles and assignments can be identical to those arising from other communication tools (*e.g.* newsgroups, see data S-3.3).

```

<roles>
  <role>
    <type>Socio-affective leader</type>
    <persons>
      <person>Fred</person>
      <person>Clara</person>
    </persons>
  </role>
  <role>
    .
    .
  </role>
</roles>

```

Figure 10. annotated list of roles arising from chat (or newsgroup) analysis

Data S-3.3 (*data related to collaborative communication tools, i.e. role emergence from newsgroups*) is derived from the transcription of all communications exchanged on newsgroups. Emerging roles are extracted from this transcription using pragmatic markers that need to be defined (see Figure 11). This data therefore consists of a list of roles arising from observed communications (see Figure 10). This list is annotated with information about which student(s) takes which role(s) and consists in a structured file (same structure as for data S-3.2). One can notice that these roles and assignments can be identical to those arising from other communication tools (*e.g.* chat service, see data S-3.2).

```

Arnaud
So, I propose the following schedule for the project:
June 3 -> legal aspects
June 3 to June 4 -> XHTML and PHP structures definition
June 3 to June 12 -> Overall, we can do the following : conception, content
structure
[...]
June 23 : final tests and presentation preparation

Myriam
Re : Project Planning
I approve your project planning

```

Figure 11. pragmatic makers identifying a 'functional leader' role in newsgroups

Data S-3.4 (data related to questionnaire synthesis) is made of answers to questionnaires (data S-2.1) synthesized in percentages and reported within an evaluation grid summarizing this information for each question.

Data S-3.5 (data related to new roles arising from learners' activity). The study of interactions made with Symba communication tools (data S-3.2 and data S-3.3), as well as detailed answers made to questionnaires (data S-2.1), allow to the evaluation of the collaborative process from a cognitive and socio-affective viewpoint. This manually derived data consists of an ad-hoc free text report (whose writing is guided by some questions) and allows the evaluation of the coherence between the different facets of each role.

In order to highlight this data, we can take an example from our experiment. We have observed the following key points: (i) In the transcription of chat messages (data S-1.2), one student has been unanimously appointed “project leader” and, therefore, one can expect that this student will play his leader role and that he was chosen by his colleagues because they think he is qualified in group management. (ii) Analysis of data S-1.2 with help of pragmatic markers (in order to produce data S-3.2 – roles emerging from chat messages) indicates that most of the interactions are organized around another student. (iii) In detailed answers to questionnaire (data S-2.1), everyone acknowledges that the initially designed “project leader” was rejected by all other team members, even if he tried to fulfill his (task based) functional role. This rejection was based on a lack of communication skills of this “project leader”.

To synthesize this situation, one can say that instructional designers have defined a predictive role of “project leader” and are expecting that this “project leader” act as a leader on their own. Although he was effectively a leader with respect to tasks they have to do, he was not completely accepted as a leader with respect to his communication skills. Consequently, one can suggest that this “project leader” role can be split in two facets: a functional one and a socio-affective one (see Figure 12).

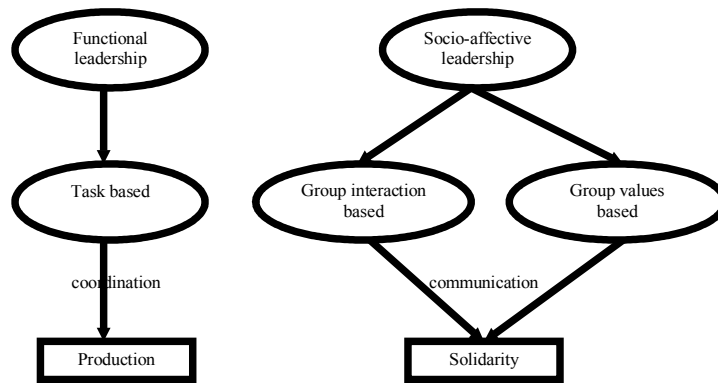


Figure 12. Functional and socio-affective leadership (Hotte, 1998)

Moreover, in our experiment we are in the context of a self-managed group. In such a context, the two facets must be simultaneously present in the student assuming the “project leader” role (in order to coordinate interactions inside the group and to facilitate communication between group members). Otherwise, if they are not simultaneously present in the same people, this leads to a group dispersal and reduces collaboration quality.

Thus, in our experiment, the data S-3.5 allows us to verify the coherence of the different facets of the roles arising from activity.

Data S-3.6 (collaboration quality) corresponds to an indicator allowing the evaluation of the collaboration quality between learners. This data consists of an ad-hoc free text report (whose writing is guided by some questions) and is manually made using reports showing role coherence arising from learners’ activity (data S-3.5) combined with the task model produced by designers (data S-4.1). This verifies the parallels between predicted roles and observed ones (at a per learner level) and requires information on whether collaboration takes place between learners or not. This is derived from the questionnaires, both synthesized form and detailed answers (that is, data S-2.1 and data S-3.4).

We will lastly describe one additional data which is used to highlight synthesized data.

Data S-4.1 (task model specified by instructional designers) corresponds to the task model as anticipated by designers (see Figure 3). That is, an indication of the activity sequence that learners are supposed to produce using the workplace organization from Symba. This task

model is expressed using IMS Learning Design (and, technically, it is an XML file conforming to IMS/LD specification, see Figure 13).

```
<imsld:learning-activity identifier="LA25">
  <imsld:title>Integration</imsld:title>
  <imsld:learning-objectives>
    <imsld:item identifierref="" identifier="LA25-obj">
      <imsld:title>
        To know integrating all previously made pieces to the Web site
      </imsld:title>
    </imsld:item>
  </imsld:learning-objectives>
  <imsld:environment-ref ref="S2-mail-service" />
  <imsld:environment-ref ref="S3-chat" />
  <imsld:environment-ref ref="S4-newsgroups" />
  <imsld:environment-ref ref="L018-outcomes-ress-detailedConception.zip" />
  <imsld:activity-description>
    Consists in integrating all previously realized pieces
    (graphical design, videos...)
  </imsld:activity-description>
  <imsld:complete-activity>
    <imsld:user-choice />
  </imsld:complete-activity>
</imsld:learning-activity>
```

Figure 13. Predictive task organization (excerpt)

5. Data analysis

Analysts, and sometimes tutors, analyze data in order to synthesize the information they contain. Results of these different analyses are then used by many actors of our e-learning system. Analysts use them in order to produce new analyses, tutors use them to evaluate learners and designers use them to improve their predictive scenario (following a reengineering cycle) and to capitalize knowledge so as to produce new scenarios (engineering purpose).

In the previous section, we have described data (either primary or derived ones) necessary to produce ‘collaboration quality’ indicator. We will now detail how those data are analyzed to produce this indicator. We recall that all dependencies between data are depicted in Figure 5.

5.1 Who analyses data, how and when?

Presently, all analyses are made by the observed uses modelers (analyzing raw data) and observed uses analysts (making analysis from analysis reports made by observed uses modelers). In this first step of our experiment, most of our analyses are done manually, at the end of a learning session. We will first detail the analysis made by *observed uses modelers*.

The analysis of **data S-1.2** (*data arising from chat*), that is, tracks produced by learners via their interactions through the chat system, is done using pragmatic markers (Cottier, & Schmidt, 2004) in order to identify emerging roles.

Analysis of **data S-1.3** (*data arising from newsgroups*) is very similar: tracks are produced by learners, by their interactions through newsgroups, and are analyzed with pragmatic markers (Cottier, & Schmidt, 2004) at the end of the session.

Data S-3.2 (*role emergence from chat*) and **data S-3.3** (*role emergence from newsgroups*) are then analyzed together. Roles lists arising from both data are merged into one list which is, then enriched with annotations (learners in role) that they contain.

Analysis of **data S-2.1** (*data arising from questionnaires*) is made by observed uses modelers and consists of synthesizing answers to questionnaires in percentages and to report them with an evaluation grid.

We will now evoke analysis made by *observed uses analysts*. The analysis they have to do mainly consists of synthesizing information from data S-2.1, S-3.4, S-3.5 and S-4.1 in order to produce ‘collaboration quality’ indicator (that is, data S-3.6).

Since **data S-2.1** (*data arising from questionnaires*) also contains detailed answers to questionnaires, it can be used in order to make a synthesis concerning collaboration inside the group. The analysis of **data S-3.4** (*questionnaires synthesis*) is carried out by the human analyst to highlight whether collaboration takes place or not (focusing on learners abilities to collaborate and to work in a group). For this purpose, questionnaires were built allowing the evaluation of variables such as: participation, collaboration, communication, work atmosphere, leadership, ... (see Figure 14).

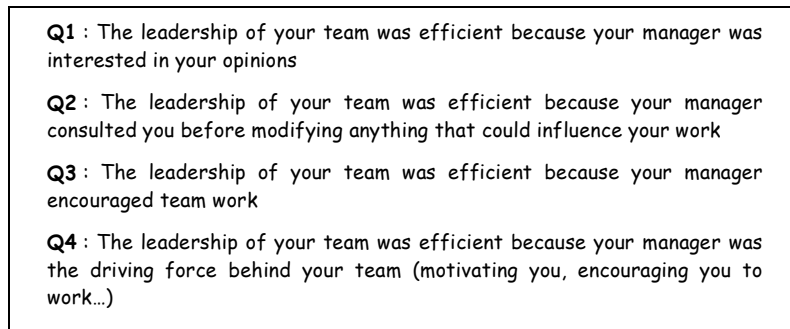


Figure 14. “Leadership” evaluation with questionnaire

They also need to analyze **data S-3.5** (*data related to new roles arising from learners’ activity*) since they need information regarding similarities between effective roles and their socio-affective facets. For example, a learner having a functional role of Project manager would ideally have an organizer or a leader socio-affective role and be rather active (whereas if he has a follower socio-affective role, he would be less successful in his task).

Finally, they need to analyze **data S-4.1** in order to compare roles arising from the activity to those that were predicted by instructional designers.

For the moment, this analysis is done at the end of a learning session. Nevertheless, obtained results suggest that it would be judicious to detect functional and socio-affective role mismatching during the session in order to react as soon as it is detected. This would imply adopting a formative evaluation rather than a summative one (especially for questionnaires). This would also imply automating role extraction with pragmatic markers. We are presently working on it, building an ontology of pragmatic markers. We have already extracted a first subset of lexical fields and dialog topics and we are currently working on extending them.

5.2 Who uses the results of the analysis, how and for which kind of action?

The results of the different analyses are used by many actors of our e-learning system. Analysts use them in order to produce new analyses, tutors use them to evaluate learners and designers use them to improve their predictive scenario (following a reengineering cycle) and to capitalize knowledge so as to produce new scenarios (engineering purpose).

Therefore, motivations in data analysis can be viewed from one of the following viewpoints: engineering, reengineering, learners regulation.

Moreover, all of our analyses are not carried out with the same goal. Some of them are only intermediary results used to produce new data, whilst others are indicators having a meaning by themselves.

We will first detail how analysts (both observed uses modelers and observed uses analysts) use the results of previous analysis in order to build new data.

First, *observed uses modelers* are in charge of formatting roles identified by analysts using pragmatic markers on chat messages (**data S-1.2**) and newsgroups messages (**data S-1.3**). They produce, respectively, data S-3.2 and data S-3.3.

Observed uses modelers also need to format percentages calculated by analyst arising from analysis of questionnaires (**data S-2.1**) using a synthesis grid.

Results of analysis of **data S-3.2** and **data S-3.3** are manually formatted in order to constitute a basis for data S-3.5. Then the results of the detailed answers analysis in **data S-2.1** is used by *observed uses analysts* to confirm roles arising from data S-3.2 and data S-3.3. For example, in our questionnaire we have asked students about “leadership”. The students have to explain if this role has been assumed by one particular student and to express their opinion about commitment of their leaders. If most student answers are similar, pointing out the same student, we can consider that this student has a ‘leader role’ and therefore this confirms that this role emerged from the activity.

We will now detail how indicators are (or can be) used from the following viewpoints: engineering, reengineering, learners regulation.

In our experiment, we have identified three indicators : acquired knowledge quality (data S-3.7), system quality (data S-3.9) and collaboration quality (data S-3.6). We will now particularly focus on *collaboration quality* (**data S-3.6**) which comes from the data S-2.1, data S-3.4, data S-3.5 and data S-4.1 joint analyses.

From an *engineering viewpoint*, this indicator will be useful for instructional designers to capitalize knowledge and produce new collaborative scenarios.

From a *reengineering viewpoint*, this indicator will also be useful to instructional designers as it allows them to improve their predictive scenario, taking into consideration effective collaboration that has been observed.

Lastly, from a *regulation viewpoint*, this indicator could be used by moderator tutors in order to correct collaboration problems that can emerge during the activity. Nevertheless, this use implies that these indicators must be computed during the learning session, which is not yet the case. It is also to be expected that these indicators could also be used by assessor tutors to attribute (at least partially) a grade to learners, but, in our experiment, assessor tutors have rejected this use as it seemed too subjective for them. More precisely, they pointed out that they were afraid of penalizing students since this indicator corresponds to an evaluation of the collaboration within the whole group (rather than an individual assessment). They nevertheless point out that such an indicator could be a great help from a regulation viewpoint (if it can be computed during the session rather than at the end of a learning session).

6. From data analysis to design patterns

Design Patterns were originally developed in architecture, then in software engineering and now one finds design patterns in communication interactions and e-learning issues: human computer interaction, Web design, pedagogical patterns, patterns for implementing an institutional e-learning centre¹. Design Patterns embody the design experience a community has developed and learned. They describe recurrent problems, the rationale for their solution, how to apply the solution and some of the trade-offs in applying the solution.

In this paper, we have already shown a way to produce ‘collaboration quality’ indicator in order to evaluate whether collaboration takes place or not in a group and to detect functional vs. socio-affective role mismatching among learners.

¹ e-LEN project, see http://www2.tisip.no/E-LEN/patterns_info.php (last consulted, April 2006).

Thus, we will now use design patterns in order to formalize both our observation problem and how to use the ‘collaboration quality’ indicator to tackle this problem.

We will use the design pattern language (Celorrio, & Verdejo, 2005) finalized by DPULS project (DPULS, 2005). We will nevertheless need to slightly modify this language so that it fits our needs better. Firstly, the DPULS language defines indicators used by described solution. We need a little more information on indicators, and, particularly, to distinguish if they are used as an input to achieve solution, or built during solution achievement, and, thus, are part of the solution. The second point is related to different viewpoints from which indicators can be used. In previous section, we have seen that indicators can be used from the following viewpoints: engineering, reengineering, learners regulation. Thus, we need to incorporate these viewpoints in our design pattern language.

6.1 Emergence of roles by tutor

Work done in order to produce data S-3.2 and data S-3.3, that is extracting emerging roles from the transcription of communications between learners using pragmatic markers (Cottier, & Schmidt, 2004) is not really dependant on the communication tool used. Indeed, work done with data arising from the newsgroups is very similar to work done with data arising from chat. The key point of this analysis is the method, that is, the use of pragmatic markers. This idea can be abstracted (or captured) with the following *design pattern*:

General		
Name	C2.1 - Emergence of Roles by Tutor	
Abstract	This pattern provides an approach to detect the emergence of roles.	
Category	Collaboration	
Context	Type of System	CSSL web-based
	Type of Situation	Collaborative (project pedagogy)
	Actors	Observed uses modeler Moderator tutor
	General Description	Your students use synchronous and/or asynchronous communication tools.

Figure 15. C2.1 DP > General section

Problem	
Statement	Human tutors can analyze communication tool tracks (exchanges between learners on <i>chat</i> , <i>forum</i> ...) in order to detect roles (either predicted or not) arising from those tools.
Tracking Focus	Actor's behavior
Analysis	The predictive roles imagined may not be confirmed; moreover during the session new (unpredicted) roles could emerge. For example, your scenario can only contain functional roles, whereas some socio-affective roles may be interesting.

Figure 16. C2.1 DP > Problem section

Solution			
Objective	Learning regulation		
Requisites	Indicators	Input	-
		Output	<ul style="list-style-type: none"> ▪ The characterization of roles arising from communication tool tracks ▪ An ontology of pragmatic markers which allows the characterization of roles that can be found in communication tool tracks
	Methods		The analysis of roles arising from communication tools tracks and the definition of pragmatic markers.
Description	<p>The sequence of activities of each group is gathered (possibly using <i>chat</i>, <i>forum</i> or <i>mail</i> tracks obtained from Learning System) by delegating one teacher by group to monitor their activities.</p> <p>Emerging roles are extracted from those gathered sequence using pragmatic markers (Cottier & Schmidt, 2004) that need to be defined. [...]</p> <p>In the end, you must arrange your pragmatic markers into an ontology in order to subsequently use them to automate role extraction from communication tools tracks.</p>		
Discussion	Heavy and time consuming analysis		
Example	<i>Symba</i> (Betbeder & Tchounikine 2003) experiment by LIUM		

Figure 17. C2.1 DP > Solution section

Related Patterns		
Set of Related Patterns	Related Pattern name	<i>C2 – Collective Organization of a Synchronous Activity</i>
	Relationship	<i>CAN BE USED BY</i>

Figure 18. C2.1 DP > Related Patterns section

Pattern Identification	
Author	V. Barré, C. Choquet, S. Iksal, N. Randriamalaka
Date	November 2005
Version	1.0
Bibliographic References	Cottier, P., & Schmidt, C.T. (2004) [...] Betbeder, M.-L., & Tchounikine, P. (2003) [...]

Figure 19. C2.1 DP > Pattern identification section

This pattern is thus an abstraction of the process used to produce data S-3.2 and S-3.3. But one can also use this DP in order to define an automated service that extracts emerging roles from communication tools tracks. In order to define such a service, you must provide two different elements: (i) communication tool tracks, and (ii) pragmatic markers (for your domain and tailored for the specific roles you want to observe) organized in an ontology (obtained using our “C2.1 - Emergence of roles by tutor” Design Pattern).

6.2 Collective organization of a synchronous activity

Another key point of our data analysis can be abstracted in a similar way: data S-3.6 (collaboration quality). First of all, the building of data S-3.6 uses data S-3.2 and data S-3.3 which are concerned by our first design pattern. Moreover, we use in this process two indicators formalized in (Manca, Persico, Pozzi, & Sarti, 2005a, 2005b). These two indicators can be used as a substitute (and as an enhancement) for data S-3.4 (questionnaire synthesis). The first one is *active participation* and aims at detecting the number of performed actions showing “active participation” by students in a given area / course. It takes into account three main activities: sending a message, uploading a document, attending a chat. The second one is *passive participation* and aims at detecting the number of “passive” acts performed by students (such as a message, downloading a document ...). This process thus leads to the following design pattern:

General		
Name	C2 - Collective Organization of a Synchronous Activity	
Abstract	This pattern provides an approach to evaluate the collective organization of a synchronous activity.	
Category	Collaboration Assessment	
Context	Type of System	CSCL web-based
	Type of Situation	Collaborative (project pedagogy) Synchronous
	Actors	Moderator tutor Assessor tutor Instructional designer
	General Description	In your CSCL system, students have to organize themselves in order to produce something (a document, a web-site, etc.) within a synchronous activity.

Figure 20. C2 DP > General section

Problem	
Statement	You want to evaluate the collective organization of a synchronous activity
Tracking Focus	Actor's behavior
Analysis	Learners use a CSCL support environment in order to (i) explicitly describe and organize what has to be done (task decomposition, role attribution, etc.), and (ii) produce something in agreement with their organization. The success of such a task relies both on the effectiveness of collaboration between learners and on an appropriate role partition among them.

Figure 21. C2 DP > Problem section

Solution			
Objective	Learner assessment Learning regulation		
Requisites	Indicators	Input	Active participation (Manca, Persico, Pozzi, & Sarti, 2005b) Passive participation (Manca, Persico, Pozzi, & Sarti, 2005b)
		Output	Collaboration quality (see data S-3.6)
	Methods	Analysis of role distribution	
Description	<p>You can use "C1 - Effectiveness of collaboration in CSCL" pattern to evaluate if and to extent collaboration is taking place.</p> <p>In order to evaluate whether role repartition among learners is appropriate or not, it is interesting to take into account that roles have many meanings, and particularly a functional meaning and a socio-affective meaning.</p> <p>When considering roles emerging from activity (see "C2.2 - Emergence of roles by system" pattern), one has to verify that functional roles are in concordance with socio-affective ones. Moreover, observed roles must be in accordance with the participation of related students (active or passive participation).</p>		
Discussion	-		
Example	<i>Symba</i> (Betbeder, & Tchounikine, 2003) by LIUM		

Figure 22. C2 DP > Solution section

Related Patterns		
Set of Related Patterns	Related Pattern name	<i>C1 - Effectiveness of collaboration in CSCL</i>
	Related Pattern type	Internal (Delozanne, Le Calvez, Merceron, & Labat, 2005)
	Relationship	CAN USE
	Related Pattern name	<i>C2.2 - Emergence of roles by system</i>
	Related Pattern type	Internal
	Relationship	CAN USE

Figure 23. C2 DP > Related Patterns section²

Pattern Identification	
Author	V. Barré, C. Choquet, S. Iksal, N. Randriamalaka
Date	November 2005
Version	1.0
Bibliographic References	Manca, S., Persico, D., Pozzi, F., & Sarti, L. (2005b) [...] Delozanne, E., Le Calvez, F., Merceron, A., & Labat, J.-M. (2005). Deliverable 32.6.01: A structured set of design patterns for the usage analysis, DPULS Project, Kaleidoscope NoE, 2005. Betbeder, M.-L., & Tchounikine, P. (2003) [...]

Figure 24. C2 DP > Pattern identification section

7. Conclusion

In a collaborative e-learning system, tracks arising from communication tools allow us to build useful indicators for all system actors. Indeed, some indicators like ‘collaboration quality’ (data S-3.6) can, at once, be used by tutors to evaluate learners, by analysts to build other indicators and by designers to evaluate the relevance of their pedagogical scenarios. From this last point of view, we have shown, in this paper, that considering emerging roles arising from communication tools tracks can be useful for reengineering purposes. For example, in our experiment, we have clarified a first reengineering cycle, and this first cycle has allowed us to enrich the predictive scenario made by designers by adding socio-affective roles arising from learning session tracks analysis. Role emergence was one key point of our reengineering process, and was in keeping with comparison of predictive scenarios and descriptive ones enriched with emerging roles.

² Pattern C1 consists in evaluating if and to extent collaboration is taking place using quantitative and qualitative analysis of interactions, Pattern C2.2 differs from pattern C2.1 by using an automated service in order to analyze communication tool tracks.

Another interesting point is that proposed indicators can be used in a more general framework than that of our experiment. Indeed, role mining from communication tools tracks can help to enlighten the effective use of the collaborative system and to push the collaboration quality indicator forward, whatever the collaborative experiment may be. These general frameworks then lead to the definition of design patterns. Moreover, in order to support the production of such generic indicators, we have defined software tools (Iksal, Barré, Choquet, & Corbière, 2004) that, once fully developed, will allow the analysis of the collected data based both on the predictive scenario and the formal description of elements to be observed. They will produce formal representations of user comportment, based on observation needs, and thus form a useful guide to implement the reengineering process.

Finally, although we have not formalized our methodology for defining Design Patterns, we think our approach could be generalized and applied to another experiences. Indeed, the participants of the DPULS project (ten research teams) have more or less employed the same methodology on their own experiments and the result was the definition of a structured set of forty Design Patterns³. We hope this result could be the first step of a process for capitalizing and sharing through the Technology Enhanced Learning community knowledge on usage analysis.

Acknowledgments

This work has been done within the framework of the DPULS project (DPULS, 2005), funded by Kaleidoscope Network of Excellence supported by the European Community.

References

Babbie, E. (1992). *The practice of social research*. Belmont, California, Wadsworth.

³ these Design Patterns are accessible at: <http://lucke.univ-lemans.fr:8080/dpuls/login.faces>

- Betbeder, M.-L., & Tchounikine, P. (2003). Symba: a framework to support collective activities in an educational context. International Conference on Computers in Education (ICCE 2003). December 2-5, 2003. Hong-Kong (China). 188-196.
- Celorrío, C., & Verdejo, F. (2005). *Deliverable 32.5.01: The design pattern language*. DPULS Project, Kaleidoscope NoE, 2005. [On-line]. Available : <http://www.noe-kaleidoscope.org/>
- Chikofsky, E. J., & Cross II, J. H. (1990). Reverse engineering and design recovery: A taxonomy, *IEEE Software*, 7(1), 13-17.
- Corbière, A., & Choquet, C. (2004a). Designer integration in training cycles: IEEE LTSA model adaptation. International Conference on Computer Aided Learning in Engineering Education (CALIE'04). February 16-18, 2004. Grenoble (France). 51-62.
- Corbière, A., & Choquet, C. (2004b). A model driven analysis approach for the re-engineering of e-learning systems. ICICTE'04. July 1-3, 2004. Samos (Greece). 242-247.
- Cottier, P., & Schmidt, C.T. (2004). Le dialogue en contexte: pour une approche dialogique des environnements d'apprentissage collectif. Colloque ARCo 2004. December 8-10, 2004. Compiègne (France).
- DPULS (2005). *Design patterns for recording and analyzing usage in learning systems*. Workpackage 32, Kaleidoscope Network of Excellence, supported by the European Community under the Information Society and Media Directorate-General, Content Directorate, Learning and Cultural Heritage Unit. Contract 507838. Consulted June, 2006, at <http://www.noe-kaleidoscope.org/>
- Henri, F., & Landgren-Cayrol, K. (2001). *Apprentissage collaboratif à distance: pour comprendre et concevoir les environnements d'apprentissage virtuels*. Sainte-Foy (Québec, Canada): Presses de l'Université du Québec, ISBN 2-7605-1094-8.
- Hotte, R. (1998). Modélisation d'un système d'aide multiexpert pour l'apprentissage coopératif à distance. Unpublished doctoral dissertation, Université Denis Diderot / Paris 7.

- Iksal, S., Barré, V., Choquet, C., & Corbière, A. (2004). Comparing prescribed and observed for the re-engineering of e-learning systems. IEEE Sixth International Symposium on Multimedia Software Engineering. December 13-15, 2004. Miami (USA).
- Koper, R., Olivier, B., & Anderson, T. (2003). *IMS Learning Design v1.0 Final Specification* [on-line]. Available: <http://www.imsglobal.org/learningdesign/index.html>.
- Jonassen, D. H., Davidson, M., Collins, M., Campbell, J., & Bannan Haag, B. (1995). Constructivism and computer-mediated communication in distance education. *Journal of Distance Education*. 9 (2), 7-27.
- Lejeune, A., & Pernin, J-P. (2004). A taxonomy for scenario-based engineering. Cognition and Exploratory Learning in Digital Age (CELDA 2004). December 2004. Lisboa (Portugal). 249-256.
- Manca, S., Persico, D., Pozzi, F., & Sarti, L. (2005a). An approach to tracking and analyzing interactions in CSCL environments. Proceedings of the E-learning Conference. 2005. Berlin (Germany).
- Manca, S., Persico, D., Pozzi, F., & Sarti, L. (2005b). *Deliverable 32.4.01: Know-how list*. DPULS Project, Kaleidoscope NoE, 2005. [On-line]. Available : <http://www.noe-kaleidoscope.org/>
- Morch, A., & Mehandjiev, N. D. (2000). Tailoring as collaboration: the mediating role of multiple representation and application units. CSCW'2000. December 2-6, 2000. Philadelphia, Pennsylvania (USA). 75-100.