Alfanet Deliverable D14 - Initial Evaluation Plan

Project Deliverable Report

D14 - Initial Evaluation Plan

Workpackage	WP1 - User Requirements and Specifications			
Task	T14. Assessment and Evaluation Plan			
Date of delivery	**Contractual**	30 09 2002	**Actual**	25 02 2003
Code name	ALFANET_D14_v1		**Version** 1.0	draft ☐ final ☑
Type of deliverable	Report (Plan)			
Security (distribution level)	IST			
Contributors	UNED, KLETT, SAGE, OUNL, EDP			
Authors (Partner)	Jesús G. Boticario, Carmen Barrera, Elena Gaudioso, Carlos Catalina (UNED) Ingeborg Hoke, Joerg Boy, Juergen A. Schmidt (KLETT) Carlos Fuentes, Mar Rodrigo, Cristina Arana (SAGE) Patricia Poelmans, Gerry Joosten, Leo Wagemans (OUNL) Maria Amelia Franco, Adalberto Moutinho, Elsa Escala, Susana Aleixo (EDP)			
Contact Person	Carmen Barrera (UNED) Universidad Nacional de Educación a Distancia Francos Rodriguez, 77 28039 Madrid - Spain		Tel: + 34 91 398 8507 Fax: + 34 91 398 66 97 Email: cbarrera@invi.uned.es	
WP/Task responsible	EDP / UNED			
EC Project Officer	Mrs. Elena Coello			
Abstract (for dissemination)	This deliverable describes the criteria and procedures that will be applied to evaluate the ALFANET system. This evaluation plan includes three different levels of assessment: Verification, Usability and Effectiveness. The evaluation of the 1st and 2nd prototype is mainly formative: technical verification and evaluating the general usability of the system. The evaluation of the final system (including all the functionality) has a more summative character: effectiveness assessment is then the focus of the evaluation.			
Keywords List	Evaluation, Validation, Assessment, Verification, Usability, Effectiveness, Effectiveness of the learning process, Educational Technology, Adaptive features			

ALFANET Project Coordination at: Software AG España, S.A. Ronda de la Luna, 4
Tel: +34 91 8079411 - fax: +34 91 8079447 - email: carana@softwareag.es Tres Cantos, E-28760 Madrid

Executive Summary

The purpose of this document is to define an initial evaluation plan that provides the criteria and procedures that will be applied to evaluate the ALFANET system. This evaluation plan includes three different levels of assessment:

- Verification
- Usability assessment
- Effectiveness assessment

The evaluation of the 1st and 2nd prototype is mainly formative: verification and evaluating the usability of the system. The evaluation of the final system (including all the functionality) has a more summative character: Effectiveness assessment is then the focus of the evaluation.

Technical Verification will be performed by developer partners at different levels: verification of existing software to integrate, each one of the software components developed in the project and the global application: the ALFANET system. The technical indicators to measure are grouped in Functionality, Robustness and Performance indicators.

Usability and Effectiveness assessment will be implemented in two pilot sites, and conducted mainly by each user partner with support of developer partners.

KLETT pilot site is oriented towards individual self-learning of a foreign language course “Spanish for beginners”, addressing especially people who want and need to learn Spanish in business context.
EDP pilot site is focused on the corporate training in the environmental area, specifically Solid Waste management, addressing people in a professional context.

The **usability study** will make use of diagnostic evaluation to assess the effectiveness, efficiency and satisfaction of users (learners, tutors, managers and authors) with the system. To assess the usability of the first prototype we will collect qualitative data by means of surveys (the greatest usability problems are expected at this stage) whereas assessing the usability of the second and final prototype by questionnaires.

The purpose of the **effectiveness study** is to obtain a global assessment on how adaptation affects the effectiveness of the learning process. The main hypotheses is “*The recommendations provided by ALFANET are effective when the learners follow them (which are dependent on their individual differences) and their learning experience through the e-learning platform enhances”."*

The methodology to do it will be based on empirical evaluation; different versions of the system will be tested: one version with adaptive features and the other one without them. After running the experiments we will take measures about the use of the system with the aim to obtain a measure of the effectiveness of each group. We can conclude about ALFANET effectiveness by comparison between results of both groups.

Initially, we have identified an exhaustive set of indicators about the effectiveness of the learning process. We can group them on:

- Perception and Subjective opinion of users (Learners satisfaction and tutors perception) to be measured quantitatively by gathering ratings via questionnaires
- Objective quality measures from users’ performance (Activity rates, Progress rates, Completion rates and Outcomes of the learning process) will be quantitatively measured by counting incidents carried out by the system (probably reported by the Auditing tools) and tracked at every chapter, lesson, interactivity, or assessment for a learner and for the entire group.

Other qualitative indicators will take into account less structured ways of obtaining opinion from the users

This Initial Evaluation plan will be refined in further project stages, and mainly the Effectiveness assessment will be deeply detailed for the next version of the document.

Table of Contents

1. Introduction

1.1 Situation

This report gathers the Initial Evaluation Plan (D14) for the ALFANET Project (IST 2001-33288).

This deliverable is the result of *T14 Assessment & Evaluation Plan* (led by UNED) within the *WP1 User Requirements and Specifications*. T14 has in charge the definition of an initial evaluation plan that provides the criteria and procedures which will be applied in *WP6 Pilots Validation* to evaluate the fulfilment of application requirements and the functionality described in the Specifications.

This deliverable will be the main start point for the tasks ***T62 and T64 Pilot Sites – Definition*** of *WP6 Pilots Validation,* where a new and enhanced version of the Evaluation Plan (D61) will be elaborated, taking into account the project's advance till then.

The results obtained from the evaluation process will be compiled in the following reports:

D62, D63, D64. Assessment of Pilot Sites for 1st, 2nd prototype and final system.

D66. Evaluation Results compiles and analyse the results of the whole evaluation, which is only due by the end of the project.

1.2 Purpose

The purpose of this document is to define an initial evaluation plan that provides the criteria and procedures that will be applied (in WP6) to evaluate the ALFANET system. This evaluation plan includes three different levels of assessment:

- Verification
- Usability assessment
- Effectiveness assessment

1.3 Overview

The Evaluation Plan covers the following points:

Chapter **2. Decision context** provides a global description of the evaluation framework and the pilot sites.

Chapter **3. Verification plan** describes the activities to determine whether or not the development of components of an application and the application as it is, fulfil the requirements established in the design phase.

Chapter **4. Usability assessment plan** describes the activities to assess the extent to which the ALFANET system can be used by students and tutors to achieve specified goals with effectiveness, efficiency and satisfaction.

Chapter **5. Effectiveness assessment plan** describes the activities to measure how ALFANET adaptive features facilitate the learning process.

The three above chapters are all organized in the following sections

- a description of objectives,
- context,
- methodology,
- assessment indicators,
- experimental design and sampling and
- presentation and evaluation of the results

Chapter 6 **Summary of timelines, yardsticks and milestones** summarizes the activities of the three above plans, and chapter 7 **Presentation and evaluation of results** describes how to show the general evaluation results integrated from the three sub-plans and the calculations to obtain the conclusions taking into account the general objectives of the plan.

2. Decision Context

2.1 Evaluation Framework

Levels of analysis

The methodology comprises the following levels of analysis, depending on the stage of the project: verification, usability assessment and effectiveness assessment.

- By **verification** we mean the execution of the necessary controlled trials and tests to verify that each module of the system works appropriately and that the set of integrated modules works as expected. *Technical indicators* inform us about the quality of the software.
- By **usability assessment** we mean assessing the extent to which the ALFANET system can be used by:
    - specified users: KLETT and EDP authors, students and tutors
    - to achieve specified goals:
      KLETT: developing and learning Spanish language course
      EDP: developing, tutoring and learning Solid Waste Management
    - with effectiveness, efficiency and satisfaction
    - in a specified context of use: KLETT: self study; EDP: in company training
- The objective of **effectiveness assessment** is to measure how ALFANET innovative (key) functionalities facilitate the learning process. It can be also referenced as measuring the effectiveness of the ALFANET system versus other traditional e-learning platforms.

General Methodology

To achieve the above objectives, the consortium has defined the following set of activities in the context of this Evaluation Plan

1. Specification of resources
    - To define the requirements of the pilot courses in order to profit from the key evaluation features
    - To define the requirements for configuring the participants in experimental groups
2. Preparation of resources
    - To prepare each one of the pilot courses, including educational material and instructional design.
    - To select the participants: learners, tutors and experimenters; configure the different teams of learners on each pilot
    - To prepare the environment and the resources to carry out the experiment (prototype installation, course installation and publishing, users registering and system configuration)
    - To prepare in advance the experimental sessions (preliminary information to provide learners and tutors, recommendations to profit interaction, collaboration and adaptation, scheduling of courses and sessions: estimated time, configuration of activities to do in a typical session).
3. Carrying out the experimental sessions
4. Gathering and analysis of results
    - To compile periodically outcome data at the end of each session / learning unit. These data can be obtained from normal system outcomes, questionnaires and interviews token by the participants, or data gathered by the experimenter, both directly observed or by means of specific queries to ALFANET.
    - Continuously analyse the obtained data, in order to obtain indicators and evaluate the expected impacts of the adaptive feature in the effectiveness of the learning process (removing suspicious measurements; averaging across all subjects and problems; organizing results table, graph).

5. To test the Hypotheses, drawing tentative conclusions about the hypotheses
6. Explaining Unexpected Results.

Scopes of the different prototypes

There are two different prototypes and one final type of the ALFANET system planned. The first prototype is intended to evaluate in an early phase a general vision of the LMS on the user side. It mainly provides a basic LMS tool, still lacking adaptation capability and auditing facilities. The first prototype includes:

- Completed functionality for the definition of Learning Contents (Authoring Tool), Knowledge Managers and Technological Resources Manager
- Basic Features of Navigation & Presentation Layer and Administration Facilities
- Use of an initial Pedagogical Model (with active and collaborative learning).

The second prototype adds to the first one advanced LMS traditional features, an initial adaptive capability and some basic auditing facilities. It includes:

- Refinement and Enhancement of completed Components.
- Completed functionality with the advanced features of Navigation & Presentation Layer and Administration Facilities; and completed functionality LMS Tracking and LMS Profiles Refinement.
- Basic Features of Multiagent Adaptive Learning and LMS Auditing Facilities.

The final system includes all the functionality.

Evaluation Stages

The evaluation of the 1st and 2nd prototype is mainly formative: verification and evaluating the usability of the system. The evaluation of the final system (including all the functionality) has a more summative character: effectiveness assessment is then the focus of the evaluation. The following scheme is summarising this:

Prototypes Levels of evaluation	1st prototype	2nd prototype	Final system
Verification	X	X	x
Usability	X	X	x
Effectiveness assessment		x	X

X means 'has to take place and has an important function'

x means 'has to take place but has a less important function'

2.2 EDP Pilot Site

2.2.1 Scenario Characterization

2.2.1.1 Pedagogical Aspects

2.2.1.1.1 Content

Solid Waste management

We choose the environmental area due to:

- the importance of this subject to EDP and to any other organizations all over the world, becoming more and more relevant to all society;
- its transversal nature crossing most of the social, economic and cultural layers;
- the natural involvement of a very diversified population in terms of knowledge and backgrounds.

Within the environmental area we selected “Solid Waste Management”, as the area where we will develop the course test. This sub-theme, besides kipping the above mentioned characteristics, covers some identified learning needs in several spheres of EDP Group.

2.2.1.1.2 Structure

The test course presents a modular structure, being each module constituted by lessons or units of learning, most of them without precedence. Having in mind that the test duration shouldn’t exceed one month, the course will have between 12 to 16 lessons with around 45 minutes each. Mandatory and optional exercises will be available to learners within each lesson. Every module will have an evaluation whose results will be complemented by those coming from a global assessment, done at the end of the course.

The content structure will be developed according to the existence of:

- Different levels of knowledge and/or interest within each learning unit
- Random connections between learning units allowing several sequences

2.2.1.1.3 Elements

We foresee that the course composition will be the following:

- Index
- Introduction
- Modules

Each module will be constituted by several learning units, each one will begin by an informative section, followed by an interpretative section and finishing with a reflective one, leading to the concept’s assimilation.

2.2.1.2 Users

2.2.1.2.1 Learners Groups

There will be target and control groups. Each one will have more or less 20 experience e-learners, with similar characteristics and profiles (age level, responsibilities, background, …).

Non-adaptive aspects, at usability and assessment levels, will be the aim of the control group.

2.2.1.2.2 Tutors

All groups will be coached by the same tutors.

2.2.2 Assessment and Evaluation Strategy

2.2.2.1 Levels and Instruments of Assessment

Having in mind the guidelines expressed in the assessment and evaluation plan, we will use different instruments for each level of assessment, according to:

- Verification

  Technical indicators

- Usability

  Questionnaire

- Effectiveness

  System indicators and questionnaire

2.2.2.2 Test Population

The *verification* level of assessment will be, naturally, performed by EDP project team, helped by some authors, tutors and learners.

Concerning *usability*, all stakeholders involved will be questioned (learners, tutors, authors and managers). Regarding *effectiveness*, test population will only involve learners and tutors once, till now, LMS platform data is not available to authors and managers.

2.2.2.3 Evaluation Moments

The Verification and the Usability assessments will be implemented in all prototypes; the Effectiveness assessment will be focused in second and third prototypes.

2.2.3 Time Schedule

- March 03: Objectives and contents identification
- April – May 03: Storyboard definition
- June – July 03: Content's development including the acceptance content tests
- September 03: First Prototype

2.3 KLETT Pilot Site

2.3.1 Description of the pilot site

KLETT is planning to develop a language course "Spanish for beginners", addressing especially people who want and need to learn Spanish in business context. An increasing number of firms is getting involved in business with Spanish speaking countries and therefore the need for employees increases to get acquainted with this language.

The course will comprise 16 lessons, each lesson containing the same structural elements:

- an introductory element (a story, a dialogue...) covering a certain topic
- presentation of grammar
- presentation of vocabulary and pronunciation
- exercises focussing on grammar
- exercises focussing on vocabulary

Language courses – especially courses for beginners – need a consequent structure/ progression of one lesson after the other, because pre-knowledge (regarding grammar and vocabulary) normally consists only of content provided in previous lessons. Nevertheless it will be possible to navigate freely between all the elements of the course, either by individual choice of the learner or by recommendations of the system, a tutor, or co-learners. This will be of special relevance for the group of false beginners, a group of learners who already have prior experience with the language but not enough to start at a higher level of the course, which requires solid knowledge of grammar and vocabulary of the lower level. These learners can choose the lesson they want to start or continue with. They can choose a certain part of a lesson (e.g. start with grammar exercises and look up general presentation of the relevant problem later). Or the learners can choose specific exercises with the assistance of a search machine. – A separate section will provide the learner with statistics showing him/ her, which parts of the course are already done and what are the results of the learner's efforts.

After every four lessons the learners will be given a compulsory assessment test. And having passed these four tests successfully the learner will get a certificate. On request the learner can be provided with additional voluntary tests.

The language course will be designed for individual self learning, because often it will be only one person in a firm starting with a language course at a given time, due to circumstances not allowing to wait until a group of learners can be gathered. With regard to this use case adaptive features will mainly refer to navigation (recommendations of what to do next).

Nevertheless the same course can be run with the full range of adaptive opportunities: The learner being provided additionally with both collaborative and tutorial support.

2.3.2 Groups Participating in the Evaluation Process

1. Testing the first prototype

During the development of the first prototype a permanent process of testing the functionality and capability of the authoring tool as well as of the technical resources manager, of the features of the navigation and presentation layer and the administration facilities takes place by authors and editors. Also the knowledge manager will be tested while making use of the pedagogical model. Therefore a permanent **verification** assessment is given by these persons. Before releasing the first prototype the developing team (authors and editors) have to test the verification formally.

Authors and editors being already experts in e-learning material development are permanently involved in the process of discussing **usability**. They will contribute to questions concerning the usability of the parts of the system used by the learners. And they will test the usability of the authoring tools. Nevertheless it is necessary that groups of learners (students) will assess the first prototype. These groups should be composed by a) students with a rich background of experience in the use of e-learning products and b) students still lacking this experience. Observation of and interviews with both groups will show the strengthens and weaknesses of the first prototype and therefore allows to draw conclusions in order to improve the system while developing the second prototype.

2. Testing the second prototype

Again – corresponding to what is said above – the **verification** will be mostly done during the content development process, authors and editors being involved on the side of the users. And also the tests focussing on **usability** might reveal aspects concerning verification. On the other hand the group of authors and editors will also perform usability tests, mainly with respect to the usability of the authoring tools.

For this phase of the assessment usability tests are the main focus. They should be performed – like it is explained above -by two groups of people (students): one group being experienced users of e-learning facilities and one group being not well experienced. The group of experienced people could comprise persons that had already tested the first prototype. This will give us the advantage to provide us with feedback about the extend of improvements comparing the first and the second prototype. The persons of the second group – the less experienced users – should have no prior experience with the first prototype.

Both groups of test users will also be questioned about **effectiveness**. To what extend and with what instruments this part of the assessment will be done needs to be defined in a later stage of the project.

3. Testing the final system

Concerning **verification** cp. what is stated for the first and second prototype.

For **effectiveness** assessment being the focus now we will need a different set of test groups on the side of the learners:

- one group without any adaptive features (no navigation proposals, no collaboration)
- one group provided with proposals for navigation
- one group provided with collaboration features
- one group provided with the complete range of adaptive features

All users (authors, editors and all students participating in the different test groups) will be questioned about **usability** features.

3. Verification Plan

3.1 Objectives

The verification process determines whether or not the development of components of an application and the application as it is, fulfil the requirements established in the design phase. These requirements can be grouped in the following aspects:

- Code rules: These rules describe how the code must be implemented in order to allow later modifications, improvements and extensions. Rules are relative to the software units naming (how to name methods, variables, etc.), comments (what and how the code must be commented) and structure (modularisation of the code logic).
- Functionality: This verification aspect determines if the functionality required for software units, at the design time, have been provided in the correct way. This point is usually checked at the integration and takes into account a normal application operation.
- Robustness: This aspect determines how the software operates in unusual situations, this is, what is its answer when it is provided wrong inputs or in extreme operation situations.
- Performance: This point takes into account not the right answers of the software but how good its behaviour is.

Verification activities are in-process activities performed concurrently with the software development and they are done usually just before and during the different integration processes, being these ones, from the littler size software units to the modules and components integration.

Due to the project character, it is considered that it cannot be defined general code rules but it is necessary that each developer partner take into account this aspect and it will be provided recommendations about it.

3.2 Context

The project character forces to take into account the following software units in order to test the entire application.

- Existing software to be integrated with the system: To verify this software components it is necessary to differentiate between two aspects: The results required from the component are the expected ones and the interfaces provided for its integration are according to the design specifications.
- Application Developments: Within this group of software units, it can be differentiated the following kinds:
    - Software Module: It represents any development done by any partner, which addresses any partial functionality of the system. Usually, it will be associated to the concept of file (or a reduced set of files). One module will always be the result of the work of one partner.
    - Software Sub-components: It represents a set of modules which, together, address one of the significant functions of the product. It will present a clear interface for integrating it with the other sub-components. One sub-component will generally be the result of the work of one partner.
    - Software Component: It represents a set of sub-components which, together, address one of the main areas of functionality of the product. It will also present a clear interface for integrating it with the other components. One component will generally be the result of the work of one or several partners.
    - ALFANET System: It represents the complete ALFANET application. It will be the result of the work of the entire consortium.

Next, it is described what should be verified, when the verification planning and tests should be defined and who is the responsible both of the planning and to performing of the test. All of this, for each kind of software unit described.

Existing software to integrate: As it is mentioned, it must differentiate between the software as is, and its communication interfaces. For the first one, the verification of code rules and internal functionality are avoided but it is necessary to test both the provided functionality, planned in the architecture, and the performance. The planning of this verification process will be done when the general architecture is defined, by the application integration responsible and performed at the first prototype integration. On the other hand, and for the communication interfaces, the plan will be defined also at the end of the architecture definition and updated if necessary during the development process. Verification of code will be performed also during the development process and, at least, at the first prototype integration for functionality and robustness aspects. If any change is done during the rest of the project it must also be tested in the next integration.

Software Modules: Verification of software module should be planned after its design has finished and the plan can be updated if it is necessary. The responsible of the plan definition will be the same one that designs the module, with the assistance of the development team. The test should be performed during the development process and when any module implementation or modification is finished, doing the necessary changes if the results are wrong.

Software sub-component: The verification of these units is relative to the integration process and must ensure that exists a right communication between the involved modules. It can also summarise the other performed tests for the corresponding modules. This process should be planned once the corresponding communication interfaces are designed and updated and when, due to newly detected needs during development process, any modification of the design is done. At the first moment, the plan responsible will be the design team of the sub-component but the development team, during the implementation, could assist it. The test will be performed, as in the software module, during the development process, at any necessary integration of the corresponding modules, being the responsible of its execution the sub-component development responsible.

Software Component: It has the same characteristics as the sub-component ones but the responsible teams of its planning and execution will be the corresponding responsible partner of the component.

ALFANET system: The responsible partner of the plan and execution of the global application verification is the responsible one of the integration and application delivery. The task of complete system verification comprises the validation of individual functionality, robustness and performance for each one of the integrated components and a first validation of the complete application functionality and performance.

In any case, all the modules, sub-components and components must be successfully tested before they are delivered to the next integration process. The documentation about the tests performed and their results should be included also as part of the software unit as is described in the “Software Engineering Guidelines” document (see appendix 2. Testing procedure).

On the other hand, critical fault detected during the tests should be communicated to its development responsible in order to do the necessary changes on the code. Once needed modifications are done, they will be repeated the affected tests again.

3.3 Methodology

Verification process can be divided in four main points which must be followed with the same order as they are described:

Technical indicators definition: These indicators will be defined in the next section of this document and will cover the objectives that has been described in previous sections. The class of measurement of this indicators and the values that they could take will be also described. As well, it will be described in which plans should be taken into account, according to the different character of each software unit and each indicator.

Verification planning: The verification plan will be defined separately for each application software unit according to the description done within the previous *Context* section. It takes into account the indicators defined, their possible values and the considerations about their application field (those aspects are described also in the next section of this document). Verification plan should define also how to obtain the value for each technical indicator and, if any of the technical indicators has no relevance in the software unit verification, it should be pointed and explained. Finally, it will be also defined the threshold values for these indicators.

Test execution: Tests will be executed according to the verification *Context* described in the previous section and the corresponding verification plan. The results of these tests will be the values described for the corresponding indicator or the ones that have been considered in the plan definition, taking into account that the final results for each plan should give a value for each indicator defined.

Verification valuation: Once the plan has been defined and the test process has been performed, every technical indicator will be evaluated according to the procedure defined in the verification plan. Its results will be compared with the defined threshold values and the necessary decisions about software unit changes, improvements or fixes to be done will be taken. If any change is done for any software unit, the test process and its evaluation will be repeated, at least, for those points that are affected by the modifications.

3.4 Technical Indicators

Based on the objectives of the verification process described within the corresponding document section (see 3.1 Objectives), it has been defined the following Technical Indicators.

Code rules: As has been described in the *Objectives* section, due to the project characteristics it cannot be defined general code rules. Nevertheless, within this point they are given general recommendations about what should take into account the code quality validation. These recommendations are divided in three main groups:

- *Naming*: Rules about how methods, variables, classes, etc. should be named. They should have as objective to provide a clear vision of the code and make easy its maintenance. They should also describe indentation rules, where the comments should be placed, where and how the variables should be declared, etc.
- *Structure*: It is relative to the code modularity and affects the software maintenance, performance and readability. It can be measured by means of methods length, functionality grouping, code flow complexity, nested deep, variable declaration grouping and avoiding hardcoded values.
- *Comments*: Code should be commented in a proper way in order to make it maintainable and readable. At least, a properly commented code should contain a description for methods and objects and the procedures that code implements in those places where the code reading is not clear enough. Comment information should provide also who is the author of the code, who has modified it, when it has been finished and modified, etc.

Functionality: There are two indicators that can be measured:

- *Functionality provided*: It is the relation between the planned functionality and the functionality that is really supported at the test phase. It can be expressed in percentage. For its measurement, the corresponding plan should take into account the grouping of the test sets within functionality aspects. Then the valuation of one functionality should be taken as supported when the related test questions are answered (it doesn't matter whether the operation is correct, but only that it is supported). The threshold will be given depending on the plan. For instance, for module verification the functionality provided should be 100%, but for application verification, the threshold could be 50% for the first prototype, 80% for the second one and 100% for the last one.
- *Correct functionality provided*: It will be calculated in the same way as the previous one, but it will take into account the result of the test sets. It will consider as correct the functionality that is provided in a proper way when the input data for the test is in the normal operation. The test questions that give unusual data as input should be only considered within the *robustness* aspect.

Robustness: This aspect considers the behaviour of the software in front of unusual operation conditions. Verification plans should take it into account when the tests are defined. It can be measured by three indicators.

- *Functionality that takes into account errors*: It means when the incorrect input data is detected and the software responds in a proper way. It can be expressed as the percentage of the provided functionality that detected incorrect inputs.
- *Functionality that gives incorrect answers*: It means the functionality that doesn't detects the incorrect input data and gives an answer. It can be expressed as percentage of the provided functionality
- *Functionality that crashes*: It means functionality that doesn't detect the incorrect inputs and causes an application crash. It can be expressed as percentage of the provided functionality and its threshold should be 0%.

Performance: It is important to take into account, for the plan definition, where this kind of tests should be executed. Performance tests should be run in a machine similar to the one that will be used to exploit the final application. For this aspect we differentiate two kinds of software units:

Modules, sub-components and components

- *Operation time*: It can be defined as the maximum time that the software unit spends in any of its functionality. It should be measured in milliseconds and the verification plan should consider within this aspect all the functionality that the corresponding software unit provides.

Application

- *Transmission input load*: It will measure, in Kb, the load that the client induces when sending a request to the application. Normal values for this indicator are less than 1Kb.
- *Transmission output load*: It will measure, in Kb, the size of the application responses. Normal values for this indicator are less than 50Kb.
- *Response times for a single operation*: It will measure, in milliseconds, the time that the application spends to give an answer, measured from the starting point to the end point of the application operation, without taken into account transmission times.
- *Maximum response times for normal operation*: It will measure, in milliseconds, the average time that the application spends to give an answer under normal operation conditions, measured from the starting point to the end point of the application operation, without taken into account transmission times. Normal operation conditions means that the application is being requested by a number of clients similar to the predicted use.
- *Overloaded operation behaviour*: It detects application or server crashes when it is considered an application use over the 125% of the normal one. The operation framework should be dimensioned in the way that these crashes are avoided under this operation condition.

These values will help to design the operation framework for the final application exploitation.

3.5 Experimental design and sampling

There are two kinds of verification plans from the template design:

Verification of functionality and robustness: It could be designed by filling the following table.

Functionality	Calling method	Input Data	Results

Where:

Function/Functionality describes the method, operation or functionality tested

Calling method gives the necessary information about the method signature of the component interface.

Input Data indicates which data has been considered as input of the method test.

Results are the expected results of the tested functionality.

Tests should contain both usual operation inputs and unusual ones. Within the result checkbox could be filled at the testing moment with one of the following results:

OK: When the functionality is provided and the answer is proper.

Wrong: When the functionality is provided but the answer is not the expected.

NP: When the functionality is not provided.

Crash: When the software unit crashes caused by an input.

Verification of application performance: It could be designed by filling the following table.

Request description	Operation Times			Load (Kb)	
	Single request	Normal	Overload	Request	Response

3.6 Presentation and evaluation of the results

As in the previous section, the presentation and evaluation of the results are divided depending on the aspects measured.

Verification of functionality and robustness: It could be designed by filling the following table.

Functionality	Result	Comments

Where:

Functionality gives a description of the verified functionality.

Result is the result of the test execution, which is calculated as the worst result of all the test questions related with the corresponding functionality. The answers, ordered from the best to the worst result are *OK*, *Wrong*, *NP* and *Crash*.

Comments. This box can be filled with the comments about what can be the problem and with the actions to be taken in order to fix the problem.

Verification of application performance: It could be designed by filling the following table:

Request description		Operation Times			Load (Kb)	
	Threshold	Single request	Normal	Overload	Request	Response
Average						
Maximum						
Special Cases						
Conclusions						

Here the average and maximum values are the corresponding calculations from the tests results in order to show how is the behaviour of the application. Fields named *Special cases* are available for those situations that should not be included in the normal calculations. For instance, it should be considered as an special case all the file uploading operations or those ones that serves large files to the client (documents or any other) that aren't parts of the application as is. In the *Conclusions* box it will be pointed how well is the application behaviour and the decisions to be taken about the operation framework dimension.

4. Usability Assessment Plan

4.1 Objectives

4.1.1 Usability definition

One of the key questions when assessing usability and determining the objectives of usability assessment is defining what we exactly mean with the concept 'usability'. Many definitions of usability exist. In Appendix 3 a presentation of some definitions is included. These definitions are found by searching the Internet on Internet usability sites.

These different definitions have some features in common. They all have the attributes: effectiveness, efficiency and satisfaction. So we propose to use the ISO 9241-11 definition further on to assess the usability of the ALFANET system.

Reference	Definition				
Cuda	Accomplishing goals effectively and efficiently	Easiness new users	Easiness Experienced users	Preferation	
Usability.gov	Error frequency and severity	Ease of learning	Efficiency of use	Subjective satisfaction	Memorability
Usabilitynet.org ISO 9241-11	Effectiveness		Efficiency	Satisfaction	

4.1.2 Objectives of usability assessment

When we use the ISO 9241-11 definition, the objectives for 'usability assessment' of the ALFANET system are the following:

"*Assessing the extent to which the ALFANET system can be used by*

- *specified users: KLETT and EDP authors, students and tutors*
- *to achieve specified goals:*
  KLETT: developing and learning Spanish language course/
  EDP: developing, tutoring and learning Solid Waste Management
- *with effectiveness, efficiency and satisfaction*
- *in a specified context of use: KLETT: self study EDP: in company training"*

4.2 Context

There are 3 different prototypes of the ALFANET system planned (see 2.1 Evaluation Framework). Usability assessment can be conducted at these 3 prototypes. We expect that usability assessment has the most important function at the first and second prototype.

Besides usability assessment itself, development of the instrumentation and analyses of the results of the usability assessment have to be planned.

Developing instrumentation	*When*	*Responsible*
1st prototype (survey)	February-September 2003	OUNL
2nd prototype (questionnaire)	October 2003-May 2004	OUNL
final system (questionnaire)	June 2004-January 2005	OUNL

Usability assessment	*When*	*Responsible*
1st prototype	After month 17 (September 2003)	KLETT/EDP
2nd prototype	After month 25 (May 2004)	KLETT/EDP
final system	After month 33 (January 2005)	KLETT/EDP

Analyses usability Assessment	*When*	*Responsible*
1st prototype	January-February 2004	UNED/KLETT/EDP
2nd prototype	September-October 2004	UNED/KLETT/EDP
final system	March-April 2005	UNED/KLETT/EDP

4.3 Methodology

To select the most appropriate method to assess usability a 'Methods Table' on http://usabilitynet.org/tools/methods.htm gives some guidance. You can select the most appropriate method depending on three conditions:

- Limited time/resources
- No direct access to users
- Limited skills/expertise

When we apply this method table to the usability assessment of the ALFANET system, we have the following conditions:

- Limited time/resources
- Direct access to users
- No limited skills/expertise

The following methods are suggested:

- Diagnostic evaluation: User based evaluation of a working system, where the primary objective is to identify usability problems. Benefits:
  - Major usability problems are identified
  - An understanding is gained of why the user has difficulties with the system
  - Approximate measures can be obtained for the users' effectiveness, efficiency and satisfaction.
- Subjective evaluation: Subjective assessment tells the evaluator how the users feel about the software being tested. This is distinct from how efficiently or effectively they perform with the software. The usual method of assessment is to use a standardised opinion questionnaire to avoid criticisms of subjectivity. Benefits:

- In a discretionary use scenario, user satisfaction is most probably the largest single key factor, which will influence the users' decision whether or not to continue with the software (other key factors may include price, technology, and brand loyalty).
- In a mandatory use scenario, poor satisfaction leads to absenteeism, fast staff turnover, and unrelated complaints from the workforce.
- Subjective Assessment complements data from efficiency and effectiveness measures.

As we have chosen to assess the effectiveness, efficiency and satisfaction of users with the system, diagnostic evaluation seems to be the most appropriate way to do this. We suggest two different ways to measure the usability of the ALFANET system. To assess the usability of the first prototype we suggest collecting qualitative data by means of surveys. One of the reasons to collect qualitative data at the first stage is that the greatest problems with the usability of the ALFANET system are expected at this stage. As we expect that the usability problems during the second prototype and the final system are less, we suggest assessing the usability of the second and final prototype by means of questionnaires. Summarized:

- Surveys:
  - in the early stage (first prototype)
  - to gather mainly qualitative information
  - structured interviews are conducted.
- Questionnaires:
  - second and final prototype
  - questionnaires
  - more quantitative data

Surveys and questionnaires for different users: authors, students and tutors have to be conducted.

4.4 Usability indicators

Below some indicators of effectiveness, efficiency and satisfaction are given:

Effectiveness

- Can users (authors, students, tutors) complete tasks, achieve goals with the product, i.e. do what they want to do?
- Etcetera …

Efficiency

- How much effort do users (authors, students, tutors) require achieving their goals? (Often measured in time)
- Quickness and accurateness
- Designed to reduce user errors and to give users (authors, students, tutors) easy ways to recover from any errors they do make?
- Ease of learning: How fast can a user (authors, students, tutors) who has never seen the user interface before learn it sufficiently well to accomplish basic tasks?
- Efficiency of use: Once an experienced user (authors, students, tutors) has learned to use the system, how fast can he or she accomplish tasks?

- Memorability: If a user (authors, students, tutors) has used the system before, can he or she remember enough to use it effectively the next time or does the user have to start over again learning everything?
- Error frequency and severity: How often do users (authors, students, tutors) make errors while using the system, how serious are these errors, and how do users recover from these errors?
- Etcetera …

Satisfaction:

- How much do users (authors, students, tutors) prefer the system to other systems?
- Are users (authors, students, tutors) feeling productive and satisfied when they're done?
- Subjective satisfaction: How much does the user (authors, students, tutors) like using the system?
- Etcetera …

Appendix 4 Usability assessment measures provides an initial list of usability indicators to be assessed by questionnaire to the different users, providing measures of utility and easy of use.

4.5 Design and sampling

We distinguish three different groups from whom information on usability is to be collected:

- Authors
- Learners
- Tutors

And this on two different pilot sites: EDP and KLETT. We suggest that all the authors, learners and tutors participate at the usability assessment.

Usability assessment takes place at three different moments:

- First prototype
- Second prototype
- Final system

4.6 Presentation and evaluation of the results

Besides reporting the results of the usability assessments in formal reports after the first, second and final prototype, quick feedback to developers is also necessary and important so they can adjust the prototypes when usability problems arise.

5. Effectiveness Assessment Plan

5.1 Objectives

The overall goal of the Effectiveness Assessment Plan is:

> *"To assess the degree to which ALFANET key features can improve the effectiveness of the e-learning process"*

Two new questions arise to address this goal.

Question 1. How can we measure the effectiveness of the e-learning process? How can we focus on the effectiveness provided by the tools included in our platform and isolate the other important environmental factors that contribute to the learning effectiveness?

Question 2. What are the key features of ALFANET? What makes ALFANET innovative and different from other e-learning tools in the market?

The first question has been addressed by many researches ([Heinecke, 1999], [Hiltz, 2000], [CapitalWorks, 2000]), and we find a very complex arena with a long way in run.

We propose to measure the effectiveness of the learning process as a combination of:

- Objective quality measures of users' performance, provided by the system.
- Subjective opinion of users (Learners satisfaction and tutors perception)

We will go in more detail about the Effectiveness measure in section 5.3.1 How can we measure the learning effectiveness?

To answer the second question, it is obvious that the main key innovation of ALFANET is the *Personalised e-learning by means of the adaptive capability (*see Annex 1 - "Description of Work" from ALFANET Project contract (IST 2001-33288), section 5.1). Prior to analyse in detail the Adaptive feature we can also differentiate other key features leveraged in pedagogical or operational aspects:

Pedagogical Level. ALFANET promotes a learning based on a learner-centered experience where the learner autonomy and activity is reinforced, where interchange between learners, tutors and other resources agents are promoted and where the tutor supports the educational process. The learners are **active**ly involved in an adaptive **collaborative** learning scenario on-line.

From this point of view we could assess the importance of these pedagogical dimensions of learning in the context of ALFANET project.

Dimension 1. Active vs Passive learning

Dimension 2. Collaborative vs Individual learning

Dimension 3. Supported vs Unsupported learning

Dimension 4. Guided by Instructional Design vs Flexible & Open learning

Operational level: there are very concrete operational issues obtained from D11. User Requirements. The Appendix A 6.1 Key Issues to Measure Effectiveness Assessment, provides a set of them.

In the context of the present document only the Adaptive feature will be addressed. With the project advance new key features can be afforded.

Next, we go in more detail about the Adaptive feature.

The aim of **adaptive learning** is to support the learner's learning process by individualising the learning event, dynamically adjusting services provided, content delivery and varying the sequence of the learning activities as the learner proceeds through the learning environment.

Taking into account what the D11. User Requirements stresses about the learner control of their learning experience, the outcomes of the adaptation will be always provided to the learners through "Learning Recommendations" that the learner can follow or not, maintaining his/her total autonomy. These recommendations are generated based on adaptation knowledge (automatically learned by machine learning methods or pre-specified in the system).

The objective of this plan is to address the issue of the importance of adaptation in the success for improving the self-learning for work.

Our main hypotheses is stated as follows:

> *"The recommendations provided by ALFANET are effective when the learners follow them (which are dependent on their individual differences) and their learning experience through the e-learning platform enhances."*

This main hypotheses will be refined in further project stages, for the next version of the document. Different recommendation types and adaptive features could be analysed separately, producing more concrete hypotheses.

Apart from the general analysis of adaptability we intend also to obtain measures of how effective are some type of the adaptation features provided:

- Automatically learned by the system vs pre-specified
- Each one of the adaptation features or by categories (presentation, navigation, collaborative, ...)

  From this analysis we can obtain that adaptive collaborative support has a greater impact on effectiveness than adaptive presentation support

5.2 Context

From the three different prototypes of the ALFANET system planned (see 3.1 Evaluation Framework) effectiveness assessment can be conducted at the second prototype and in the final system. We expect that effectiveness assessment has the most important function at the final system, when all adaptive features are incorporated.

From the tests performed in the first prototype, the ALFANET developers will profit from all the data captured by the system in order to afford the technological evaluation of user model and machine learning techniques.

UNED leads the overall evaluation plan, which implies to analyse the tests performed at the pilot sites extracting the quantitative and meaningful measures of the effectiveness of the e-learning process with ALFANET.

EDP and KLETT will organise the experiments at each one of their pilot sites.

Besides effectiveness assessment itself, development of the instrumentation and analyses of the results of the effectiveness assessment have to be planned.

5.3 Methodology

From the analysis of existing literature about the Evaluation in the Educational area ([Chin, 2001], [Langley, 1998], [Reeves, 2000], [Paramythis, 2001], [Heinecke, 1999], [Hiltz, 2000], [CapitalWorks, 2000]), we obtain the following main conclusions:

- We realize the high complexity of educational technology and their effectiveness evaluation.

- Evaluation methods has undergone a major transformation in the last decades. It has changed from monolithic to pluralist conceptions, to multiple methods, multiple measures, multiple criteria, multiple perspectives, multiple audiences, and even multiple interests. Methodologically, evaluation moved from primary emphasis on quantitative methods, in which the standardized achievement test employed in a randomised experimental control group design was mostly highly regarded, to a more permissive atmosphere in which qualitative research methods were acceptable
- In order to get at the complexities of the learning process multiple measures (quantitative and qualitative) will be used. Evaluations should not focus on simple outcomes measures such as posttests but should also focus on complex metrics describing the learning process; we will adopt multifaceted approaches to evaluation, reduce the reliance on standardized test scores as the primary evaluation outcome. Evaluation designs will rely less of participants self reported attitudes and more on observations of participants actions within learning contexts.
- Although we can define student learning as the retention of basic skills and content information as reflected on norm referenced and criterion referenced standardized tests, educational technology benefits include also preparing students for jobs, increasing student interest in learning, increasing student access to information and making learning an active experience
- It is very important to have a stronger description of what is the technological innovation (in our case adaptation); we must invest time documenting what adaptive features will be provided.
- Not less important is to conduct implementation evaluations prior to outcomes evaluations. We will perform an exhaustive Technical Verification of the Adaptive Modules to assure that adaptation has been fully implemented and provides a solid base to generate users' recommendations before trying to determine its effectiveness.
- The Evaluation Model most commonly used in the Educational area is the Kirkpatrick's 4-Level Model that is described in Appendix A5.1 The Kirkpatrick Model for Summative Evaluation. Our evaluation approach is focused on level 1 Reaction and level 2 Learning of this model.
- The evaluation of adaptive educational systems is commonly based on Empirical Evaluations (see Appendix A5.2 Empirical Evaluation)

Modular Assessment of the adaptation cycle

The evaluation of adaptive systems should not treat adaptation as a "monolithic" process, rather, adaptation should be "broken down" into its constituents, and each of these constituents should be evaluated separately where necessary and feasible [Paramythis, 2001]. Thus, in the scope of the Technical Verification of the Adaptive Modules, we will conduct an exhaustive modular assessment of each one of the adaptation stages in order to assure that adaptation as been fully implemented and that it provides a solid base before trying to determine its effectiveness.

The individualisation of the learning is based on the learner's individual differences: knowledge level, interest, the performance history, etc. All them constitute the User Model.

The adaptation in the system is done through user-model acquisition from the student data available and interaction with the system. In other words, the adaptation is based on capturing a lot of data from the user in order to define the user-model, which represents the user's knowledge state, preferences and goals. The system adapts automatically through the use of machine learning techniques and modifies students and content models in accordance with the student's interactions.

Firstly, we must assure the **accuracy of the User Model.** This stage will include the evaluation of:

- the architecture of the modelling which will be built with standard data.
  We will assess how different User Model characteristics affects the adaptation. The results of the assessment can include which ones are the more relevant characteristics for adaptation.
- the data obtained by the system (directly captured and throw machine learning techniques) that constitute the model of each learner are considered correct by s/he.

Next we will assess the **accuracy of the Adaptation Model** and the machine learning techniques, in order to assure that the adaptation knowledge learned by the system are correct, or at minimum to obtain a positive valuation by experts of the Instructional area.

Adaptive Tasks

We want to obtain a global assessment on how adaptation affects the effectiveness of the learning process.

We can demonstrate our main hypotheses: "*The recommendations provided by ALFANET are effective when the learners follow them (which are dependent on their individual differences) and their learning experience through the e-learning platform enhances*"

The methodology to do it will be based on empirical evaluation (see Appendix A5.2 Empirical Evaluation); different versions of the system will be tested: one version with adaptive features and the other one without them. After running the experiments we will take measures (see Section 5.4.2 Indicators for Effectiveness Measurement) about the use of the system with the aim to obtain a measure of the effectiveness of each group. We can conclude about ALFANET effectiveness by comparison between results of groups with and without adaptive features.

A deeper analysis will be done analysing also the **Acceptance of Adaptability**: if the learners accept or no the recommendations provided by the system.

Considerations for Experimental Design

The following factors that **favour adaptability** will be also taken into account when preparing each one of the pilot sites, (including educational material, instructional design and recruitment of learners), to profit from the key evaluation features:

- Learners with diversity of needs
- Instructional design offers different routes through courses where learners have different interests and knowledge levels.
- Educational Material characterised with different levels of difficulty; Educational Material with a rich conceptual map and inter-relationship.
- Changes in the course environment

To measure the user acceptance and impact analysis, the parameters will be quantitative and qualitative:

- Quantitative items will consist in measures obtained from the system and questionnaire surveys.
- Qualitative items will take into account less structured ways of obtaining opinion from the users

5.3.1 How can we measure the learning effectiveness?

When we talk about effectiveness in general we talk about to achieve what is intended. In the context of our project, the measurement of the effectiveness of the learning process is a very complex problem, which is a hot issue in the educational evaluation field.

There exists a lot of factors that contribute to learning effectiveness, and that are not the concern of our project. From the literature we can identify the following elements:

- Quality of educational materials (Course Structure)
- Quality of instruction
    - Pedagogical model: instructional design
    - Educational & Technological services provided
- Quality of learner (in special Motivation)
- Adequacy of learner characteristics with contents, instruction.

It will be difficult to isolate all these confounding factors that can affect our study, but we must be conscious of them in order to fix them as possible, and identify the advantages of our platform.

Some intends are made in order to define a Learning Effectiveness Index [CapitalWorks, 2000]; it will be desirable but utopic to define a magic formulae with the measure in which each one of these issues contribute to making the learning process effective. This is not our key issue, we consider such area interesting but it is not our current focus.

The topic we deal now is how can we measure the effectiveness of a learning process.

Experimental studies in nearly every discipline concern some form of behaviour. This means they require some measure of that behaviour to play the role of a dependent variable in the experiment, that is, the variable that is affected by the experimental manipulations.

The analysis of available (current and actual) literature related to the evaluation of adaptive educational systems [Langley, 1998] provides us four general types of measures that seems appropriate for our study:

Measures of Efficiency (adaptive features will contribute to accomplish the platform objectives more rapidly and with less effort ?):

- time the user takes to complete his interaction
- effort the user must exert; this metric concerns the number of user actions that occur during solution of a giving problem; typically is measured by number of keystrokes or mouse clicks

Measures of Quality (adaptive features will contribute to improve the quality of the learning process?)

- objective measure quality for a domain; in our educational domain quality measures of learner outcomes, such us personal aptitudes and acquired knowledge

Measures of User Satisfaction (reliance on some separate measures of user satisfaction to determine the quality of the system's behaviour):

- present each user with a questionnaire that asks about their subjective experience
- find out if a person will continue to use the system
- giving the user control over the use certain systems features detect whether the user actively selects them.

Measures of Predictive Accuracy (because the user model in an adaptive interface makes predictions about user responses to the system's advice, it is important to rely on predictive accuracy as a surrogate measure for efficiency and quality. Moreover, accuracy is the most widespread measure in machine learning):

- percentage of recommendations positively followed by users.
- frequency and accuracy of predicted actions.
- accuracy of machine learning tasks
- correctness of learner's model

The consortium general view is that objective quality measures from the educational field (grades, results of self-assessment, unit-of-learning and/or course tests) will not be very much affected by the adaptive features. Furthermore, the consortium supposes that no significant differences could be initially appreciated from the two experimental groups, because the influence of the above mentioned issues will be much more important than the adaptive features. However, the study will take into account also these quality measures.

Efficiency messures are not the focus of our effectiveness study, it is not the objective of our ALFANET platform to accomplish the learning process more rapidly and with less effort. Nevertheless, these measures have been initially included as part of our study, due that are easily provided by the system, and can be of interest for researching the impact of adaptive features on the systems' efficiency.

The opinion of the users will be the most relevant measures provided by the experimental study to determine the quality of the system's behaviour, but these measures will be complemented with the other objective measures.

As summary, we can identify a set of performance indicators that address the issue of learning effectiveness.

- Outcomes of the learning process: results of self-assessment, unit-of-learning and/or course tests.
- Activity rates, Progress and Completion rates addressing issues as "are people moving through the course?" Have they started? The activity rate increases along time? What "percent completion" have they achieved?
  All these measures would probably reported by the Auditing tools and tracked at every chapter, lesson, interactivity, or assessment for a learner and for the entire group.

- Perception, Subjective opinion of users (Learners satisfaction and tutors perception) To what extend did the adaptive features contribute to effectiveness of learning (for the group with adaptive features) or would adaptive features have contributed to the effectiveness of learning (for groups without adaptive features).

  For example we can ask the perceived learning by the participating students

  1= I learned more than I expected

  2= I learned as much as expected

  3= I learned less than I expected

  4= I learned nothing

This list of indicators is only an initial reference to be considered. Section 5.4.2 Indicators for Effectiveness Measurement provides a more exhaustive analysis of these indicators. They will be refined in further project stages, for the next version of the document.

5.4 Expected Impacts and Indicators

5.4.1 Key Issues to Measure Effectiveness Assessment

D11 “User requirements” provide issues that directly or indirectly are expected to positively influence the effectiveness of the learning process. A list of these issues is contained in Appendix A 6.1 Key Issues to Measure Effectiveness Assessment. For the time being we exclusively refer to issues influencing adaptation derived from D11 “User requirements”. The items of this list needs to be revised at a later stage of the project when it is clear, which of the requirements are implemented in the ALFANET system and when all functionalities related to adaptation are defined.

In the following we describe two quantitative measurements to be carried out: Chosen are issues that either can be measured by data provided by the system, and/ or can be rated via questionnaires for users calling up how learners/ tutors judge the impact of a certain issue on the effectiveness of the learning process / tutoring. The questionnaires will be designed according to the Likert scale. Depending on decisions about methodology other additional measurements may be possible.

5.4.2 Indicators for Effectiveness Measurement

Key issues derived from “User requirements” will be measured quantitatively. Data will be taken by

- checking what the system recorded and translated into recommendations for the learner (provided either automatically by the system or by a tutor). It will also be checked to what extend learners accepted the recommendations of the system. Also it will be checked what information the system provided for tutors and to what extend tutors react as a follow-up of the data provided.
- a questionnaire for every user group (learners, tutors) asking for the personal evaluation, for every issue. On the side of the learners it will be asked: How do you rate its impact on the effectiveness of your learning progress? On the side of the tutors it will be asked to what extend the data provided by the system eased their tutoring of the course.

Additionally learners and tutors can be asked about their opinion: What issues they think to be helpful in the learning process, what they don’t like and what they would like to have additionally.

The following issues to be measured either by counting incidents carried out by the system or by gathering ratings via questionnaires refer exclusively on data taken from D11 “User requirements”. In a later stage of the project , in D61 Evaluation Plan, they need to be reviewed and adapted to the current state of the system.

Ad 1. Measurement to be taken from the system (counting of incidents)

The statements from user requirements were firstly translated into facilities that the system will provide .Then the incident has to be defined whose occurrence can be counted ("How often does .... ?")

The issues will be classified according to the user groups of learners and tutors.

Corresponding lists are contained in the Appendix A 6.2 Indicators for Effectiveness Measurement.

Ad 2 Questionnaire

The questionnaire will ask for agreement/ disagreement to a number of statements. Therefore the statements from "User requirement" need to be translated into sentences calling up agreement or disagreement (for learners: "... contributes to my effectiveness of learning", for tutors: "... contributes to my ease of tutoring the course"). The rating will be according to the Likert scale in 5 steps (fully agree, agree, indifferent, disagree, fully disagree).

The issues again will be classified according to the user groups of learners and tutors. Corresponding questionnaires for learners and tutors, derived from D11 are contained in the Appendix A 6.3 Questionnaires for Learners and Tutors.

5.5 Experimental design and sampling

Concerning effectiveness there will be two groups from whom information must be collected:

- Learners
- Tutors

The effectiveness assessment will be done in both pilot sites: EDP and KLETT.

To evaluate effectiveness we will use two different instruments:

- System indicators
- Questionnaires

The way effectiveness is measured is through the existence of control groups.

The effectiveness assessment, due to its characteristics, will only take place in the second prototype and in the final system.

5.6 Presentation and evaluation of the results

The results presentation of the effectiveness assessments will be made in formal reports after the second prototype and after the final system.

In order to correct eventual problems that can occur, the developers will also be inform, so they can take the appropriate corrective actions.

6. Summary of Timelines, Yardsticks and Milestones

This section describes the main activities to be performed during the evaluation of ALFANET, in the context of the project milestones. The purpose of this section is just to show a global view of the different dimensions of the evaluation activities, sequenced in the time and assigned to the responsible partners.

Task / Subtasks Description	Responsible Partners			Dates
	Global T61	Pilot 1 T62, T63	Pilot 2 T64, T65	
To refine the Evaluation Plan (Contributions for D61)	UNED			July 03
To describe in more detail ALFANET adaptive features To refine indicators, mainly to refine the effectiveness plan Specification of resources • To define the requirements of the pilot courses in order to profit from the key evaluation features • To define the requirements for configuring the participants in experimental groups				
To prepare Pilot Sites		KLETT	EDP	August 03
• To prepare each one of the pilot courses, including educational material and instructional design. • To select the participants: learners, tutors and experimenters; configure the different teams of learners on each pilot • To prepare the environment and the resources to carry out the experiment (prototype installation, course installation and publishing, users registering and system configuration) • To prepare in advance the experimental sessions (preliminary information to provide learners and tutors, recommendations to profit interaction, collaboration and adaptation, scheduling of courses and sessions: estimated time, configuration of activities to do in a typical session).				
EDP Timelines:				
Objectives and contents identification			EDP	March 03:
Storyboard definition			EDP	April – May 03
Content's development including the acceptance content tests			EDP	June – July 03
Developing Usability instrumental	OUNL			Feb.-September 03
Technical Verification	**(T61)**			July-September 03
Existing Software Edubox OpenACS Developed Software Modules: Authoring Tool Technical Verification of Adaptive Modules. Detailed Plan ALFANET system and Integration	 OUNL UNED Each Partner ACEBNet UNED SAGE			

Task / Subtasks Description	Responsible Partners			Dates
	Global T61	Pilot 1 T62, T63	Pilot 2 T64, T65	
First Prototype	**M2**	**(T62)**	**(T64)**	**September 03**
Installation of the 1st prototype in Pilot sites		KLETT	EDP	September 03
Preparation and deployment of the course environment (configuration and administration)		KLETT	EDP	October 03
Support to preparation, installation and deployment		OUNL	SAGE ACEBnet	September-October 03
		(T63)	**(T64)**	
Test Execution (**usability surveys**) First Introduction to participants		KLETT	EDP	October-December 03
Usability Test Analysis	UNED (T63,T65)	KLETT	EDP	January - February 04
Compilation of results from both pilot sites (D62)	KLETT (T63)	KLETT	EDP	April 04
Developing instrumental Usability Effectiveness	(T61) OUNL UNED			October 03-April 04
Technical Verification	(T61) Each Partner SAGE			October 03-April 04
Second Prototype	**M3**	KLETT	EDP	May 04
Test Execution (**usability**, effectiveness)		KLETT	EDP	June - November 04
Test Analysis	UNED (T63,T65)	KLETT	EDP	September-October 04 October-November 04
Compilation of results from both pilot sites (D63)	EDP (T65)	KLETT	EDP	December 04
Developing effectiveness instrumental	UNED			July - September 04
Technical Verification (updated versions)	(T61) Each partner SAGE			June - November 04
Final System	**M4**	KLETT	EDP	January 05
Test Execution (**effectiveness**, usability)		KLETT	EDP	January - February 05
Test Analysis	UNED (T63,T65)	KLETT	EDP	February – March 05
Compilation of results from both pilot sites (D64)	UNED (T63,T65)	KLETT	EDP	March 05
Final Evaluation Results (D66)	UNED			April 05

Yardsticks to reach at each one of the project milestones will be defined in next version of this document.

7. Presentation and evaluation of results

We propose to obtain a big number of data from the tests in order to perform a serious statistical analysis that provide us realistic conclusions.

We must define in advance the formulae to compute impacts from indicators and the thresholds for decision making (reference patterns)

We will achieve a final conclusion concerning the impact of the adaptive features, in the learning effectiveness of the ALFANET system.

But also a lot of secondary measures also transcendent.

After the verification and demonstration stages, the results will be presented in brief tabulations, charts and figures to show the results in the most useful way. The key results will stand out clearly. Any numerical result that is a population estimate will be expressed with the appropriate confidence interval and confidence level.

For the evaluation of results it will be taken into account the following topics:

- An explanation for any deviations at verification and demonstration stages from the plan will be provided.
- An evaluation of the effect of such deviations on the verification and demonstration results will be presented.
- An evaluation of the verification and demonstration results against the criteria confirmed in the Final Evaluation Plan will be presented.
- A confirmation of the assessment objectives and evaluation methods will be presented.
- A contribution of evaluation results to the exploitation plan will be presented.

Appendix 1. References

Chin, D. N. (2001). Empirical Evaluation of User Models and User-Adapted Systems. *User Modelling and User--Adapted Interaction, 11 (1-2),181-194,. Kluwer Academic Publishers.*

CapitalWorks (2000) multi-client study in "Developing and Applying a Learning Effectiveness Index" (including Cisco, IBM Global Services, Nortel Networks, Anexsys, EDS-Centrobe, and the U.S. Department of Veterans Affairs (Veterans Health Administration).

Langley, P., & Fehling, M. (1998). The experimental study of adaptive user interfaces (Technical Report 98-3). *Institute for the Study of Learning and Expertise, Palo Alto, CA.*

Reeves, Thomas C. (2000). Enhancing the Worth of Instructional Technology Research through 'Design Experiments' and Other Development Research Strategies. *Paper presented at American Educational Research Association Conference,* April 2000

Paramythis, A., Totter, A., & Stephanidis, C. (2001). A modular approach to the evaluation of adaptive user interfaces. *In S. Weibelzahl, D. N. Chin, & G. Weber (Eds.), Empirical Evaluation of Adaptive Systems. Proceedings of workshop at the Eighth International Conference on User Modeling*, UM2001 (pp. 9--24). Freiburg.

Heinecke, W. F., Blasi, L., Milman, N., & Washington, L. (1999). New directions in the evaluation of the effectiveness of educational technology. Whitepaper presented at The Secretary Conference on Educational Technology, Washington, D.C. Retrieved July 27, 2002, from

Hiltz, S.R.; Coppola, N.; Rotter, N.; Turoff, M. and Benbunan-Fich, R. (2000) "Measuring the Importance of Collaborative Learning for the Effectiveness of ALN: A Multi-Measure, Multi-Method Approach." *Journal of Asynchronous Learning Networks*, Vol. 4. Issue 2.

Appendix 2. Testing procedure

Next, it is included the description of the testing procedure done within the "Software Engineering Guideliness" document.

A 2.1 Performing the ALFANET Testing Plan

Requirements:

- Software modules must be tested according with a predefined plan, before integration.
- It will be necessary to make integration tests and the support documentation will be generated according to a determined strategy.
- It will be necessary to check that a strategy of integration of the software components has been determined and that is followed.

Purpose:

This procedure describes the way and responsibilities in performing tests of the different Software Units.

Roles and Responsibilities:

The roles identified for Testing Procedure are the following:

- Software Test Team: person or team from a partner that test the software keeping trace of it. The partner delivering the changed unit is responsible of its testing.
- Upper Level Partner: The partner that receives the code is responsible of integrating it with the rest of the elements which are related to it, but not of the actual validation of the received element.

Entry criteria:

This procedure applies from the very beginning of implementation-evaluation process in ALFANET project, before the production of the first prototype that corresponds to WP4, WP5 and WP6.

All Software Units will be thoroughly tested and validated whenever new actualisation or versions of the elements are delivered.

Inputs:

- Software version produced and documented.
- Test Plan.

Steps:

1. Each SDP is responsible of the proper tests of their software units, according to the methods they assume as valid for the test.
2. Associated to these private tests, a standard test procedure will be considered. Test will be documented indicating the operation/functionality, the expected results, the input data used and the obtained results.
3. Perform the test plan according with the test level (module, sub-component or component level). And to provide proper feedback to the Software Unit Developer.

Outputs:

- Document of test produced, the operation/functionality, the expected results, the input data used and the obtained results.

Exit criteria:

- Once the integrator partner has passed the test, the out log is obtained for the proposed input

Appendix 3. Definitions of usability assessment

http://www.csulb.edu/centers/cuda/ : Centre for Usability in Design and Assessment (CUDA), California State University, Long Beach

Usability is a measure of the quality of a user's experience interacting with a product or web site. It involves users' ability to do what they want and need with the product or site. Four main questions about the usability of a product or web site can be posed:

1. How effectively and efficiently can users accomplish their goals? Usable products and sites:
   - support users in accomplishing their goals
   - improve productivity by allowing users to do so quickly, efficiently, and accurately
   - are designed to reduce user errors and to give users easy ways to recover from any errors they do make.
2. How easy is it for new users to learn how to use the product or site? Usable products and sites:
   - allow new users to get up to speed and accomplish meaningful work in a relatively short period of time
   - match the way in which users approach their tasks
   - are easy to remember
   - require less training
   - allow users to return to peak levels of performance more quickly after absences.
3. How easy is it for experienced users to use the product or site? Usable products and sites:
   - are designed to require a minimum number of keystrokes or mouse-clicks
   - allow experienced users to customize actions and streamline processes to maximize their efficiency
   - present information in ways that support rather than hinder decision making
   - allow users to move easily from one task to another
4. How much do users prefer the system to other systems? Usable products and sites:
   - leave users feeling productive and satisfied when they're done
   - can reduce staff turnover and absenteeism by making users feel that they are adequately supported in their jobs.

http://usability.gov/

Usability is the measure of the quality of a user's experience when interacting with a product or system — whether a Web site, a software application, mobile technology, or any user-operated device. Usability is a combination of factors that affect the user's experience with the product or system, including:

- Ease of learning:

How fast can a user who has never seen the user interface before learn it sufficiently well to

accomplish basic tasks?

- Efficiency of use:

Once an experienced user has learned to use the system, how fast can he or she

accomplish tasks?

- Memorability:

If a user has used the system before, can he or she remember enough to use it effectively?

The next time or does the user has to start over again learning everything?

- Error frequency and severity:

How often do users make errors while using the system, how serious are these errors, and

How do users recover from these errors?

- Subjective satisfaction:

How much does the user like using the system?

http://usabilitynet.org

Usability means making products and systems easier to use, and matching them more closely to user needs and requirements.

The international standard, ISO 9241-11, provides guidance on usability and defines it as:

The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

Usability is about:

- Effectiveness - can users complete tasks, achieve goals with the product, i.e. do what they want to do?
- Efficiency - how much effort do users require doing this? (Often measured in time)
- Satisfaction – what do users think about the products ease of use?

….which are affected by:

- The users - who is using the product? e.g. are they highly trained and experienced users, or novices?
- Their goals - what are the users trying to do with the product - does it support what they want to do with it?
- The usage situation (or 'context of use') - where and how is the product being used?

Usability should not be confused with 'functionality', however, as this is purely concerned with the functions and features of the product and has no bearing on whether users are able to use them or not. Increased functionality does not mean improved usability!

Appendix 4. Usability assessment measures

All the Items to be assessed by questionnaire

A 4.1 Authors

INDICATORS
Ease of:
Recycling of learning units
Reutilization of learning units
Reutilization of objects
Use of authoring tool

A 4.2 Learners

INDICATORS
Ease of:
Interaction with the contents
Navigation in the system
Access to the objectives of each unit of learning
Access to the course index
Access to tests answers
Access to up-to-data information
Use of Personal Storage Area
Use of e-mail
Use of Discussion Forums
Use of Brainstorming Area
Use of Chat room
Use of Videoconference
Use of Virtual Cafeteria
Use of instant messenger applications
Use of Newsgroups
Use of Virtual Seminars
Use of Bulletin Board
Use of shared navigation
Use of shared workspaces
Use of shared favourite links
Use of shared whiteboard
Research in Virtual Library

Personalization of the learning area
Utility of:
Route's access to the progress on the course
Complementary contents
Auto-evaluations tests to the progress on the course **Adaptive characteristic**: Whenever you feel a need to have an auto-evaluation did the system respond immediately?
e-mail
Discussion Forums
Brainstorming area
Chat room
Videoconference
Virtual Cafeteria
Instant messenger applications
Newsgroups
Virtual Seminars
Agenda of events
Bulletin Board
Virtual Library
Shared navigation
Shared workspace
Shared favourite links
Shared whiteboard
Self-assessment exercises
Personalized area

A 4.3 Tutors

INDICATORS
Ease of access to all the learner's information provided by the system (test answers, fulfilled objectives, contacts, FAQ's, etcetera)
Utility of access to all the learner's information provided by the system (test answers, fulfilled objectives, contacts, FAQ's, etcetera)

A 4.4 Managers

INDICATORS
Ease of access to the tutor's and learner's information provided by the system (fulfilled objectives, contacts, etcetera)
Utility of access to the tutor's and learner's information provided by the system (fulfilled objectives, contacts, etcetera)

Appendix 5. Educational Evaluation

There exists different Evaluation Models in the research literature of the Educational area. The most commonly used is the Kirkpatrick's 4-Level Model that is described above. Our evaluation approach will be focused in level 1 Reaction and level 2 Learning of this model. Other educational evaluation models considered are:

- **Formative** Evaluation: Focuses on improving the online learning experience.
- **Summativ**e Evaluation: Focuses on the overall success of the OL experience (should it be continued?).
- **CIPP** Model Evaluation: Framework of Context, Input, Process, and Product.

A 5.1 The Kirkpatrick Model for Summative Evaluation

In 1975, Donald Kirkpatrick first presented a four-level model of evaluation that has become a classic in the industry:

- Level One: Reaction
- Level Two: Learning
- Level Three: Behaviour
- Level Four: Results

These levels can be applied to technology-based training as well as to more traditional forms of delivery. Modified labels and descriptions of these steps of summative evaluation follow.

Level One: Students' Reaction

In this first level or step, students are asked to evaluate the training after completing the program. These are sometimes called smile sheets or happy sheets because in their simplest form they measure how well students liked the training. However, this type of evaluation can reveal valuable data if the questions asked are more complex. For example, a survey similar to the one used in the formative evaluation also could be used with the full student population. This questionnaire moves beyond how well the students liked the training to questions about:

- The relevance of the objectives.
- The ability of the course to maintain interest.
- The amount and appropriateness of interactive exercises.
- The ease of navigation.
- The perceived value and transferability to the workplace.

With technology-based training, the survey can be delivered and completed online, and then printed or e-mailed to a training manager. Because this type of evaluation is so easy and cheap to administer, it usually is conducted in most organizations.

Level Two: Learning Results

Level Two in the Kirkpatrick model measures learning results. In other words, did the students actually learn the knowledge, skills, and attitudes the program was supposed to teach? To show achievement, have students complete a pre-test and post-test, making sure that test items or questions are truly written to the learning objectives. By summarizing the scores of all students, trainers can accurately see the impact that the training intervention had. This type of evaluation is not as widely conducted as Level One, but is still very common.

Level Three: Behaviour in the Workplace

Students typically score well on post-tests, but the real question is whether or not any of the new knowledge and skills are retained and transferred back on the job. Level Three evaluations attempt to answer whether or not students' behaviours actually change as a result of new learning.

Ideally, this measurement is conducted three to six months after the training program. By allowing some time to pass, students have the opportunity to implement new skills and retention rates can be checked. Observation surveys are used, sometimes called behavioural scorecards. Surveys can be completed by the student, the student's supervisor, individuals who report directly to the student, and even the student's customers.

Level Four: Business Results

The fourth level in this model is to evaluate the business impact of the training program. The only scientific way to isolate training as a variable would be to isolate a representative control group within the larger student population, and then rollout the training program, complete the evaluation, and compare against a business evaluation of the non-trained group. Unfortunately, this is rarely done because of the difficulty of gathering the business data and the complexity of isolating the training intervention as a unique variable. However, even anecdotal data is worth capturing.

A 5.2 Empirical Evaluation

Empirical evaluation [Ching, 2001] refers to the appraisal of a theory by observation in experiments. The key to good empirical evaluation is the proper design and execution of the experiments so that the particular factors to be tested can be easily separated from other confounding factors. These factors, which are under the control of the experimenter, are termed independent variables because their values can be varied independently of other variables by the experimenter. Dependent variables are variables whose values depend on the values of other variables. They include response variables or recorded measures such as the frequency/extent of certain behaviours (e.g., system usage), qualities of a behaviour in a particular situation, number of errors, error rate, time to complete a task, pro-portion/quality of tasks achieved, interaction patterns, learning time/rate, and/or subjective evaluations (e.g., user satisfaction).

Some dependent variables are difficult to measure or can only be measured indirectly such as cognitive load measured through blood pressure or pupil dilation.

In an ideal experiment, only the independent variables are varied and everything else is fixed so that any changes in the dependent variables can be directly attributed to the variations in the independent variables. Unfortunately, such ideal control is almost impossible. If different participants are used for the different independent variable cases, then individual differences (intelligence, reading ability, spatial reasoning, perceptual abilities such as colour blindness, poor eyesight, poor hearing, knowledge, etc.) will typically influence the dependent variables much more than the independent variables. If the same participant is involved in all independent variable conditions, then there is the problem that the earlier conditions will affect the later conditions due to practice effects (e.g., a participant may be able to perform a task faster the second time around).

There are also potential problems with different times, locations, or other environmental conditions influencing the dependent variables.

To overcome such problems, participants are randomly assigned to groups in order to average out the effect of `nuisance variables' on the dependent variables.

Of course, in order for averaging to work properly, large numbers of participants are needed. Statistical techniques for analysis of variance (ANOVA) are used to determine whether differences in dependent variable values among groups are due to the different independent variable treatments or due to random fluctuations.

To improve the sensitivity of experiments (and thus reduce the number of participants needed), crossed designs use the same participants for multiple dependent variable conditions. For example, the same participant uses both the user-adapted system and the no-UM system. Crossed designs control for practice effects by varying the order of dependent variable conditions for different participants and participants are randomly assigned to the different orders.

Appendix 6. Effectiveness assessment measures

A 6.1 Key issues to measure effectiveness assessment

Key	Requirement	Measured by system	Measured by quest.
CC1.1	A time limit must be set for each unit of learning as well as a recommended duration		x
CC4.1	The content development tool must allow the existence of activities that constrain the learner's evolution in each learning unit		x
IL2.1	The objectives of each learning unit must be easily accessed by the learners	x	x
IL2.2	The learner must be able to access, at any moment, to all the information available, such as the course index, ...	x	x
IL2.3	The platform must allow learners to test their knowledge level during each learning unit	x	x
IL2.4	There must be a tool allowing the learners to perceive the evolution along the course	x	x
IL2.5	The system must allow learners to select the route that better fits their learning style		x
IL2.6	The system must help learners to select the units of learning needed to fulfil their objectives	x	x
IL2.7	The system must offer a list of possible complementary contents to be studied, that users can select	x	x
IL2.8	The system must allow evaluation tests answers available to learner's access		x
IL2.9	The system must allow on-line auto-evaluation available at learners' request	x	x
IL3.1	The system must propose to the learners, other learners that are in the same item of the course, in order to solve problems together	x	x
IL3.2	The system must provide a list of learners that are on-line and that have already solved the problem	x	x
IL3.4	In the case of wrong answers or failure in the proposed activities, the platform must allow and suggest the learners, the use of alternative learning activities	x	x
IL3.6	The system must allow up-to-date information, such as recent news you had not read yet, new messages in the forum, etc.	x	x
IL3.9	The platform must allow (propose) self-assessment exercises to check if the contents were appropriately learned whenever the system decides so	x	x
IL4.4	The system must have a virtual library explored by a state-of-the-art search engine	x	x
IL4.5	The system must have a list of Frequently Asked Questions and their answers	x	x
IL4.6	The system must have a Course Assistant tool (virtual tutor), giving some clues to help solving problems	x	x

IL4.8	The system must allow Private Favourite Links study facility	x	x
IL4.9	The system must allow an e-mail for all / for each fellow; Discussion Forums / Brainstorming area; Chat room; Videoconference services and a Virtual Cafeteria	x	x
IL4.10	The system must provide information through Shared Applications; Instant messenger applications; Shared Navigation; Shared Workspaces; Shared favourite Links; Shared Whiteboard (Notes); Newsgroups; Virtual Seminars; Agenda of Events and a Bulletin Board	x	x
T1.2	The system must provide indications about the participation of learners	x	x
T1.3	The system must have a tool allowing the tutor to easily see, for a group of learners, the objectives that were fulfilled in each learning unit of a course;	x	x
T1.4	The system must have a tool presenting for each learner, the number of accesses, the time spent and the evaluation test results in order to allow the tutor to assess the learner's performance		x
T1.5	The tutors must have access to the answers the learner gives to each question in an evaluation test	x	x
T2.3	The system must help the tutor to schedule and re-schedule the activities of the course		x
T2.5	The system must help the tutor in detecting useful information for learners, such as the most relevant didactic units, the most difficult concepts, the most frequently asked concepts, etc.	x	x
T2.8	The system must allow the tutor to adapt to each learner the standard structure of the course	x	x
T2.9	The system must allow tutors to decide what type of collaboration activities can be performed	x	x
T2.10	The system must generate groups and subgroups according to criteria specified by tutors and using information stored for each learner	X	

A 6.2 Indicators for Effectiveness Measurement

No.	To be measured on the side of the learners	Key of UR
1	*The system provides the learner on request with the learning objectives (LO) of every unit.* How often does the learner click to the LOs of the learning unit?	IL2.1
2 2.1 2.2	*The system provides the learner on request with several information, e.g. course index.* How often does the learner click to - *course index* - *...*	IL2.2
3	*The system creates and provides the learner with statistical data for every unit of exercises in a learning unit like* - *number/ percentage of exercises done (and also not yet done)* - *number/ percentage of right answers (and also of wrong answers)* How often does the learner click to the statistics to check his performance?	IL2.3 IL2.4

4	*The system provides the student with a list of his learning objectives (per learning unit) via a search engine according to LOs.*	IL2.6
4.1	How often does the student click to this kind of search engine?	
4.2	How often does he choose content related to this LO?	
5	*The system provides compulsory content as well as additional content.*	IL2.7
5.1	How often does the learner click to the list of additional content?	
5.2	How often does he choose to work on additional tasks?	
6	*The system allows the learner to ask on-line (the tutor or via forum) for auto-evaluation material.*	IL2.9
6.1	How often does the student do this request?	
6.2	How often does he apply this auto-evaluation?	
7	*The system provides learners (on request) with a list of other learners working on the items of the course.*	IL3.1
7.1	How often is this list asked for?	
7.2	How often does the learner successfully contact another learner?	
8	*The system provides the learner (on request) with a list of learners actually on-line that have already solved a certain problem.*	IL3.2
8.1	How often is this list asked for?	
8.2	How often does the learner successfully contact another learner?	
9	*In case of bad performance of an activity the system proposes the learner additional material.*	IL3.4
9.1	How often does this proposal occur?	
9.2	How often does the learner accept to apply this material?	
10	*The system provides the learner with information on recent news.*	IL 3.6
10.1	How often does the learner click into this list?	
10.2	How often does he open the news?	
11	*The system automatically provides self-assessment exercises.*	IL3.9
	How often does the learner apply these assessments?	
12	*The system offers a virtual library with a state-of-the art search engine.*	IL4.4
	How often does the learner use this facility?	
13	*The system offers a list of FAQs with answers.*	IL4.5
	How often does the learner use this facility?	
14	*The system provides a Course Assistant tool (virtual tutor).*	IL4.6
14.1	How often does it give advice to a learner?	
14.2	How often does the learner accept the recommendations?	
15	*The system provides the learner with a private favourite link study facility.*	IL4.8
	How often does the learner use this facility?	
16	*The system provides the learner with several collaborative facilities.*	IL4.9
16.1	How often does the student use the e-mail?	
16.2	How often does the student use the chat room?	
16.3	How often does the student use the brainstorming area?	
16.4	How often does the student use the discussion forum?	

16.5	How often does the student use the videoconference service?	
15.6	How often does the student use the virtual cafeteria?	
17	*The system provides several information facilities.*	IL4.10
17.1	How often does the student use the instant messenger application?	
17.2	How often does the student use the shared navigation?	
17.3	How often does the student use the shared workspace	
17.4	How often does the student use the shared favourite links?	
17.5	How often does the student use the shared whiteboard?	
17.6	How often does the student use the newsgroups?	
17.7	How often does the student use virtual seminars?	
17.8	How often does the student use the agenda of events?	
17.9	How often does the student use the bulletin board?	

From other point of view the system could provide the following main indicators from the learners; some of them are in bold, meaning that they include adaptive characteristics.

INDICATORS
Route Diversity Rate
Mean time by course/module/unit of learning (comparison between adaptive group and control group)
Mean time by unit of learning and by profile **(within adaptive group)**
Pre-assessment learning route acceptance rate
Complementary contents acceptance rate
Complementary contents acceptance rate, by profile
Successful contacts rate suggested by the system for solving problems with pears
Successful contacts between learners that are online
Number of adjustments to the foreseen route
Proposed alternatives activities acceptance rate
Number of wrong answers not subjected to alternative activities, suggested by the system (by learner)
Proposed composition of discussion / work groups acceptance rate
Number of self assessment tests proposed by the system versus number of self assessment tests requested by the learners
Virtual tutor suggestions rates vs Virtual tutor suggestions acceptance rate
Number of collaborative activities introduced by the tutor vs Acceptance rate of collaborative activities introduced by the tutor

No.	**To be measured on the side of the tutors**	**Key of UR**

18	*The system provides the tutor with data about the participation of every student.*	T1.2
18.1	How often does the tutor click on these lists?	
18.2	How often does the tutor react as a follow-up to these data?	
19	*The system provides the tutor with data of a learning group (or the individuals of this group) about the status of having fulfilled the objectives.*	T1.3
19.1	How often does the tutor click on these lists?	
19.2	How often does the tutor react as a follow-up of these data?	
20	*The system provides the tutor with data of every learner showing the* - *number of accesses* - *time spent* - *results of evaluation tests.*	T 1.5
20.1	How often does the tutor click on these lists?	
20.2	How often does the tutor react as a follow-up of these data?	
21	*The system provides the tutor with data of every learner (or learning group) showing* - *the most difficult concepts* - *the most FAQs* - *the most relevant didactic units (how to detect this by machine data?)* How often does the tutor click on these lists?	T2.5
22	*The systems allows the tutor to adapt the structure of the course to an individual learner.* How often is this facility applied?	T2.8
23	*The systems allows the tutor to allocate certain types of collaborative activities to certain students (or groups of students).* How often is this facility applied?	T2.9
24	*The system allows the tutor to generate and to change groups and subgroups of learners.* How often is this facility applied?	T2.10

A 6.3 Questionnaires for learners and tutors

No.	Questionnaire for learners	Key of UR
1.1	The time limit for every learning unit contributes to my effectiveness of learning.	CC1.1
1.2	The recommended duration of studying for every unit of learning contributes to my effectiveness of learning.	
2	Activities that constrain my evolution in every learning unit contributes to my effectiveness of learning.	CC4.1
3	The easy access of the LOs of every learning unit contributes to my effectiveness of learning.	IL2.1
4	The opportunity to access at any moment all information available, e.g. course index, contributes to my effectiveness of learning.	IL2.2
5	The opportunity to test my knowledge level during each learning unit contributes to my effectiveness of learning.	IL2.3
6	The opportunity to perceive my evolution along the course contributes to my effectiveness of learning.	IL2.4
7	The opportunity to select my favourite route of learning fitting best to my learning style contributes to my effectiveness of learning.	IL2.5
8	The system's support to select the learning units I need to fulfil the learning objectives contributes to my effectiveness of learning.	IL2.6
9	The system's offer of possible complementary content to be studied contributes to my effectiveness of learning.	IL2.7
10	The opportunity to access the evaluation test answers contributes to my effectiveness of learning.	IL2.9
11	The system's proposal of other learners that are in the same item of the course whom I may contact to solve a problem contributes to my effectiveness of learning.	IL3.1
12	The system's offer of a list of learners that are on-line and have already solved the problem I'm just working on contributes to my effectiveness of learning.	IL3.2
13	The system's offer of alternative learning activities in case of problems on my side contributes to my effectiveness of learning.	IL3.4
14	The opportunity to get up-to-date information like recent unread news or new forum messages contributes to my effectiveness of learning.	IL3.6
15	The system's proposal of self-assessment exercises to check if the content is appropriately learned contributes to my effectiveness of learning.	IL3.9
16	The access to a virtual library to be explored by a state-of-the-art search engine contributes to my effectiveness of learning.	IL4.4
17	The access to a list of FAQs plus their answers contributes to my effectiveness of learning.	IL4.5
18	The virtual tutor giving me some clues to help solving problems contributes to my effectiveness of learning.	IL4.6
19	The opportunity to use the Private Favourite Links study facility contributes to my effectiveness of learning.	IL4.8
20.1	The access to personal e-mail contributes to my effectiveness of learning.	IL4.9
20.2	The access to discussion forums/ brainstorming areas contributes to my effectiveness of learning.	
20.3	The access to chat rooms contributes to my effectiveness of learning.	

20.4	The access to videoconference services contributes to my effectiveness of learning.	
20.5	The access to a virtual cafeteria contributes to my effectiveness of learning.	
21.1	Being supplied with shared applications contributes to my effectiveness of learning.	IL4.10
21.2	Being supplied with the facility of instant messenger applications contributes to my effectiveness of learning.	
21.3	Being supplied with the facility of shared navigation contributes to my effectiveness of learning.	
21.4	Being supplied with the facility of shared workspaces contributes to my effectiveness of learning.	
21.5	Being supplied with the facility of shared favourite links contributes to my effectiveness of learning.	
21.6	Being supplied with the facility of shared whiteboard (notes) contributes to my effectiveness of learning.	
21.7	The access to virtual seminars contributes to my effectiveness of learning.	
21.8	The access to an agenda of events contributes to my effectiveness of learning.	
21.9	The access to a bulletin board contributes to my effectiveness of learning.	
22	Knowing that the tutor is provided with indications of my participation contributes to my effectiveness of learning.	T1.2
23.1	Knowing that the tutor is provided with data of my number of accesses contributes to my effectiveness of learning.	T1.4
23.2	Knowing that the tutor is provided with data of my time spent in a course contributes to my effectiveness of learning.	
23.3	Knowing that the tutor is provided with data of my evaluation test results in order to assess my performance contributes to my effectiveness of learning.	
24	Knowing that the tutor gets access to my answers to every question in an evaluation test contributes to my effectiveness of learning.	T1.5
25	The opportunity for the tutor to schedule and re-schedule my activities in the course contributes to my effectiveness of learning.	T2.3
26.1	Knowing that the tutor is provided by the system with information on the most relevant didactic units for me contributes to my effectiveness of learning.	T2.5
26.2	Knowing that the tutor is provided by the system with information on the most difficult concepts for me contributes to my effectiveness of learning.	
26.3	Knowing that the tutor is provided by the system with information on the most frequently asked concepts by me contributes to my effectiveness of learning.	
27	The opportunity for the tutor to adapt the standard structure of the course to my personal conditions contributes to my effectiveness of learning.	T2.8

From other point of view the following main indicators can be assessed by questionnaire from the learners:

INDICATORS
Complementary contents suggested in right moment
Immediate possibility of self assessment whenever needed
On time warning, in risk situations
Immediate suggestion given by the virtual tutor, when asked for

No.	Questionnaire for tutors	Key of UR
1	The indications about the participation of the learners provided by the system contribute to ease my tutoring of a course.	T1.2
2	The opportunity to check for a group of learners the objectives that were fulfilled in each learning unit contributes to ease my tutoring of a course.	T1.3
3.1	The data on the number of accesses of every learner provided by the system contributes to ease my tutoring of a course.	T1.4
3.2	The data on time spent by every learner provided by the system contributes to ease my tutoring of a course.	
3.3	The data on evaluation test results (in order to allow me to assess the learner's performance) of every learner provided by the system contributes to ease my tutoring of a course.	
4	Having access to every learner's answers to every question in an evaluation test contributes to ease my tutoring of a course.	T1.5
5	The system (the respective features of the system to be specified) contributes to easily schedule and re-schedule the activities of the course.	T2.3
6.1	Being provided by the system with information about the most relevant didactic units for learners contributes to ease my tutoring of a course.	T2.5
6.2	Being provided by the system with information about the most difficult concepts for the learners contributes to ease my tutoring of a course.	
6.3	Being provided by the system with information about the most frequently asked questions of the learners contributes to ease my tutoring of a course.	
7	The system (the respective features of the system to be specified) contributes to easily decide what type of collaboration activities can be performed.	T2.9