Science of Computer Programming 88 (2014) 54–81

Contents lists available at ScienceDirect

Science of Computer Programming

www.elsevier.com/locate/scico

Evaluating a graphical notation for modeling collaborative learning activities: A family of experiments

Ana I. Molina *, Miguel A. Redondo, Manuel Ortega, Carmen Lacave

Departamento de Tecnologías y Sistemas de Información, Escuela Superior de Informática de Ciudad Real, Universidad de Castilla-La Mancha, 13071-Paseo de la Universidad, 4, Ciudad Real, Spain

HIGHLIGHTS

- The use of graphical notations can facilitate the specification and design of learning systems.
- Our interest is centered on the design of CSCL systems and the modeling of group learning activities.
- We propose the use of the CIAN notation for modeling this type of learning activities.
- We describe three empirical studies to measure the adequacy of that notation to model collaboration.
- The results denote positive perceptions about the use of the CIAN notation for this purpose.

ARTICLE INFO

Article history:
Received 1 May 2013
Received in revised form 14 February 2014
Accepted 17 February 2014
Available online 26 February 2014

Keywords:
Educational modeling
CSCL
Notation assessments
Empirical study
Eye tracking

ABSTRACT

It is increasingly common to use languages and notations, mainly of a graphical nature, to assist in the design and specification of learning systems. There are several proposals, although few of them support the modeling of collaborative tasks. In this paper, we identify the main features to be considered for modeling this kind of activities and we propose the use of the CIAN notation for this purpose. In this work, we also try to empirically analyze the *quality* (in particular the *understandability*) of that notation. To this end, three empirical studies have been conducted. In these experiments we used several sources of information: subjective perception of the designers, their profiles and their performance on a set of understandability exercises, as well as the physical evidence provided by an *eye tracker* device. The results obtained denote positive perceptions about the use of the CIAN notation for modeling collaborative learning activities.

1. Introduction

The use of specification languages and standards for the design and modeling of *eLearning* systems is an open challenge that aims to promote the application of practices from *Software Engineering* (SE) and *Requirements Engineering* (RE) in the development of educational software. Thus, the development process of such applications could become more of an engineering and less of a traditional one. In this sense, the use of specifications or higher level abstraction notations, as well as the use of educational modeling languages (EMLs), constitutes an open research line [1]. Their use provides a number of advantages, among which we can mention that apart from promoting the reuse of design solutions, they provide better documentation and understanding of the designs. Moreover, these graphical notations are a good communication tool among members of development teams of this type of systems [2] (teams characterized, in many cases, by their multidisciplinarity).

* Corresponding author.
E-mail addresses: AnaIsabel.Molina@uclm.es (A.I. Molina), Miguel.Redondo@uclm.es (M.A. Redondo), Manuel.Ortega@uclm.es (M. Ortega), Carmen.Lacave@uclm.es (C. Lacave).

http://dx.doi.org/10.1016/j.scico.2014.02.019

There are many proposals of *educational modeling languages* in the literature. Among them, we highlight the proposals of Martínez-Ortiz et al. [3], PoEML (*Perspective-oriented EML*) [4] and E2ML (*Educational Environment Modeling Language*) [5]. Each one provides interesting aspects, being semantically rich and flexible, mainly due to their graphical nature. In particular, among the EMLs, our focus is on those that support the modeling of *computer supported collaborative learning* (CSCL) [6] systems. In this paradigm, there are a number of specific features that must be specified in a suitable and precise way: coordination, information sharing, collaborative activities, distribution of responsibilities, etc. In this regard, we noted that the aforementioned languages do not include the necessary support for *modeling of group activities* [7]. Next, we state the minimum features that we consider should be incorporated into a modeling language for these kinds of systems:

- *Organizational issues modeling*. Due to the fact that we want to specify CSCL environments, they should support the specification of *actors*, *roles*, *software agents*, as well as the grouping of the aforementioned, giving rise to *groups*; that is, groups of actors with homogeneous responsibilities, or *work teams*, consisting of several roles. Relationships between members of the organization should be able to be specified. You should at least be able to describe *acting* relationships (actors playing roles), *inheritance* (specialization of roles) and *hierarchical* relationships (that affect the delegation of responsibilities between roles).
- *Group learning activities modeling* (modeling of different levels of abstraction). The notation must support the modeling of the *structure of group learning activities*, i.e. the division of complex activities into simpler ones.
- *Learning activities flow modeling*. They should support the modeling of the flow of learning activities which can be related to each other by a set of *temporal operators*. We need a complete set of operators which enables us to specify complex group learning scenarios. We are not only interested, however, in specifying *order* between activities (i.e. *qualitative temporal information*), but we also need to specify *quantitative temporal information* (periods of time, dates, etc.) [8]. In addition, the tasks flow can be enriched with the specification of *information passing* between activities. And finally, in order to specify coordination completely, the notation must support the specification of *notification*, as well as triggering *events* between activities. Thus, activation and coordination between activities can be specified.
- Support (preferably in a graphical manner) for *resource* specifications (*information* and *software*) used in the *context* of the learning activities. And, in the case of group learning activities, the ability to specify *shared context* [9].
- Support to *express graphically and jointly* the relationship between elements specified (*activities*, *roles* and *resources*).

In previous works, we have proposed a notation called CIAN (*Collaborative Interactive Application Notation*), which allows modeling of group work and human–computer interaction issues [10]. CIAN can be used for collaborative learning modeling systems, adding higher levels of abstraction in the development of educational applications in specified standards, such as IMS-LD [7]. It has been shown that this language can be used to specify CSCL activities and systems, even in mobile contexts [11]. In addition, CIAN includes instrumental support (by means of a tool called CIAT) that facilitates the edition and validation of diagrams created with this notation [12].

Most notations and techniques for requirement specification (including CIAN) have a major deficiency which is the absence of its suitable *assessment*. Therefore, the main contribution of this paper is the description of several empirical studies to evaluate CIAN. These studies are based on the *subjective perception* of designers [13], as well as on physical evidence obtained using an *eye tracker* device [14]. In these experiments, we began characterizing the CIAN notation using the *cognitive dimensions framework* [15]. As a result of this heuristic characterization, we identified the positive and negative aspects of our proposed notation. With the aim of contrasting these aspects with the opinions of designers and software engineers, we performed several empirical studies. In the first one, the participants were final-year Computer Science students. In this experiment, we used an opinion questionnaire which we designed ourselves. This experience allowed us to obtain the initial empirical feedback about our proposal. The second experiment involved a more representative sample of participants (professional in software engineering) and used a more contrasted and well-founded survey for assessing requirements modeling methods, proposed by [13]. In this second study, we also contrasted the use of the CIAN notation with another notation which also allows modeling cooperative tasks (the CTT notation) [16,17]. However, as with the first experiment, the results obtained are based on purely subjective perceptions of designers about both notations. The use of subjective surveys has the inconvenience of the existence of biased answers. With the aim of solving this problem, we propose the use of a more objective source of information: the use of an *eye tracker* device. The use of *eye tracking* techniques provides objective evidence (complementary to other sources of information) that allows us to perform a more complete analysis and comparison of results. In the final experiment, we performed a controlled experiment in which we incorporated this last objective source of information.

This paper is organized as follows: first, a review of graphical notations for modeling CSCL environments is included. Then, the CIAN notation is briefly introduced. Next, the main approaches for evaluating modeling languages (mainly of a graphical nature) are reviewed. In Section 5, the experiments carried out to evaluate the use of CIAN are described. And finally, conclusions are drawn and the main lines of future work are exposed.

2. Related works: Graphical notations for specifying CSCL environments

In this section, we present and provide references for some of the main contributions in the field of graphical modeling of *e-learning* and CSCL environments.

During the last decade, there was an increasing interest in proposing standards, reference models and specification techniques in *e-learning* environments. They all have the aim of supporting the design and development of this kind of application and also aim to reinforce reuse and interoperability. Among these languages, we center our interest on the one called *Educational Modeling Languages* (EMLs) [18,19]. These specification techniques allow the description of teaching and learning interactions and activities in a formal way.

In order to support and facilitate instructional learning design (LD) many **visual learning design tools** have been proposed [20]. Among the more highlighted LD authoring tools we find ASK-LDT (*ASK Learning Designer Toolkit*) [21], LAMS (*Learning Activity Management System*) [22], MOT+ [23] and OPENGLM (*Open Graphical Learning Modeler*) [24].

Among the learning design languages, we are especially interested in the *Visual Instructional Design Languages* (VIDL) [1] which provide diagrammatic notations that facilitate communication and thinking for practitioners. Among these VIDLs we highlight E2ML [5], e-LD [3,25], PCeL [26,27] and PoEML (*Perspective-oriented EML*) [4,28]. All these languages propose the use of specific notation (symbols and rules) for designing learning activities. Most of them have a direct translation to the concepts supported by the IMS-LD standard [29,30]. IMS Learning Design (IMS-LD) is the most widely extended formal EML. IMS-LD defines a modeling technique and XML binding for describing roles and activity sequences within an environment of learning objects and services. Properties, conditions and notifications are provided at further levels. The primary goal is to provide a means for exchange and (semi) automatic execution of learning designs. However, this specification technique has certain faults. One is the nature of this specification technique: it is a textual and XML-based representation, difficult to use for people with a non-technical background. For this reason, the use of VIDLs is a good alternative. VIDLs provide a higher abstraction level for the specification of learning scenarios. They are semantically richer, easier to understand and more flexible, thanks to their graphical notation.

Other approaches have been proposed for learning process specification. Some authors propose the use of languages for graphical representations of processes such as the BPMN (*Business Process Modelling Notation*) standard [31] that could be useful in providing common interoperable representations of learning flows. In [32], the authors examine the possibility of using BPMN as a common representation notation for learning flows which are modeled using the BPEL (*Business Process Execution Language*) [33] and present an algorithm for transforming BPEL Workflows to IMS Learning Design Level A learning flows. In addition, some of the diagrams of UML [34] (e.g. use cases or interaction diagrams) can also be used for supporting communication in *e-learning* development by modeling particular instructional situations.

However, although all these VIDLs allow specification of the learning process, their use in describing collaborative learning scenarios presents certain faults. There are issues such as the *support for modeling of cooperative procedures, collaborative activities, coordination restrictions* and *spaces for the sharing of information*, that become requirements to be considered during the specification of collaborative learning processes. In relation to the IMS-LD standard, its major weakness from a CSCL perspective is its strict *sequencing model*. In addition, if multiple individuals have to collaborate, this has to be done using a *service* in the environment associated with the activity in which they must work together. In [35], IMS-LD is criticized from a CSCL point of view and a proposal is presented considering new components as necessary in the IMS-LD sequencing and act models. Among the aforementioned VIDL proposals, we highlight the PoEML approach which integrates workflow and groupware aspects into educational modeling, and focuses on a separation of eleven different perspectives of educational practices (e.g. social, organizational, temporal, etc.). Another interesting proposal is the CPM (*Cooperative Problem-based Learning Metamodel*) [36,37], which is a modeling language that targets the design of cooperative problem-based situations. It is proposed as a UML profile to express educational scenarios. However, since it was designed for educational engineers who have mastered the UML formalism, it can be quite difficult for non-technical users to handle.

We can also consider using proposals from the field of CSCW specification for specifying CSCL systems. The modeling of group work has been of interest in the areas of Human–Computer Interaction (HCI) as well as in Software Engineering (SE). Among the most relevant contributions in the field of HCI, we can highlight the *Group Task Analysis* (GTA) framework [38], the CUA (*Collaborative Usability Analysis*) notation [39] and a task analysis method called MABTA (*Multiple Aspect Based Task Analysis*) [40]. In the HCI field, one of the most widespread notations is the CCT (*ConcurTaskTrees*) notation [16]. This notation, initially proposed for modeling interactive applications, also supports the modeling of cooperative behaviors [17], by means of the use of a rich set of temporal operators from LOTOS (*Language of Temporal Ordering Specification*) [41], and the creation of a hierarchical task model. In the fields of CSCW and workflow systems, we can find the *Action Port Model* (APM) notation [42] which is taken as a reference by notations such as RML, TaskMODL and DiaMODL proposed by Traetteberg [43], and the *Proclets* proposal [44]. As for approaches derived from or completely framed within SE, we can find two extensions of UML notation: COMO-UML [45] and UML-G [46]. A more detailed review and comparative evaluation of different notations for collaborative modeling can be found in [47]. There are also specific contributions from the field of CSCL. Among them, we highlight the LDL language [48] and ACEM [49].

It is necessary to point out that, at the moment, we are interested in modeling collaborative learning processes. We are aware that specification of CSCL environments should integrate multiple perspectives or viewpoints, not only from computer science but also from the education and psychology fields [50,51] if we want to promote effective learning experiences. Thus, there are other important aspects to be considered in CSCL environments such us equal participation, individual responsibility or positive interdependencies [52]. In this work, however, we center on specification of learning processes, that is, the sequence of learning activities and collaborative tasks and flows. As we discussed in [7], we propose the use of CIAN (*Collaborative Interactive Applications Notation*) notation [10] for this purpose. Most of the elements in CIAN have a direct mapping to the main elements supported by the IMS-LD standard and add a higher level of abstraction in the specification of collaborative learning scenarios. In the next section, we present this notation and its use for collaborative learning processes specification.

3. CIAN: A graphical notation for modeling of collaborative learning activities

In order to fulfill the modeling necessities for the specification of CSCL systems, which we discussed in the previous section, we propose the use of a graphical notation called CIAN [8]. This notation supports the modeling of CSCW (*Computer Supported Cooperative Work*) systems. This notation was initially designed to model aspects of group work, but it can be used to specify *educational* and *sequencing models* of learning activities in CSCL environments. The graphical artifacts proposed by CIAN have an almost direct correspondence with the main elements supported by the IMS-LD (*IMS-Learning Design*) standard (*roles, learning activities, methods* and *resources*), as we described in [7], allowing them to be displayed together (a requirement that we pointed out in previous sections). In addition, this notation enhances the standard by including new elements (a comprehensive set of *temporal operators*, *collaborative task* specifications, *notifications*, *qualitative temporal information*, specification of *shared context*, *conditional flow of activities*, etc.).

CIAN proposes the use of several diagrams for specifying collaborative systems. In this work, we propose to use the so-called *process model* proposed by CIAN for specifying *collaborative learning processes*. For more details about the notation, the reader should consult [10].

One of the most highlighted features of CIAN is that this specification technique allows specifying in a distinctive way the *cooperative* tasks and the purely *collaborative* ones. In many contexts, these terms are often used as synonyms. Dillenbourg [53] clarifies the difference between these two concepts. *Cooperation* entails the division of work to be carried out so that each person is responsible for his or her own portion of the work. Members of the group pursue the same goals but act independently on their own tasks, or perform the same task but in separate parts of the shared context [9]. *Collaboration* entails the mutual commitment of the participants as well as a coordinated effort to solve a problem. Collaboration is, therefore, a superior activity in which, in addition to cooperating, the members of the team have to work together on common tasks and towards a common outcome. The result obtained moves through different states to reach a state of final results obtained by the group. In the final product, it is difficult to determine the contribution of each member of the group. The collaboration assumes that the various members work within an area of common representation (the *shared context*) [9]. Taking this distinction into account, we can consider that cooperation and collaboration imply different ways of understanding the division of tasks (which affects task modeling), the participation of the different roles in the development of these tasks (which affects the task and role modeling) and the product obtained as a result of this joint activity (which affects the data model). Furthermore, cooperation involves the inclusion of special coordination tasks at the end of the cooperative activity to enable the group to collect their individual contributions in the final product (group solution), as well as decision-making or agreements in the production process. In this latter case, we are talking about the existence of protocols for interaction and coordination among the group members.

To illustrate the potential of the CIAN notation for modeling collaborative learning processes, we show (Fig. 1) the *process model* that specifies the main workflow of a CSCL system, called DomoSim-TPC [54]. This system supports the collaborative learning of house automation in a problem solving approach, in which students discuss and design models by means of different representations and simulate their models in shared workspaces. The learning approach is structured into two differentiated phases: (i) *Planning*, and (ii) *Design and Simulation*. In the first phase, students reflect on the elements that should be part of the solution to the problem and plan the general actions that they should carry out to build it. In the second phase, they execute the plan, refining the design and defining the properties of the elements of the solution. Once the design has been attained, they verify its behavior by means of simulation. The *Design and Simulation* process is organized into five tasks that the students carry out in synchronous collaboration [55]. These are: (i) *collaborative design*, to construct a solution with a modeling tool; (ii) *work distribution*, to organize and distribute the modeling work among the users; (iii) *parameterization*, to define the values for the general variables of the solution through agreement among the students; (iv) *cases and hypotheses*, to define cases in which the simulation hypothesis should be tested; (v) *simulation*, to experiment with the model by means of simulation in order to check its behavior in the defined cases. These tasks are carried out following a specific protocol in shared workspaces that integrate direct manipulation mechanisms, communication and coordination support, and awareness techniques.

In Fig. 1 we can see the *process model* that specifies the main workflow of DomoSim-TPC. The process model shows the flow of execution of *activities* (1.A) which fall into four categories (1.B): *individual*, *cooperative*, *collaborative* and *abstract* (1.C). *Configure experiences* is an individual task in which only one role can appear. However, *Planning* and *Design and Simulation* are complex group work tasks. This notation also allows us to specify different abstraction levels. For this, *abstract* tasks can be used (such as *Design and Simulation* task) that are modeled in another diagram in more detail (1.G).

For all the tasks, the roles involved in their execution (1.D), the objects manipulated (1.E) and their access modifiers are indicated (C for indicating creation of objects, R for reading and W for writing). Also *domain independent support tools* (1.F) can be specified using the corresponding icon.

In this model, the workflow can be enhanced by a wide set of temporal operators, such as the *information-passing sequence*, shown in this example (1.H). In the diagram, we have included temporal and data dependencies between the tasks *Configure experiences* and *Planning*, indicating that the *Activities* data is transferred, and that the relationship between

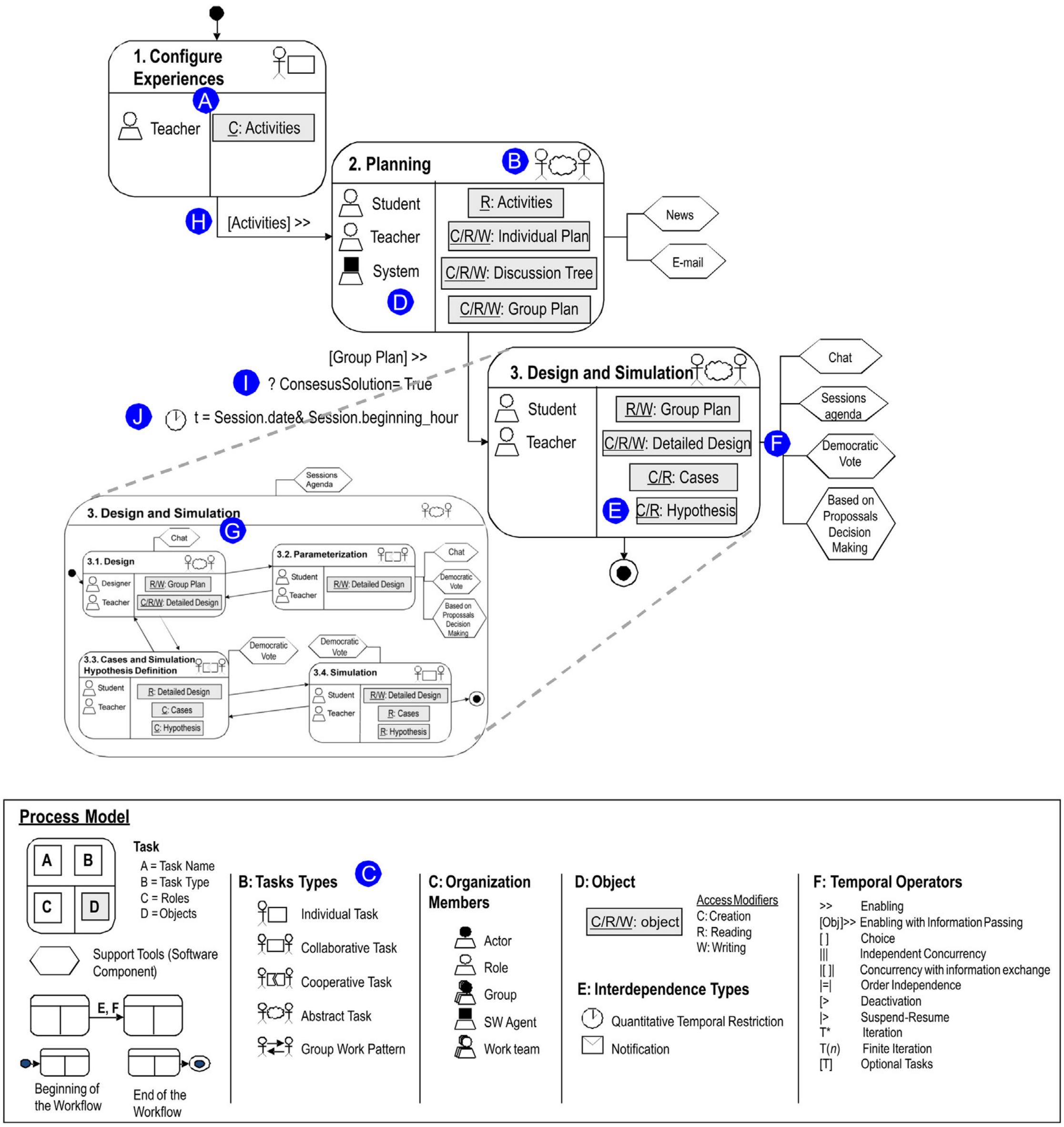


Fig. 1. Example of *process model* specified using the CIAN notation.

the tasks is sequential (≫). Between the tasks *Planning* and *Design and Simulation* we have specified *constraints* (1.I) and *qualitative temporal conditions* (1.J).

The edition of diagrams in this notation and the validation of the correctness of CIAN models are supported by a CASE tool called CIAT (*Collaborative Interactive Application Tool*) [12]. This tool has been developed using the technologies provided in the context of the *Eclipse Modeling Project*[1]: EMF[2] (*Eclipse Modeling Framework*) and GMF[3] (*Graphical Modeling Framework*) which allow for the creation of graphical modeling tools based on *Ecore* meta-models. Fig. 2 shows a screenshot of CIAT. In the figure we can see the process model of Domosim-TPC (Fig. 1) edited using the CIAT environment.

[1] http://www.eclipse.org/modeling/, last visited on 1 November 2013.
[2] http://www.eclipse.org/modeling/emf/, last visited on 1 November 2013.
[3] http://www.eclipse.org/modeling/gmf/, last visited on 1 November 2013.

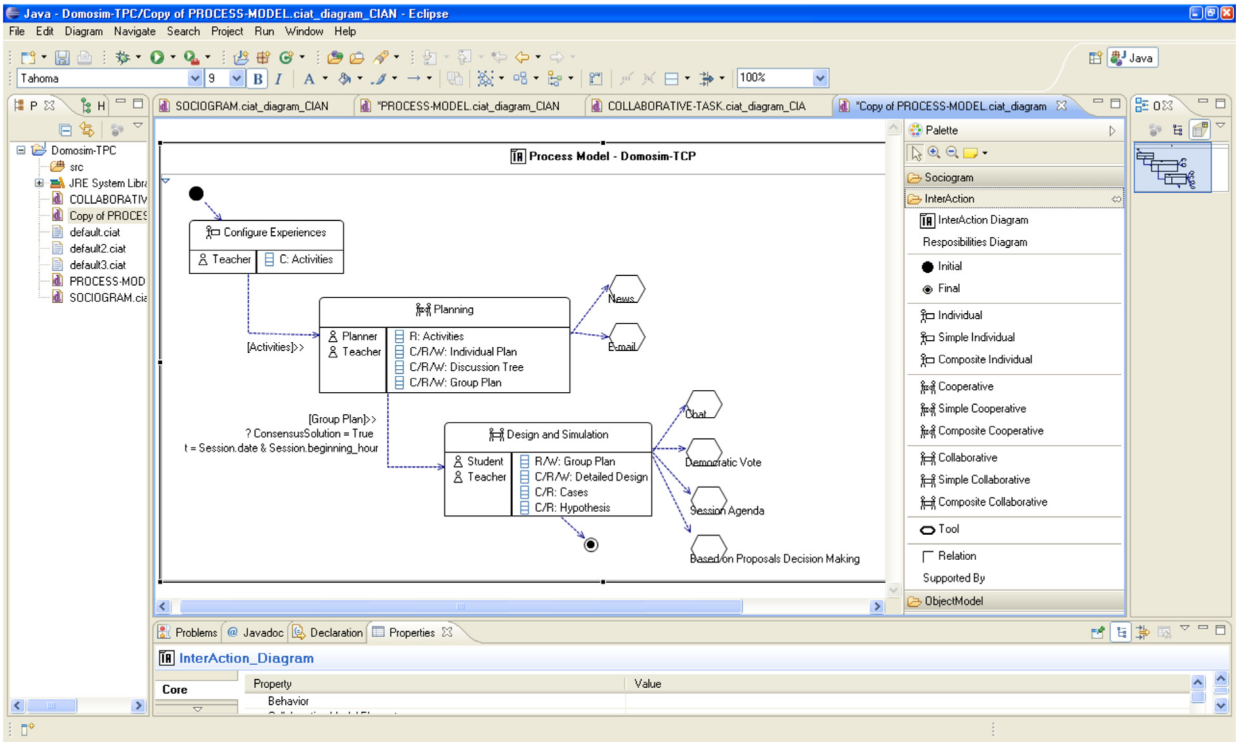


Fig. 2. Example of *process model* specified using the CIAT case tool.

4. Methods and techniques for assessing graphical notations quality

One important aspect to be considered when a new specification language is proposed, whatever its scope, is to assess its *quality* [56,57]. This issue is addressed in the area of *Empirical Software Engineering* from which two main approaches are proposed. One of them recommends the use of *objective metrics* on certain aspects of the notation and models [58]. The other suggests the use of *subjective measures* such as *perceived ease of use*, *perceived usefulness* or the *usage intentions* of the potential users of these specification languages [59].

We consider as *subjective metrics* [58] those that are obtained from user or expert opinions (expressed through a formal or informal scoring system). Because they are *perception-based*, these metrics can be considered as *qualitative measures* and cannot be automatically calculated. *Objective metrics* are indirect measurement that does not require human judgment for calculating them. For example, they can be based on internal model properties [58] or, as we propose in this paper, they can be obtained during an *eye tracking* session. The measurement of objective metrics can be automated, and they can be considered as indirect indicators of certain *quality attributes* of conceptual models (as *understandability* or *usability*).

In the second of these approaches, we find some relevant contributions among which are those that address the definition of *quality attributes* to be measured (*usability*, *simplicity*, *orthogonality*, etc.) [60]. Most of the contributions of this first group are based on the *ISO 9126* quality model [61]. This standard proposes two main characteristics: the *usability* and the *maintainability* of languages. *Usability*, in this context, is defined as a set of attributes that bear on the effort needed for use and on the individual assessment of such use by a stated or implied set of users. Usability is measured from three sub-characteristics: *learnability*, *understandability* and *operability*. The other quality characteristic is *maintainability*, defined as a set of attributes that rely on the effort needed to make specified modifications. Maintainability includes four sub-characteristics: *stability*, *analyzability*, *modifiability* and *testability*. In the literature we also find proposals of frameworks [62] and metrics (mainly based on structural properties of diagrams) to assess conceptual data models in an objective manner. Most of these proposals have been applied for assessing UML diagrams [63].

Other authors propose *guidelines* or recommendations to be considered when creating modeling languages [64]. In [65], a set of *principles* for designing cognitively effective visual notations are listed. Meanwhile, in [66], the authors describe the worst practices used in the definition of modeling languages. All these principles and guidelines can be used not only to build visual languages, but also as assessment techniques (*heuristics*) for improving and comparing notations.

We also highlight some contributions that propose evaluation methods based on the *understanding* of the models [67,59] and *user perception* [68,13]. These authors propose assessing visual languages more subjectively. All these proposed methods verify the suitability of the language (and predict its practical acceptance), based on the *effort* that is required to apply a modeling method, as well as the *perceived quality* of the artifacts produced. In the three experiments described in this paper we apply this type of evaluation.

Finally, it is necessary to point out a new line of work that proposes the use of new sources of information to evaluate the *understandability* and/or *comprehension* of the models by the designers. In this line of work, there are studies that propose using *eye tracker* devices [69–71]. The use of *eye tracking* techniques allows us to obtain a physical evidence which can complement the information provided by previous methods. Using this method, we can analyze the visualization and scanning patterns of visual representations and diagrams (such as the CIAN models).

In continuation, we describe three empirical studies performed to evaluate the use of the CIAN notation for modeling collaborative learning activities.

5. Evaluating CIAN for collaborative learning activities modeling

In order to empirically evaluate the CIAN notation, a family of experiments was conducted. In this section the details of these empirical studies are described and discussed. Their objective was to analyze the use of CIAN for collaborative tasks modeling and the degree of acceptance of this notation by designers as well as the evidence provided by an *eye tracker* device during a set of comprehension exercises.

5.1. Experiment 1: Assessing CIAN as a modeling language of collaborative learning tasks

In this section, we evaluate the language both analytically, using the *Cognitive Dimensions* framework [15], and by means of an empirical evaluation with real users.

5.1.1. Applying the Cognitive Dimensions framework to the CIAN language

The *Cognitive Dimensions* (CD) framework [15,72,73] is a popular and widely used framework to evaluate visual notations and languages [74–76]. Using CD, we can characterize non-interactive artifacts, such us tables, graphs, programming language, etc. So, we have used it to characterize the main aspects of the visual language we have proposed (CIAN). The application of this framework delivers an evaluation along different dimensions. The result of the analysis indicates how the artifacts under analysis address these dimensions. In this paper, we address the assessing of our notation on activities of *understanding*. Thus, we have selected a subset of six dimensions that we consider can have an effect on these kinds of activities:

- *Hidden dependencies*. Important links between entities are not visible.
- *Closeness of mapping*. Closeness of representation to domain.
- *Role expressiveness*. The purpose of an entity is readily inferred.
- *Visibility*. Ability to view components easily. This aspect is related to what kinds of things are more difficult to see, identify or find.
- *Secondary notation*. Ability to add information to the notation outside its formal syntax. This aspect refers to the use of visual variables not formally specified in the notation to reinforce or clarify meanings, e.g., use of color to highlight information.
- *Hard mental operations*. High demand on cognitive resources.

Table 1 depicts the conclusions obtained from applying the framework to the language.

Table 1
Results of the application of the Cognitive Dimensions framework to CIAN.

Cognitive dimension	CIAN
Hidden dependencies	All of the relationships between main concepts and elements (*task-roles-resources*) in CIAN are visible. The relationships between tasks (*task–task*) in the process modeling are also visible.
Closeness of mapping	It has been included when possible. For example, in the case of the icons proposed to represent roles, groups, notifications or temporal restrictions, we have selected a visual representation whose appearance suggests their meaning.
Role expressiveness	All the nodes in the language are different, easy to interpret and distinguish. We consider that the role of each visual element in a CIAN model is obvious and that the particular elements are easy to interpret.
Visibility	We consider that all the graphical elements in CIAN are easy to see, identify and find.
Secondary notation	There is no notation apart from the visual language itself which adds information to the diagram. For example, color is not used as an additional cue to express extra information in the models. However, in the case of task, the layout of nested elements (roles, object) indicates a relationship between the task and the resources and subjects involved in its execution. This can be considered a particular use of secondary notation.
Hard mental operations	Since all the main elements (roles, tasks, objects, relationships, notifications, etc.) are integrated into a single diagram, the users do not need to remember any additional concepts in order to work with the diagrams.

Once we had characterized our notation, we performed several experiments in which we assessed several aspects of the CIAN notation for modeling collaborative learning scenarios and tasks. We began with a first experiment whose main objective was to obtain an initial feedback (their subjective perception) from potential users about the *complexity* of the CIAN notation and its ability to visually show the relationship between the main elements to be modeled in a collaborative process (*tasks-roles-resources*). This aspect is related to the *hidden dependencies* dimension of the CD framework. The survey used in this first experience was created by us and the sample was formed by students taking a Computer Science Degree.

5.1.2. Use evaluation

The objective of this first experiment was to assess certain features of the CIAN notation related to its *ability for modeling collaborative tasks* and the *complexity* of its models. In this first study, 59 undergraduate students from the College of Computer Science (ESI) of the University of Castilla-La Mancha (UCLM), Spain, were involved. Students participated in a seminar on the CIAN notation, in which documentation and modeling examples were provided. After the seminar they performed a modeling exercise with this notation: "*modeling of a collaborative system to support the procedures and defense of the Degree Ending Project*".

In this first experiment certain aspects were assessed, using a questionnaire, some of them related to general opinion about use of graphical requirements modeling methods and some more specific aspects of CIAN diagrams for CSCL modeling.

Of the more general aspects, we asked about students' *preference* regarding the *graphical notations* over textual specifications. 85.75% expressed their preference for the notation of graphical nature compared to 11.86% who preferred textual notations, while the remaining 3.39% did not indicate any preference. These results are consistent with the majority of studies that assess domain specific languages [77] and the design guidelines and recommendations related to this issue [65,64]. In accordance with Green et al. [78], we agree that graphical representations are, in most cases, inherently superior to textual representations.

Next, we included some questions related to specific aspects of *process model* provided by CIAN for modeling collaborative processes. In relation to the *structural complexity* of the notation, 73.2% of the participants felt that the *number of icons* proposed by the notation was suitable. The number of graphical elements provided by a certain notation is also an important issue to be analyzed. This is related to the so-called *graphic economy* [65] which recommends including a manageable number of different graphical symbols in a notation. In the case of CIAN, participants considered that the size of the visual language provided by CIAN was suitable.

The questionnaire filled out by the participants included more specific questions related to the *process model* supported by CIAN (Fig. 1). Regarding this diagram, 72.88% of the students felt it was very useful to show the relationships graphically and jointly between *tasks–roles–resources*. In fact, this is one of the most valued aspects of this notation. This result is consistent with the fact that the CIAN notation presents low support to the *hidden dependencies* cognitive dimension (as we pointed out in Table 1).

This first experiment allowed us to obtain an initial empirical evaluation of several aspects of the notation. As we have previously commented, in this first study the participants were undergraduate students (final-year Computer Science students). This fact can affect the *external validity* of this first study (i.e. the generalization of the findings). In this regard, some authors suggest that students could be considered as the next generation of professionals [79]. They argue that, under certain conditions, there is not great difference between this type of student and professionals, considering that their ability to understand requirements models is comparable to that of typical novice analysts [80,81,13]. However, in order to solve this possible limitation, we performed a second experiment in which software engineering professionals were involved.

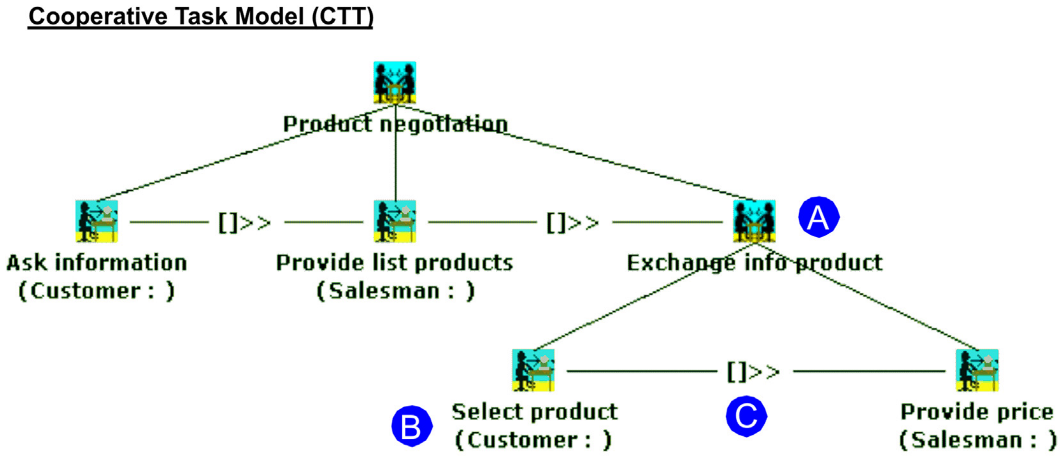
Another limitation of this first experiment is the fact that we used a questionnaire which we created ourselves. We consider it more suitable to use an evaluation method that is better contrasted, based on theoretical foundations and empirical studies. Therefore, we suggest the use of the evaluation method proposed by Abrahao et al. [13]. This method allows evaluating the *quality* of requirements modeling methods based on *user perceptions*. This method consists of a theoretical model which explains the relevant dimensions of quality for requirements modeling methods, together with a practical instrument to measure these dimensions. This proposal allows predicting the likelihood of a particular method being accepted in practice, based on the quality of the produced requirements artifacts and the user perceptions of the quality of the method.

Finally, and as pointed out in Cheng and Atlee [82], if practitioners are to consider adopting a given requirement modeling method, they must know how effective it is compared to other similar methods. In the next experiment, we propose to compare the CIAN notation with another notation which also allows modeling collaborative tasks (the CTT notation) [17].

5.2. Experiment 2: CIAN vs. CTT for modeling collaborative learning tasks

As we have commented above, a second experiment was designed to study the *understandability* of the CIAN notation and to compare it with other notations for specifying group work activities. The CTT (*ConcurTaskTrees*) notation was chosen, one which is considered a *de facto* standard in the Human–Computer Interaction area [16]. The choice of the CTT notation is based on its ability to specify cooperative behaviors by means of the creation of a hierarchical task model [17].

CTT includes several types of tasks and a set of temporal operators that allows tasks that are on the same level of the hierarchy to be related so that sequence, choice, concurrence, etc. between them can be identified. CTT has four main types of tasks, the most important being the *interaction* tasks (which represent the tasks that are initiated by the user) and the *application* tasks (representing the response actions to these interactions). The other two types are the *abstract* tasks (which



Tasks Types

Icon	Type	Definition
	User Task	Tasks performed by the user. Usually, they are cognitive activities like thinking a strategy tosolve a problem.
	Abstract Task	Tasks which requires complex activities or that could be decomposed in simpler ones.
	Application Task	Tasks performed by the application. For example: presenting results to the user.
	Interaction Task	Tasks performed by the user interacting with the system. For example: selecting an option.
	Cooperative Taak	Tasks that imply actions by two or more users in order to be performed.

Temporal Operators

>>	Enabling
[Obj]>>	Enabling with Information Passing
[ ]	Choice
|||	Independent Concurrency
|[ ]|	Concurrency with information exchange
|=|	Order Independence
[>	Deactivation
|>	Suspend-Resume
T*	Iteration
T(*n*)	Finite Iteration
[T]	Optional Tasks

Fig. 3. Example of a cooperative task model (extracted from [17]).

allow management of the levels of abstraction) and *user* tasks (which are of a cognitive or motor nature, and which precede the interaction tasks). In the case of specification of cooperative behaviors, the notation includes the use of a fifth type of task, the *cooperative task* (Fig. 3.A), defined as a task that requires activities from two or more users to be accomplished for its performance. For each task in the cooperative model which cannot be further decomposed into other cooperative tasks, designers have to specify the following information: task name, roles of users involved, type of communication protocol (for example, synchronous, asynchronous and broadcast), roles cardinality, the media used to interact and communicate, the logical objects that the task needs to manipulate and some additional informal comments which can be added to further describe the task or some of its main features. One of the main faults of this graphical notation is that the majority of this data is specified using the supporting tool (CTTE) [16,17], but is not graphically shown in the diagrams. In cooperative models, we can only specify the type of the task (cooperative or not), the role of the users involved in its execution (only one role), the hierarchical decomposition of tasks and the temporal operators that relate tasks to each other. In this sense, and considering the CD dimensions (Section 5.1.1) we can assume that CTT notation for modeling of cooperative behaviors presents *hidden dependencies*. Only *task–role* (Fig. 3.B) and *task–task* (Fig. 3.C) relationships can be visually shown in the models.

In Fig. 3 we can see an example of a CTT cooperative model whose purpose is to indicate the relationships among tasks performed by different users. Negotiating a price is a cooperative task because it requires actions from both a *customer* and a *salesman*. Cooperative tasks are represented by a specific icon with two persons interacting with each other. In the cooperative model, cooperative tasks are decomposed until we reach tasks performed by a single user that are represented with the icons used in the single user parts.

In Table 2 we compare the *process model* provided by the CIAN notation and the *cooperative model* included in CTT proposal for the modeling of collaborative learning scenarios.

In this second experiment, 14 professionals participated (12 men and 2 women) with an average age of 32. Participants were enrolled in a course on Human–Computer Interaction taught by the Department of Information Technologies and Systems at the University of Castilla-La Mancha (UCLM), Spain, and their participation was voluntary. All of them were familiar with CTT notation for specifying mono-user interactive systems, although they were not aware of its use for cooperative modeling.

The assessment done was based on the *subjective perception of designers* and in its development the evaluation method proposed by Abrahao et al. [13] was followed. We performed this second experiment basing its performance on the ex-

Table 2
Comparison between CTT and CIAN for the modeling of collaborative systems.

	CIAN notation	CTT notation
Syntaxis type	Graphical notation	Graphical notation
Types of organizational members supported in the models	Roles, actors, software agent, group, work team	Roles
Group work Task modeling *concepts* (only graphically supported by models)	Roles, abstract tasks, cooperative tasks, collaborative tasks, individual tasks, notifications, temporal restrictions, constraints, objects	Roles, abstract tasks, cooperative tasks, individual tasks
Group work Task modeling *relationships* (only graphically supported by models)	Workflow (*task–task* relationships), hierarchical relationships between tasks, *task–role–resources* relationships	Workflow (*task–task* relationship), hierarchical relationships between tasks, *task–role* relationships
Cooperation/Collaboration distinction	Supported	Not supported
Qualitative temporal operators (order)	Rich set of *temporal operators* (from LOTOS)	Rich set of *temporal operators* (from LOTOS)
Quantitative temporal operators	Graphically supported (dates, periods of time)	Graphically not supported (definition of *preconditions* using the CTTE environment)
Notifications and events	Graphically supported	Graphically not supported (definition of *preconditions* using the CTTE environment)
Resources specification	Graphically supported (shared context, information-passing, domain independent support tools, access modifiers of data objects specification)	Graphically not supported (definition of logical *objects* used in the context of a task using the CTTE environment)
Supporting tool for edition and checking the semantic validity of these models of models	Yes (CIAT)	Yes (CTTE)
Express graphically relationship between main concepts	*Role–task–resources* (objects and support tools), *task–task*	*Role–task, task–task*
Modeling of several roles executing the same group work task	Graphically supported	Not supported

perimental method proposed by Ciolkowski et al. [83] and guided by the experimental process of Wohlin et al. [84]. The general goal of this experiment was to test the usefulness of our models empirically. We use the *Goal–Question–Metric* (GQM) template [85] for goal definition. We formulate the experimentation goal as follows:

> *To* ***analyze*** collaborative learning models represented with CIAN
> ***with the purpose of*** testing this notation
> ***with regard to*** the ease of use, usefulness and intention to use of the models,
> ***from the point of view of*** software engineering professionals
> ***within the context of*** the specification of collaborative learning applications.

Thus, our **research questions** can be: $RQ_{\#1}$:*"Is CIAN perceived as easy to use and useful for CSCL requirements modeling?"* and $RQ_{\#2}$:*"Is there an intention to use CIAN in the future for designing CSCL systems?"*. We base this empirical study on the *user perception* with regard to the quality of the CIAN notation. We consider the dimensions of *quality* for requirements modeling methods proposed in [13]. For this purpose, we use the *measurement instrument* provided by these authors for assessing requirement modeling methods (Table 3). We have done certain adjustments and modifications in the drafting of the items (adapting the questions to CIAN and CTT notations and models).

As we have mentioned, for performing this second study we followed the experimental process of Wohlin et al. [84], graphically expressed in Fig. 4. The 14 participants were given both theoretical and practical seminars on CTT and CIAN notations (*preparation phase*), with a duration of approximately 30 minutes each. Following that, they accomplished two understanding exercises of two diagrams which specified two collaborative learning workflows, one using the CIAN notation and the other the CTT notation (*development phase*). The development took place in a single room. It was controlled in such a way that no interaction between subjects occurred. The order of performing these exercises was randomized for each participant, in order to limit *learning effects*. For each of these exercises, participants annotated the starting and finishing time. To avoid a possible *ceiling effect*, there was no time limit for the performance of the tasks. Next, they analyzed and tried to understand the diagram. Their *level of understanding* was measured by answering six statements (*true/false*) related to the collaborative process specified. These questions were related to the relationship between tasks (workflow), the relationship between tasks and the roles involved in their execution, and the understanding of constraints between tasks (notifications).

Table 3
Items in the survey for measuring the perception-based variables (extracted to [13]).

Item	Item statement
PEOU1	The requirements modeling method is simple and easy to follow
PEOU2	Overall, the requirements models obtained by the method was easy to use
PEOU3	It was easy for me to understand what the requirements model was trying to model
PEOU4	The requirements modeling method is easy to learn
PU1	I believe this requirements modeling method would reduce the time required to understand software requirements
PU2	Overall, I found the requirements modeling method to be useful
PU3	I believe this requirements modeling method is useful for building a conceptual model of a software system
PU4	I believe that the requirements specifications obtained with this method are organized, clear, concise and non-ambiguous
PU5	I believe this requirements modeling method has enough expressiveness to represent functional requirements
PU6	Overall, I think this requirements modeling method provides an effective means of describing requirements specifications
PU7	Using this requirements modeling method would improve my performance in describing requirements specifications
ITU1	If I am working in a company in the future, I would use this requirements modeling method to specify functional requirements
ITU2	It would be easy for me to become skillful in using this requirements modeling method
ITU3	I intend to use this requirements modeling method in the future
ITU4	I would recommend the use of this requirements modeling method

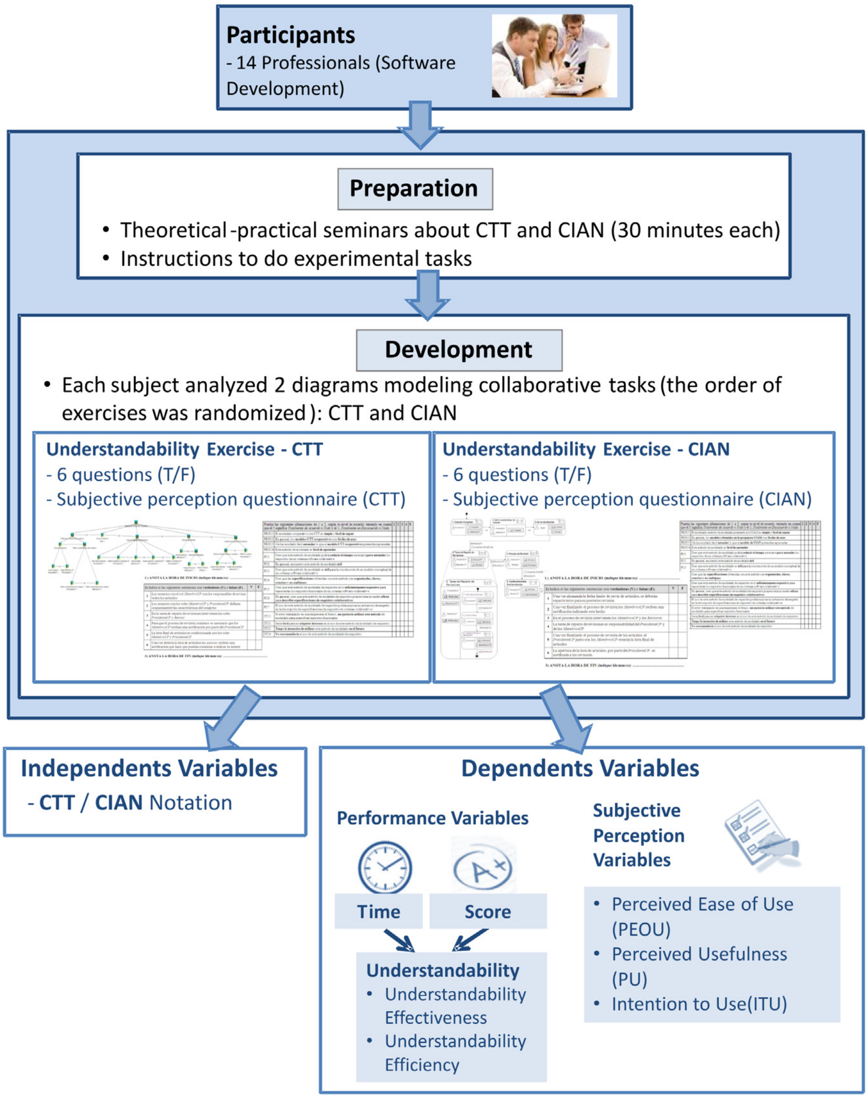


Fig. 4. Experimental design.

Table 4
Summary of the dependent and independent variables considered in the second experiment.

Type	Subtype	Name	Measure
Independent variables		Notation	CTT or CIAN diagram
Dependent variables	Performance variables	Time	Time spent to understand the model and to answer the questionnaire
		Score	Number of correct answers
	Performance-based variables	Understandability effectiveness	Average number of correct answers divided by the total number of questions
		Understandability efficiency	Average number of correct answers divided by the time spent to understand the model and to answer the questionnaire
	Perception-based variables	Perceived Ease of Use (PEOU)	Mean of the items that measure each of the items PEOU1…PEOU4 obtained in the survey proposed in [13] (Table 3)
		Perceived Usefulness (PU)	Mean of the items that measure each of the items PU1…PU4 obtained in the survey proposed in [13] (Table 3)
		Intention to Use (ITU)	Mean of the items that measure each of the items ITU1…ITU4 obtained in the survey proposed in [13] (Table 3)

Table 5
Results of comparative study (CTT vs. CIAN).

	CIAN notation[a]	CTT notation[a]
Time (seg)	**282.64** (**99.65**)	355.36 (86.61)
Score	**5.36** (**0.63**)	4.86 (1.03)
Understandability effectiveness	**0.89** (**0.11**)	0.81 (0.17)
Understandability efficiency	**0.24** (**0.15**)	0.17 (0.07)
PEOU	**3.68** (**0.51**)	3.39 (0.66)
PU	**3.78** (**0.41**)	2.71 (0.63)
ITU	**3.29** (**0.60**)	2.64 (0.71)

[a] We show the *mean* scores and the *standard deviations* (in parentheses).

These understanding tasks allowed us to obtain two measures for understandability (Table 4): *Understandability Efficiency*, and *Understandability Effectiveness* for the two notations to compare (CIAN and CTT).

The *development phase* included two tasks: an understanding task and a post-task survey. In the *post-task survey*, for each of the two exercises, participants rated on a Likert scale (from 1 to 5) their level of agreement or disagreement with 15 statements used to measure the *perceived ease of use* (PEOU), *perceived usefulness* (PU) and the *intention to use* (ITU) for the specific (CIAN or CTT) notation (Table 4).

In Table 4 we summarize the dependent and independent variables considered in this second experiment. The *time* spent solving each exercise, in addition to the *score* of the six questions, were used to measure the *understandability effectiveness* and *understandability efficiency* of the models by the designers. The *understandability effectiveness* is calculated as the average number of correct answers divided by the total number of questions. The *understandability efficiency* is the average number of correct answers divided by the time spent to understand the model and to answer the questionnaire. The *perception-based variables* (PEOU, PU and ITU) are calculated as the mean of the items that measure each of these three aspects in the surveys proposed in [13].

After the experiment took place, we collected and analyzed the experimental data. Table 5 shows the results of the experiment. For each of the variables considered, the mean values and standard deviation (in parentheses) are shown. Due to the small sample size in this second experiment, we realized a descriptive analysis of the data collected. In addition to the numeric analysis, a graphical analysis was performed with boxplots (Fig. 5).

We now discuss the values related to task *performance*. We can see how the *time* spent solving CTT notation exercises ($M = 355.36$, $SD = 86.61$)[4] is, on average, higher than the time taken to understand and solve the CIAN notation exercises ($M = 282.64$, $SD = 99.65$). The difference between the time spent in analyzing each type of diagram was statistically significant ($F = 4.25$, $p = 0.04 < 0.05$,[5] in an analysis of variance with a significance level of 0.05 and a confidence level of 95%). The *score* obtained by participants is greater for the CIAN model ($M = 5.36$, $SD = 0.63$) compared to the CTT version ($M = 4.86$, $SD = 1.03$). These results can be related to the different support for the *cognitive dimensions* discussed in Section 5.1.1 provided by CIAN and CTT. We consider that both notations fulfill the *closeness of mapping* dimension, using visual representations whose appearance suggests their meaning (the so-called *principle of semantic transparency* [65]). However, the support for *role expressiveness* is different in each notation. In the case of CTT, we consider that the visual elements are more difficult to interpret, with this fact affecting the time needed for recognition. We believe that the graphical elements of

[4] M = Mean; SD = Standard Deviation.
[5] F = F of Snedecor.

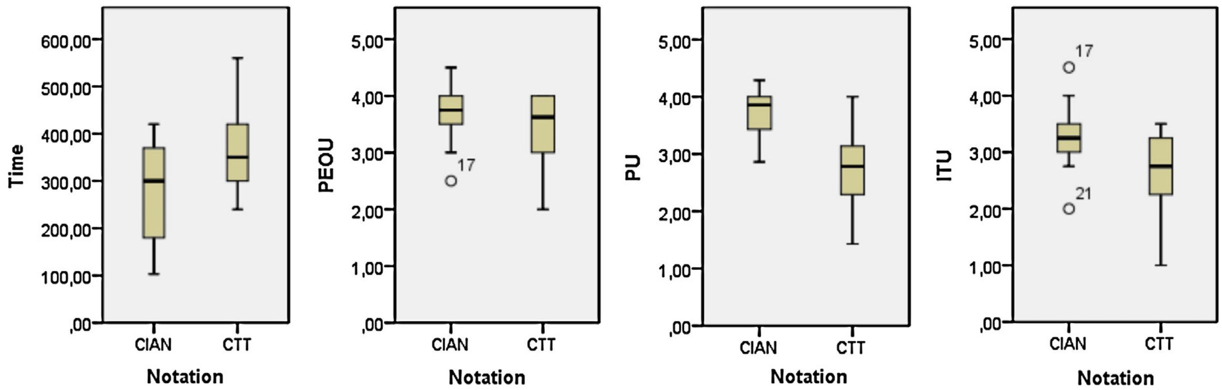


Fig. 5. Boxplots of time, PEOU, PU and ITU variables (CTT vs. CIAN).

CIAN are better with respect to their *clarity* (they are more legible), *simplicity* (without unnecessary embellishments, colors or details) and *discriminability* (for example, the icons provided to distinguish the type of group work tasks). We consider CIAN supports in a better way the so-called *principle of perceptual discriminability* [65], which states that different symbols should be clearly distinguishable from each other. In general, the greater the visual distance between symbols, the faster and more accurately they will be recognized. In the case of CTT, the different types of tasks (for example, interactive and cooperative tasks) are visually very similar, making it more complex to differentiate between them.

As for the *measures of subjective nature*, we see that the CIAN notation is better valued by the participants. The scores for the three variables considered (PEOU, PU and ITU) for the CIAN notation are higher ($M = 3.68$, $M = 3.78$ and $M = 3.29$, respectively) in all cases compared to those obtained by the models in the CTT notation ($M = 3.39$, $M = 2.71$ and $M = 2.64$, respectively). In the case of PU and ITU, the differences were statistically significant ($F = 28.1$, $p = 0.00 < 0.05$ and $F = 6.72$, $p = 0.02 < 0.05$, in an analysis of variance with a confidence level of 95%). Thus we can conclude, as expected, that the majority of the participants found the CIAN notation quite useful and easy to use for modeling collaborative learning processes. This is supported by their efficiency and effectiveness in performing the understandability tasks.

We can, therefore, conclude that the professionals who participated in this experiment appreciated CIAN more and they achieved higher performances when the notation used to specify collaborative tasks was CIAN as opposed to the CTT notation version. The majority of the respondents were very positive about the use of the CIAN notation for modeling CSCL systems in the future.

This second empirical testing was designed with the aim to alleviate some threats to the *internal validity*. Thus, the understanding tasks were randomly assigned to the subjects in a different order with the aim of controlling the possible *differences among subjects*. Also, the *universe of discourse* (the case scenario used in the CIAN specifications) was well-known by the participants. Regarding the *fatigue effects*, we have to point out that the average time spent in completing the understanding tasks was approximately 6 minutes in the case of the CTT exercises and 4.7 minutes in the case of the CIAN exercises. Hence, fatigue did not have an influence on the result obtained. The experiment was carried out with subjects who had never done a similar experiment before, thus avoiding *persistence effects*. In relation to *subject motivation*, we have to mention that subjects were highly committed to this research.

However, there is an important threat to the internal validity of this kind of study: the *researcher and participant bias*. The participants sometimes say what they think the researcher wants to hear. Although we can reduce this threat by informing the subjects that their answers would be treated anonymously, this fact does not guarantee that their answers are completely sincere and objective. To deal with this fault, we propose including a more objective source of information in the next experiment: the use of an *eye tracker* device. The use of *eye tracking* techniques provides objective (physical) evidence that allows us to perform a more complete and objective analysis and comparison of results. In the final experiment we performed, a controlled experiment in which we included *eye tracking* techniques combined with a more subjective measurement instrument (like the one used in this second experience).

5.3. Experiment 3: CIAN vs. CTT for modeling collaborative learning tasks. Analysis based eye tracking *techniques*

Finally, we describe the last experiment performed for assessing the CIAN notation. The ***objective*** of this third experiment was to evaluate in greater detail three particular aspects of CIAN: the "*ease of understanding the flow or order between tasks*", the "*ease of locating and understanding the relationship between tasks and the roles that perform those tasks*" and, finally, the "*ease of locating and understanding the specification of notifications*", comparing these features in the two notations used in the second study (CIAN and CTT). To evaluate each of these aspects, several comprehension exercises in CIAN and CTT notations were designed.

In this empirical study, we considered several sources of information to assess both notations: the *subjective perception* of designers, their *profile* and some evidence of a physical nature provided by an *eye tracker* device [14]. In this last experiment, we want to measure in an objective way the *cognitive effectiveness* of the studied notations. This concept can be defined as the speed, ease and accuracy with which a representation can be processed by the human mind [86]. This provides an operational definition of visual notation "goodness" that can be empirically evaluated [65]. The *cognitive effectiveness* of a certain graphical representation correlates with the *cognitive load* that it imposes. The *eye tracking* technique is an indirect method used to measure the *cognitive load* or *mental effort* that intends to try to understand a particular image or diagram.

The concept of ***eye tracking*** refers to a set of technologies which monitor and record the way a person looks at a particular scene or image and, more specifically, on what areas they fix their attention, for how long and in the order in which he/she visually explores the graphical material provided. The *eye tracking* technique has been applied in various disciplines and areas of study: marketing, advertising, and evaluation of user interfaces (mainly web pages) [14,87]. Recently, several authors have proposed the use of this technique to provide new empirical evidence in the study of graphical notations [69–71]. We believe that there is great potential in using this objective source of information. The use of *eye tracking* techniques allows us to add evidence of a physical nature to the evaluation of aspects such as diagram *understandability* and the *visual effort* applied by users to try to comprehend graphical specifications. Using an *eye tracker* device allows answering questions such as: "*Where do people look when trying to understand and analyze a graphical model?*", "*Where are they fixing their attention?*", "*How do they explore and examine the diagram?*". We can complement the data provided by more subjective sources of information (for example, the designers' subjective perception collected by questionnaires) and contrast them with a more objective source of information (as the one provided by an *eye tracking* device).

In the design and development of this final *eye tracking* test, we followed the ***methodology*** proposed by Nielsen and Pernice [88] for conducting *eye tracking* empirical studies. This methodology includes recommendations about the recruiting of users, number of questions in pre- and post-tests, facilitator (investigator) interruptions during the session, etc. In addition, this document includes recommendations about more specific aspects like the chair that must be used during the test, the participant's seating position, the most suitable lighting conditions, aspects related to the calibration phase, etc.

Choosing the number of test participants needed in this kind of study is an important aspect to be considered. It is necessary to point out that some people cannot be eye tracked at all for various reasons. *Eye tracking* technology has been developed to work well with people who have healthy eyes and normal visual acuity. Thus, for example, people wearing bifocals are typically difficult to track. Moreover, the use of contact lenses or the length of the person's eyelashes can make *eye tracking* more difficult. In this experiment, the sample population has been restricted to "*suitable eye tracking individuals*" [88]. Individuals with contact lenses, glasses and/or poor trackability were excluded from the tests. This restriction affects the sample size, because the initial number of participants recruited may decrease [89]. When performing an *eye tracking* study we have to be prepared to lose about 5% of the *eye tracking* recordings due to technical reasons or because participants cannot be tracked. Hence, at least 5% extra recruitment is recommended in order to end up with enough good quality recordings for analysis.

With regard to the most suitable place to run this kind of study, Nielsen recommends performing *controlled experiments* and conducting the *eye tracking* sessions in a laboratory setting. This is because, for example, in a laboratory setting the lighting conditions can be arranged in a way that does not interfere with the *eye tracker*. Other environmental interruptions can also be controlled.

To perform this last study, we used the *Usability Lab* ***equipment*** of the CHICO (*Computer–Human Interaction and Collaboration*) research group of the University of Castilla-La Mancha (UCLM), Spain. Apart from the common resources found in any computer laboratory, this one includes the equipment necessary for the usability and accessibility testing of interactive systems. For this, it includes an *eye tracker* device (Tobii X60 model), common systems found in *testing* and *interview rooms* (cameras, microphones and a PA system) and an *observation room* for monitoring tests. For the design, development and analysis of the data obtained during the *eye tracking* sessions, we used the *Tobii Studio* software version 3.0.2.

In the next subsections, we explain the details of this last experiment. Before carrying out the final experiment, a *pilot test* in which some professors and researchers participated was performed. This test was used to refine some details of the experiment (materials supplied, questionnaires, and duration), as well as to realize *eye tracker* calibration testing. Conducting the pilot test allowed us to prepare the *experimental tasks* and the conditions required to successfully run the *eye tracking* session.

5.3.1. The pilot test

In the ***pilot test,*** eight professors and researchers from the College of Computer Science (ESI) of the University of Castilla-La Mancha (UCLM), Spain, voluntarily participated. As a result of this first phase, we made some decisions about the instructions that would be included in the final test, the way to supply the questionnaire and how to record the subjects' responses.

Four of the professors that participated in the pilot test acted as ***test facilitators*** during the final study. This pilot phase served as preparation for these facilitators and as a basis for the creation of a *test session script*. During the test session, the facilitator has to keep track of many things simultaneously. Hence, most researchers recommend developing a test session script which contains everything the facilitator has to do and say to the participants during the session. By using a script, we can make sure that test results are not influenced by participants receiving different instructions or proceeding through the test session in different ways. While participants are working on their tasks, the facilitator also observes what they are doing and takes notes about possible misconceptions or problems during the session.

As a result of the pilot test, some *eye tracking* testing conditions which can interfere in the session were controlled. Thus, the testing room was prepared for carrying out the final test. When testing, it is important to check whether the *eye tracker*

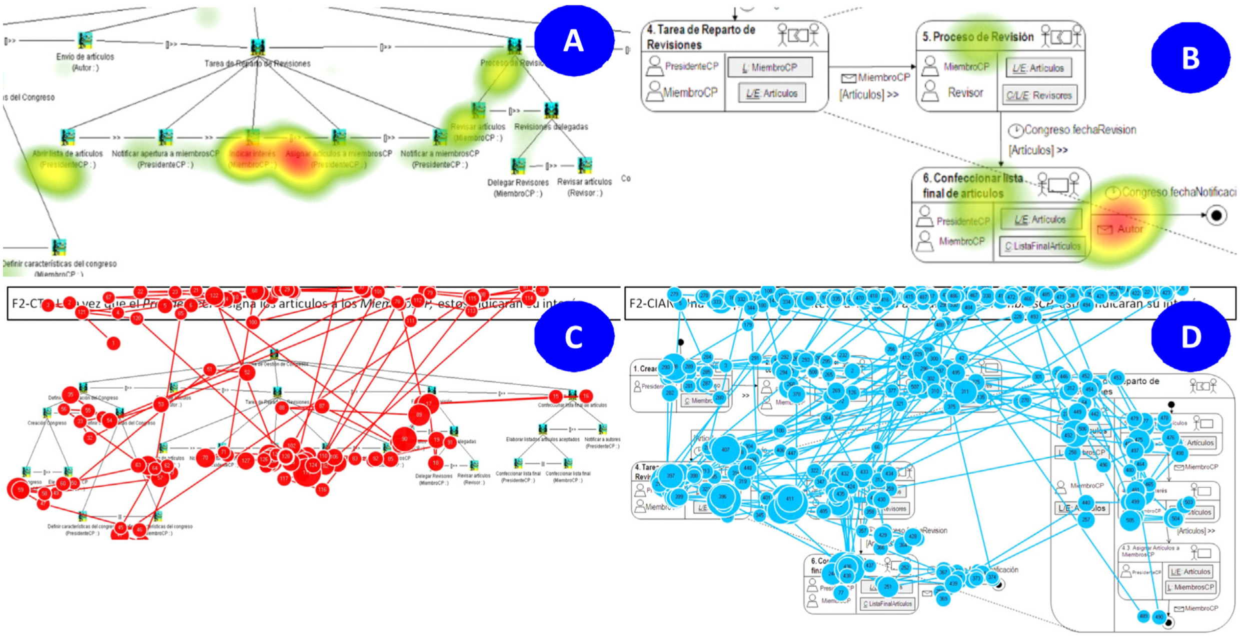


Fig. 6. Examples of *heat maps* and *scan paths* generated during the pilot test.

will work under the given lighting conditions before beginning with participants. Finally, this phase allowed us to decide which *eye tracking* measures to consider in the final test and the interpretation of each of them. To this end, we applied a *retrospective thinking aloud* method (RTA) [90,91]. More details about the measures selected and their interpretation are discussed in the next section.

5.3.2. Eye tracking measures and interpretation

When we run an *eye tracking* session, we can perform a *quantitative analysis* of the test results (when the sample size is adequate) or a *qualitative* one [88]. In this last case, we can use two representations: *heat maps* and *scan path* or *gaze plot*, which allows us to show graphically the visual behavior of a user or a group of users. In Fig. 6, we can see examples of these representations, obtained during the pilot test. Fig. 6.A shows the *heat map* generated by one of the participants in a comprehension task of the flow (order) between two tasks. This representation includes the so-called "*hot spots*" (marked in red) that show where users have paid most attention (in this case on the two tasks whose order had to be checked). Fig. 6.B shows an extract of the *heat map* generated in a task of *notification* location in a CIAN notation diagram. For its part, the *scan path* allows the identification of *diagram analysis patterns* and searches which are more or less efficient. For example, in Fig. 6.C we can see the *scan pattern* on a CTT tree in which the subject had to locate two tasks to check their order (the *heat map* in Fig. 6.A corresponds to this same task). In the figure (Fig. 6.C), we can see how the user progressed through the various depth levels of the tree to locate the tasks that he was asked about. Finally, in Fig. 6.D, we can see the pattern generated by one of the participants who had difficulty in locating the answer to a question about the tasks order in a CIAN diagram with two abstraction levels. We can see how in this case a long and dense *scan path* is generated, which indicates the effort involved in locating the answer.

If we want to perform a quantitative analysis, we can use several metrics that we can collect during an *eye tracking* session [87]. These metrics can be used to measure the participants' *visual effort* in tasks of analysis and diagram comprehension. In Table 6, we show a selection of what we consider most useful for this purpose as well as its interpretation. Most of them are related to the number and duration of the so-called *fixations* (obtained when the eye remained stationary, focused on an area of interest for a certain period of time).

It is necessary to point out that a metric can be interpreted in different ways. So for example, longer fixations can mean a user found a particular area interesting but it can also mean that they found the area difficult to interpret [92]. Hence, it is important to attempt to supplement *eye tracking* data with additional information gained from the participants about their experiences during the *eye tracking* session. To decide the final interpretation of each metric, we applied the *retrospective thinking aloud* (RTA) method during the pilot test, commonly used for usability testing [93]. In this method, once the tasks were completed by participants, they replayed the *eye tracking* session again and verbalized their experiences while doing the tasks. Applying this method helped us to determine the most suitable interpretation for the metrics gathered during the *eye tracking* session.

In addition, to analyze the data obtained in an *eye tracking* session it is necessary to define the so-called "*areas of interest*" (AOI) of the image displayed. These *areas of interest* (AOIs) are for the evaluation team, and depend on the specific task to

Table 6
Eye tracking metrics – interpretation.

Metric	Interpretation
***Behavior Patterns* (*Analysis of the scan path*)**	
Scan path complexity (*duration, length and density*)	A longer-lasting *scan path* indicates *less efficient scanning/searching*. Smaller *spatial density* indicates a more direct search.
Scan path direction	This can determine a participant's *search strategy* (e.g. *top-down* vs. *bottom-up* scan paths). "*Sweep*" denotes a scan path progressing in the same direction.
Performance and efficiency measures	
TTA: *Time to answer*	Indicates the *time* spent by the subject until answering the question.
Cognitive processing measures	
TNF: *Total number of fixations* (*in all the diagram*)	A higher number of fixations indicates *confusion* (*less efficient search* of the solution to the task).
AFD: *Average fixation duration*	A longer duration indicates that the participant needs more time to *understand* the *individual objects*. This can indicate *difficulty* in *extracting information*.
Solution* (*or target*) *findability measures	
Number of participants who were fixated on the target	Indicate how *quickly* an element is *located* in the image. It can indicate *where* a particular content *is expected to be found*.
TTFFT: *Time to first fixation on the target*	
NFPFFT: *Number of fixations prior to first fixation on the target*	
Solution* (*or target*) *recognizability measures	
NVT: *Number of visits to target* (*area that contains the solution*) *before answer*	Measures of *recognizability* of the searched target (in our case, the solution to the task)
TFFFTA: *Time from first fixation on target to answer*	

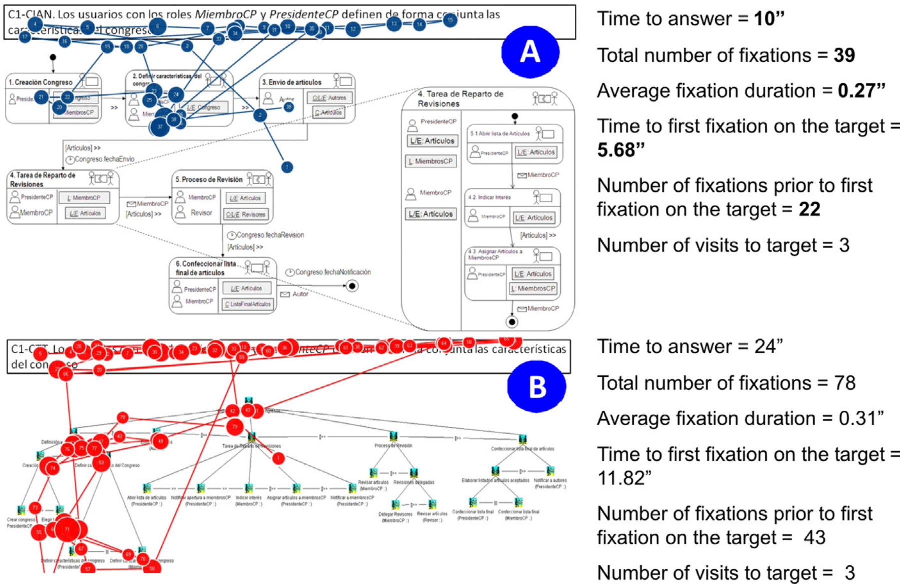


Fig. 7. *Scan paths* and some *metrics* generated during the pilot test (by two particular participants) in a *task of localization of the relationship between tasks and roles*.

be performed. So, in our case, the AOIs are the areas of the diagrams containing the information required to answer each of the comprehension tasks to be solved by participants.

To better understand the analysis and interpretation of these metrics, we show two examples. In the first one (Fig. 7), we show an exercise which attempts to measure the "*ease of locating and understanding the relationship between tasks and the roles that perform those tasks*". We can see the *scan paths* of two of the participants in the pilot test who answered the same question but analyzing different diagrams (in CTT and CIAN notations). The figure also shows some of the metrics collected. In this particular case, and for these two particular participants, the results indicated that the one who analyzed the CIAN version spent less *time* on localizing the solution of the task and responding (10″). The metrics that we use to measure the *cognitive load* (the *total number of fixations* and the *average fixation duration*) has lower values in the case of the CIAN diagram (39 fixations and 0.27″, respectively) compared to the CTT diagram (78 fixations and 0.31″, respectively). The time to first fixation on the area of the diagram that contains the solution and the number of fixations prior to first fixation on this area

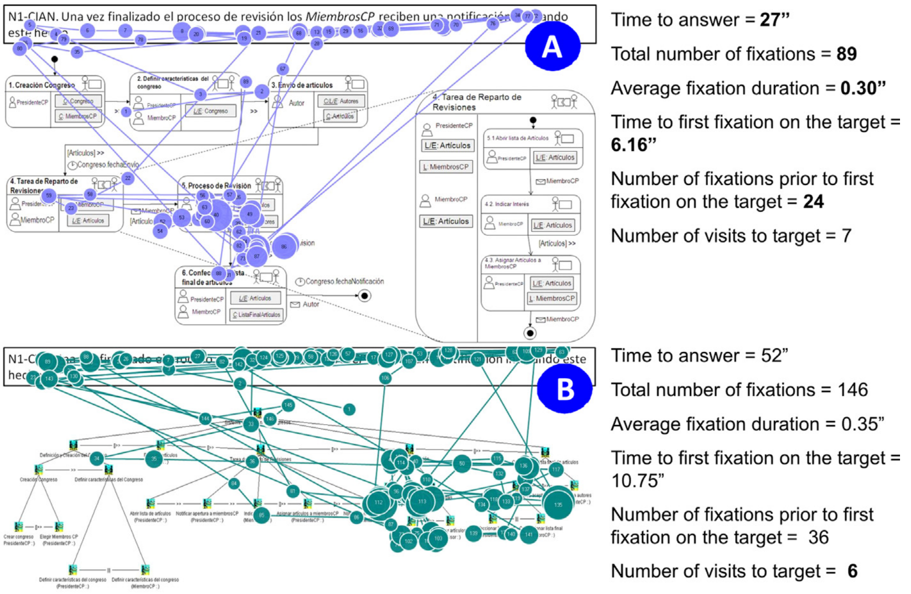


Fig. 8. *Scan paths* and some *metrics* generated during the pilot test (by two particular participants) in a *task of localization of notifications*.

are also lower in the case of the CIAN notation (5.68″ and 22 fixations, respectively). This fact indicates that the participants who analyzed this version found the answer to the question more quickly than those who analyzed the CTT version (who spent 11.82″ and generated 43 fixations). In relation to the *number of visits to target* (that can be interpreted as *recognition* of the solution), the values obtained did not indicate any difference (in both cases the number of visits was the same: 3).

We can try to explain these results considering the characterization of both notations made in Section 5.1.1, using the cognitive dimensions framework. In this case, the task was “*locating and understanding the relationship between tasks and the roles that perform those tasks*”. With regard to specification of relationship between roles and task, both notations show this relationship graphically. In this regard, neither of them presents *hidden dependencies*. In the case of CTT, the name of the task and the name of the role appear together and close in the diagram. In the case of CIAN, a task graphically includes some elements (concepts related with the performing of the task): the roles involved in its execution and the object manipulated. All these graphical models form a unique and more complete symbol, in which the layout of the task makes this relationship more explicit (*secondary notation*).

In addition, regarding the roles representation, in the case of CIAN the notation provides a specific icon to represent them. This graphic symbol is combined with the name of the role (in text). In the case of CTT, the name of the role appears close to the task (in text), but without a particular icon accompanying the name. According to *dual coding theory* [94], using text and graphics together to convey information is more effective than using either one on its own. The symbol used to specify roles in CIAN facilitates *finding* and *recognizing* more directly this particular element in the whole diagram.

In Fig. 8, we can see the *scan paths* and the metrics of two of the participants in the pilot test who answered a question related to the “*locating and understanding of notifications*” in CTT and CIAN diagrams. As we can see in Fig. 8, in the case of CTT (8.B) it is necessary to scan several levels of the diagram to locate the notification task, while in the case of CIAN (8.A) the location of this element is more direct. We can also interpret the metrics in the same way as we have done in the previous example. Additionally, in this case, the participants spent less *time* on localizing the solution of the task and responding (27″). The *cognitive load*, measured using the *total number of fixations* and the *average fixation duration*, has lower values in the case of the CIAN diagram (89 fixations and 0.30″, respectively) compared to the CTT diagram (146 fixations and 0.35″, respectively). The time to first fixation on the area of the diagram that contains the solution and the number of fixations prior to first fixation on this area are also lower in the case of the CIAN notation: 6.16″ and 24 fixations in CIAN vs. the 10.75″ and 36 fixations in the case of the CTT diagram. This fact indicates that the participants who used the CIAN version found the answer to the question in a more straightforward way than those who analyzed the CTT version.

Trying to explain the reason for these results, we can again associate them with the support of cognitive dimensions of both notations. In the case of notification specification, CIAN provides a specific icon to express this kind of restriction. CTT does not include a particular graphical representation of this kind or constraint, making it necessary to specify it as a specific task in the workflow. The use of a specific icon, in the case of CIAN, facilitates the identification and recognition of this element in the diagram. In this regard, we consider that the *visibility* cognitive dimension of this particular concept is better in the case of CIAN than in the case of CTT. In this sense, and in order to maximize visual expressiveness of the notation, graphical encoding (using a specific icon) is preferred and is more efficient than textual encoding.

Once the exercises were created and the metrics to be obtained were selected, we proceeded to the final test. In the next subsection, we explain all the details of this experiment and the results obtained from it.

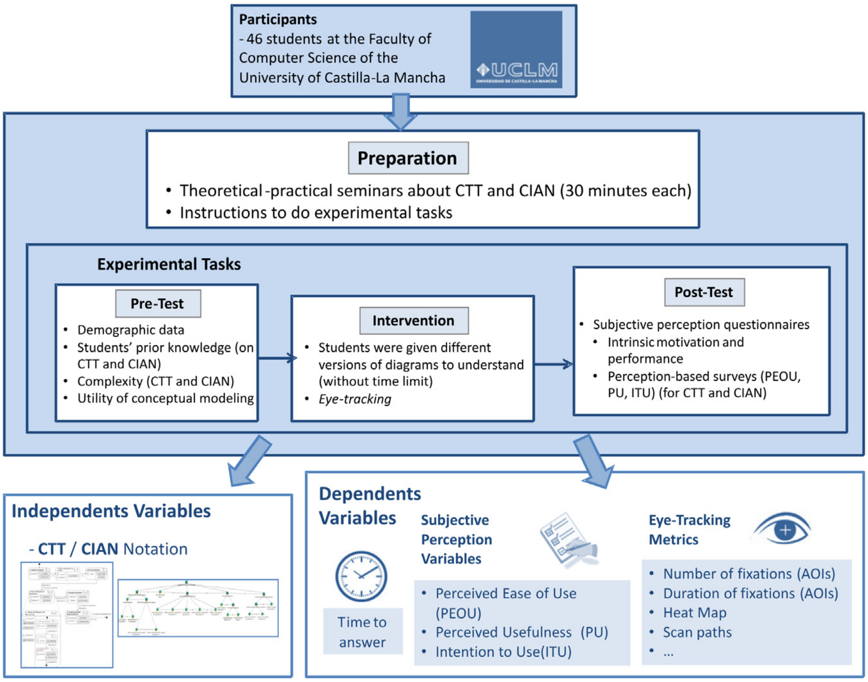


Fig. 9. Experimental design.

5.3.3. Experimental design

Fig. 9 graphically illustrates the *experimental design* followed in this study. Each of the participants was summoned at different times to perform the experimental task. The duration of the test for each subject was about 10–12 minutes. In the design and development of this *eye tracking* test, we followed the *methodological recommendations* of Nielsen and Pernice [88]. For each student, the test consisted of several phases: *pre-test*, *eye tracker calibration*, *intervention phase* and *post-test*. The objective of each of these stages, as well as the information collected in each, is described below. In Table 7, we summarize the dependent and independent variables considered in this experiment.

Firstly, the participants took a ***pre-test*** (on paper) by which researchers could gather certain data about the students' *profile*:

- ***Demographic information***: gender, age.
- ***Prior Knowledge* (*PK*) *on CTT* (*for collaborative modeling*) *and CIAN notations***. In order to assess *prior knowledge* of participants in the two notations, they were asked to rate their knowledge on a Likert 5-point scale ranging from "*I don't know at all*" (associated with a score of 1) to "*I know very well*" (associated with a score of 5) for one statement related to their knowledge about these notations.
- ***Complexity* of *CTT* (*for collaborative modeling*) *and CIAN notations***. Participants were also asked to rate the complexity of each of the notations studied on a 5-point scale (ranging from 1 which means "*very simple*' to 5 which means "*very complex*').
- ***Utility of conceptual modeling.*** Finally, participants indicated their opinion about the utility of using conceptual modeling (notations) for specifying software systems in general, and collaborative systems in particular. Both aspects were rated using a 5-point scale ranging from "*Hardly useful*" (associated with score 1) to "*Very useful*" (associated with score 5).

Then, the participants moved on to the ***intervention phase***. The experiment rules and the characteristics of the materials were explained before this phase began. No other explanation was given during the intervention. Before visualizing and analyzing the diagrams, the *eye tracker* was calibrated, passing by to record the session. *Tobii Studio* software records the task and tracks the time, in seconds, taken by each student to complete the task (**TTA**: ***Time to Answer***). Each of the participants spent the *time* he/she considered necessary, with no limitation in this regard.

After completing the intervention phase, all participants were asked to fill out the ***post-test*** which consisted of several sections:

- ***Selection of notation* (*CIAN vs. CTT*) *which more easily located the three aspects studied.*** In the *post-test*, the participants chose the notation (CTT or CIAN) that he/she considered more easily located the answer to the questions related to the

Table 7
Summary of the dependent and independent variables considered in the first experiment.

Type	Subtype	Name	Measure
Independent variables		Notation	CTT or CIAN diagram
Dependent variables	Performance variables	Time to answer (TTA)	Indicates the *time* spent by the subject until answering the question
	Perception-based variables	Perceived Ease of Use (PEOU)	Mean of the items that measure each of the items PEOU1...PEOU4 obtained in the survey proposed in [13] (Table 3)
		Perceived Usefulness (PU)	Mean of the items that measure each of the items PU1...PU4 obtained in the survey proposed in [13] (Table 3)
		Intention to Use (ITU)	Mean of the items that measure each of the items ITU1...ITU4 obtained in the survey proposed in [13] (Table 3)
	Eye-tracking metrics	Total number of fixations (TNF)	Total number of fixations in all the diagram
		Average fixation duration (AVD)	Average value of duration of each individual fixation within an AOI
		Time to first fixation on the target (TTFFT)	This metric measures how long it takes before a test participant fixates on an AOI for the first time
		Number of fixations prior to first fixation on the target (NFPFFT)	This metric measures the number of times the participant fixates on the media before fixating on an AOI for the first time
		Number of visits to target (NVT)	This metric measures the number of visits within an AOI. Each individual visit is defined as the interval of time between the first fixation on the AOI and the next fixation outside the AOI
		Time from first fixation on target to answer (TFFFTA)	This metric measures the time from first fixation on an AOI until the subject answers

Table 8
Intrinsic *motivation* and *performance* subjective ratings.

Item	Measure (*5-point Likert scale*)
COM1. I think I did this activity well.	Competence (COM)
EFF1. I put a lot of *effort* into doing this activity well.	Effort (EFF)
EFF2. It was *important* for me to do this task well.	Effort (EFF)
PRE. I felt *nervous* while doing this activity.	Pressure (PRE)
CHO. I willingly chose to participate in this activity.	Choice (CHO)

understanding of "*the relationship between tasks and the roles that perform those tasks*", "*the specification of notifications*" and "*the flow/order between tasks*".

- **Intrinsic *motivation* and *performance* subjective ratings.** Next, participants completed a set of 5 questions (5-point Likert scale) designed to measure their *competence* (**COM**), *effort* (**EFF**), *pressure* (**PRE**) and *choice* (**CHO**) during the activity [95] (Table 8).
- ***Perception-based survey.*** Finally, for each of the two notations (CIAN and CTT), participants used a Likert scale (from 1 to 5) to show their level of agreement or disagreement with 15 statements used to measure the *perceived ease of use* (**PEOU**), *perceived usefulness* (**PU**) and the *intention to use* (**ITU**) for each notation. The survey used to measure these three aspects for each notation is the same one used in the second experiment (based on the questionnaire proposed in [13]).

5.3.4. The final test: Results and discussion

The ***subjects*** in the final experiment were 56 undergraduate students enrolled in the third and fourth years of the Bachelor of Computer Science Degree at the University of Castilla-La Mancha in Ciudad Real, Spain. The participation in this experiment was voluntary. The average age of the participants was 23. As we have mentioned before, some extra recruiting is needed to ensure enough data is collected due to the fact that about 5% of the participants cannot be eye tracked.

Participants attended a seminar on CTT and CIAN (for cooperative modeling) notations. Some examples of understandability exercises were also given during this previous phase. Then, they solved several comprehension exercises. In this experiment, we wanted to evaluate three aspects: the "*ease of understanding the flow or order between tasks*", the "*ease of locating and understanding the relationship between tasks and the roles that perform those tasks*" and, finally, the "*ease of locating and understanding the specification of notifications*". To evaluate each aspect, an understandability exercise of each type was presented. Half of the participants received the diagrams in CTT and the other half in CIAN. The subjects were randomly assigned to each group. While they were completing the exercises of interpretation and analysis of the diagrams, the *eye*

Table 9
Participants profile (some *pre-test* results).

	CIAN notation[a]	CTT notation[a]	F*	p*
Prior knowledge (PK)	2.14 (0.70)	**3.14 (0.72)**	55.28	**0.00**
Complexity (Comp)	**3.11 (0.78)**	2.84 (0.87)	2.95	0.08

[a] We show the *mean* scores and the *standard deviations* (in parentheses).
* Results of the *analysis of variance* for the data with a confidence level of 95% ($\alpha = 0.05$).

Table 10
Utility of conceptual models (*pre-test*). Intrinsic *motivation* and *performance* subjetive measures (*post-test*).

	Mean (SD)[a]
Conceptual model utility	3.43 (0.83)
Conceptual model utility (for specifying collaborative systems)	3.63 (0.80)
Competence (COM)	3.63 (0.80)
Effort (EFF)	4.21 (0.72)
Pressure (PRE)	2.32 (1.16)
Choice (CHO)	4.82 (0.62)

[a] We show the *mean* scores and the *standard deviations* (in parentheses).

tracker equipment scanned and recorded their eye movements performed while scanning the image and looking for the answers to the comprehension activities.

We then analyzed and interpreted the information collected during the experiment. As discussed previously, it includes values obtained from different sources. Some of them are of an objective nature (TTA and the metrics provided by the *eye tracker* device), while others are obtained from subjective surveys (PK; PEOU, PU and ITU for the CTT notation; PEOU, PU and ITU for the CIAN notation and the intrinsic motivation and performance variables: COM, EFF, PRE, and CHO).

Regarding the ***profile*** of respondents, it is necessary to indicate that most of them had prior knowledge of using the CTT notation for specifying interactive systems. The participants were recruited from two subjects (*Human–Computer Interaction* and *Groupware Engineering*), and thus they had some knowledge of *task modeling*, with CTT being a *de facto* standard in this field. However, they had no prior knowledge about its use for modeling cooperative behaviors [17]. We indicated and commented on the difference between CTT for classical use (to specify interactive behavior) and its use for modeling cooperative behaviors. However, we believe the students did not consider this difference when answering this question. In Table 9, we can see the mean value assigned by participants to prior knowledge on both notations ($M = 2.14$, $SD = 0.70$ for CIAN; $M = 3.13$, $SD = 0.72$ for CTT), with the difference being statistically significant ($F = 55.28$, $p = 0.00$).

A *pre-test* was used not only to ascertain the participants' profiles, but also to determine their subjective perception about the *complexity* of both notations. Most participants considered the CIAN notation as more complex ($M = 3.11$, $SD = 0.78$) than CTT ($M = 2.84$, $SD = 0.87$). However, previous knowledge in CTT could have influenced their answer and the difference was not statistically significant ($p = 0.08 > 0.05$). However, some students mentioned that CIAN has a higher number of icons and perhaps they associated completeness as complexity.

In the *pre-test*, we also asked about the perceived ***utility of conceptual modeling*** for specifying software systems in general, and for specifying collaborative systems in particular. In this regard, participants scored the utility of using notations for modeling with a mean of 3.43 ($SD = 0.83$), valuing the utility for modeling specifically collaborative systems with a mean of 3.63 ($SD = 0.80$) (Table 10).

When the participants had completed the activity they filled out the *post-test* surveys. In Table 10, we can see some of the mean values assigned to ***subjective perception measures***. We can observe that participants considered themselves very *competent* in performing this kind of activity ($M = 3.63$, $SD = 0.80$). They made an *effort* to properly complete the activity ($M = 4.21$, $SD = 0.72$), not feeling very *nervous* when doing it ($M = 2.32$, $SD = 1.16$). Most of them willingly *chose* to participate in this session ($M = 4.82$, $SD = 0.62$).

As we have mentioned before, we are interested in knowing the *subjective perception* of the participants in relation to the *perceived ease of use* (PEOU), *perceived usefulness* (PU) and the *intention to use* (ITU) for each notation. This data was collected in the *post-test*. In Table 11, we can see the results obtained for the CIAN and CTT notations. This data indicated that better scores were obtained for the CIAN notation compared to CTT for the three measures. In the case of variable PEOU, the score for CIAN was on average 3.71 ($SD = 0.72$) as opposed to 3.51 ($SD = 0.58$) for the CTT notation. However, the difference is not statistically significant ($p = 0.11 > 0.05$). The score for *perceived usefulness* of the CIAN notation was 3.90 ($SD = 0.56$) compared to 3.44 ($SD = 0.58$) for the CTT notation, the difference in this case being significant. Finally, the ITU score was again higher for CIAN ($M = 3.58$, $SD = 0.69$) versus CTT ($M = 3.16$, $SD = 0.73$), the difference here also being significant ($F = 9.53$, $p = 0.00 < 0.05$). This positive subjective perception about CIAN differs from the fact that participants had higher prior knowledge of CTT. In fact, we made a correlation analysis between the measures gathered in the *pre-test* and the

Table 11
Subjective perception measures (*post-test*).

	CIAN notation[a]	CTT notation[a]	F^*	p^*
Perceived Ease of Use (PEOU)	**3.71 (0.72)**	3.51 (0.58)	2.64	0.11
Perceived Usefulness (PU)	**3.90 (0.56)**	3.44 (0.58)	18.47	**0.00**
Intention To Use (ITU)	**3.58 (0.69)**	3.16 (0.73)	9.53	**0.00**

[a] We show the *mean* scores and the *standard deviations* (in parentheses).
* Results of the *analysis of variance* for the data with a confidence level of 95% ($\alpha = 0.05$).

Table 12
Eye tracking metrics generated in the *task of localization of the relationship between tasks and roles*.

Eye Tracking measures	CIAN notation[a]	CTT notation[a]	F^*	p^*
Time To Answer (TTA)	**44.98** (**24.81**)	46.17 (17.32)	0.035	0.85
Total number of fixations (in all the diagram) (TNF)	81.13 (48.39)	**76.74** (**33.88**)	0.13	0.72
Average fixation duration (AFD)	**0.32** (**0.06**)	0.36 (0.06)	3.70	0.06
Time to first fixation on the target (TTFFT)	**6.74** (**6.28**)	20.09 (9.66)	30.89	**0.00**
Number of fixations prior to first fixation on the target (NFPFFT)	**23.78** (**21.75**)	66.52 (29.64)	31.09	**0.00**
Number of visits to target (area that contains the solution) before answer (NVT)	7.96 (3.84)	**3.61** (**2.25**)	21.92	**0.00**
Time from first fixation on target to answer (TFFFTA)	38.24 (25.33)	**26.08** (**14.40**)	4.01	**0.04**

[a] We show the *mean* scores and the *standard deviations* (in parentheses).
* Results of the *analysis of variance* for the data with a confidence level of 95% ($\alpha = 0.05$).

subjective perception measures, obtaining a negative correlation between the complexity score assigned to CIAN and PEOU ($r = -0.41$, $p < 0.01$[6]) and PU ($r = -0.33$, $p < 0.05$[6]) scores assigned to this notation.

To obtain all the data mentioned so far, we considered the data provided by the 56 participants that voluntarily participated in this study. We will now discuss the data obtained with the *eye tracker* device and relating the metrics obtained with the data collected using the surveys. We should point out that we eliminated 10 participants from the initial sample, considering a final sample of 46 participants. The reason is that, as we have discussed above, some people cannot be eye tracked at all for various reasons (people wearing bifocals, contact lenses, and/or poor trackability). All of these were excluded from the tests. This restriction affects the sample size, because the initial number of participants recruited decreased. Only participants with a *calibration accuracy* higher than 90%[7] ($M = 93.1\%$, $SD = 5.29$) in the session were selected in the final sample [96].

As previously mentioned, in this study we were interested in testing various aspects related to the *understandability* of each notation. One of these is the "***ease of locating and understanding the relationship between tasks and the roles that perform those tasks***" in the CTT and CIAN notations, to check if the latter was more understandable to users. We will now comment on the results obtained in this comprehension exercise (Table 12). In the *post-test*, we asked the participant which notation (CIAN or CTT) he/she considered more easily located the answer to this kind of task. Of the 46 respondents, 73.91% selected CIAN as easier. Now we will contrast this subjective opinion with the data and metrics collected during this exercise.

One of the values that allows us to compare the *understandability* of the diagrams is the *time* spent by participants on interpreting and analyzing each model. That is, the time spent by the subject on answering the question (*time to answer* metric, TTA). The test results indicated that the time for locating the answer was lower in the case of CIAN ($M = 44.98$, $SD = 24.81$) compared to CTT diagram ($M = 46.17$, $SD = 17.32$). In fact, participants appreciated the use of icons representing roles ( ), because they helped participants to more quickly locate the answer. As we inferred from the pilot test, the use of this specific graphical symbol facilitates *finding* more directly this particular element in the whole diagram. In this sense, CIAN presents a better *visibility* (Section 5.1.1) than the CTT notation.

Regarding TTFFT (*time to first fixation on the target*) and NFPFFT (*number of fixations prior to first fixation on the target*) metrics which indicated how quickly the solution is located in the diagram, participants obtained lower values in the case of CIAN ($M = 6.74$, $SD = 6.28$; $M = 23.78$, $SD = 21.75$, respectively), with the difference in the data obtained in the case of the CTT notation being statistically significant ($p = 1.49\text{E}{-}06$ and $p = 1.41\text{E}{-}06$, respectively; with both values being lower than 0.05). These objective *findability* measures are consistent with the perception of the majority of the participants, who considered that CIAN facilitated the finding of the answer to this kind of understandability tasks.

As regards *cognitive processing* measures, we can conclude that the layout of CIAN diagrams generates less efficient searches due to the fact that the *total number of fixations* (TNF) when exploring a CIAN model ($M = 81.13$, $SD = 48.39$) is higher than when exploring a CTT model ($M = 76.74$, $SD = 33.88$). However, the other measure that we can use to

6 Minimum significant correlation coefficient r for sample size $n = 56$ is 0.26 at $p = 0.05$; minimum significant correlation coefficient r for sample size $n = 56$ is 0.34 at $p = 0.01$; minimum significant correlation coefficient r for sample size $n = 56$ is 0.43 at $p = 0.001$.

7 The *quality* of the recording is calculated using the number of *eye tracking* samples that were correctly identified. 100% means that both eyes were found throughout the recording; 50% that one eye was found for the full recording or both eyes during half the time. Note that the eyes cannot be found when a person is looking away from the screen; this will result in a lower percentage.

Table 13
Eye tracking metrics generated in the *task of localization of notifications*.

Eye Tracking measures	CIAN notation[a]	CTT notation[a]	F^*	p^*
Time To Answer (TTA)	**35.30** (**16.53**)	38.83 (19.11)	0.45	0.51
Total number of fixations (in the entire diagram) (TNF)	67.57 (44.13)	**61.61** (**32.60**)	0.27	0.61
Average fixation duration (AFD)	**0.31** (**0.04**)	0.38 (0.07)	14.81	**0.00**
Time to first fixation on the target (TTFFT)	**10.21** (**5.35**)	14.33 (6.60)	5.41	**0.02**
Number of fixations prior to first fixation on the target (NFPFFT)	**30.57** (**15.57**)	40.30 (16.77)	4.19	**0.04**
Number of visits to target (area that contains the solution) before answer (NVT)	**6.17** (**4.68**)	8.43 (4.73)	2.66	0.11
Time from first fixation on target to answer (TFFFTA)	**22.70** (**17.62**)	30.41 (19.09)	2.02	0.16

[a] We show the *mean* scores and the *standard deviations* (in parentheses).
* Results of the *analysis of variance* for the data with a confidence level of 95% ($\alpha = 0.05$).

determine the *cognitive load* during the comprehension task (the *average fixation duration*) is lower in the case of CIAN ($M = 0.32$, $SD = 0.06$). This last measure, as is shown in Table 12, indicates (when the duration is longer) that the participants need more time to understand the individual objects. This occurs in the case of CTT diagrams. Therefore, we can conclude that, for this type of understandability exercise, the CIAN notation produces less efficient scanning patterns, although the extracting of information when the solution is located is easier.

The last two measures (NVT and TFFFTA) related to the *recognizability* of the solution, are better in the case of CTT ($M = 3.61$, $SD = 2.25$ and $M = 26.08$, $SD = 14.40$, respectively), with the difference between the two metrics being statistically significant. Considering the values of the *findability* measures (TTFFT and NFPFFT) and the *recognizability* ones (NVT and TFFFTA), we can conclude that the use of icons to represent roles facilitates the finding of the answer in the case of CIAN, although the subject needed more time to be sure before answering (he/she needs to visualize more elements). In the case of CTT, the location of the answer is not so direct but when the subject has located the solution of the comprehension task, he/she is sure of the answer more quickly.

With regard to the specification of the "*relationship between tasks and the roles that perform those tasks*", neither CIAN nor CTT presents *hidden dependencies* (Section 5.1.1) showing explicitly and visually the relationship between roles and tasks. In the case of CIAN, the use of graphical icons for representing roles facilitates *locating* them (that is, the *visibility* of these elements is improved). The relationship between roles and tasks are expressed in CIAN using a more complex graphical element (roles are nested elements in a task representation). Roles and tasks appear together in the diagram, with the *layout* of the graphical representation of the task acting as a *secondary notation* which graphically expresses the relationship between these concepts. The complexity of the elements used for representing tasks can explain the values of the *recognizability* metrics provided by the *eye tracker*, which are worse in the case of CIAN.

Regarding the "***ease of locating and understanding notification specifications***" significant differences were noted between the two notations considered. In the case of CIAN, an icon exists (✉) to graphically specify notifications in the flow of activities. To model this element in CTT, it is necessary to create a notification task which is not graphically different from other tasks. In this case, we found significant differences in the interpretation of both notations.

In the question related to the ease of locating notifications (filled out in the *post-test*), 35 of the 46 participants (76.09%) considered that it was easier to locate notifications in CIAN models. However, the measures provided by the *eye tracker* allow us to refine this subjective perception (Table 13). In relation to the *time to answer*, the lower values have been obtained by the CIAN version ($M = 35.30$, $SD = 16.53$), although the difference is small and not significant ($p = 0.51 > 0.05$).

Regarding the two metrics related to the *cognitive processing* during the analysis exercise (*total number of fixations in the diagram* and *average fixation duration*), again, the values obtained are consistent with those obtained in the previous type of exercises. The *total number of fixations* is higher in the case of CIAN ($M = 67.57$, $SD = 44.13$) while the *average fixation duration* is lower ($M = 0.31$, $SD = 0.04$), with the difference in the last case being statistically significant ($F = 14.81$, $p = 0.0004 < 0.05$).

In relation to *findability* measures (TTFFT and NFPFFT), the better scores are obtained by the CIAN notation ($M = 10.21$, $SD = 5.35$ and $M = 30.57$, $SD = 15.57$), with the differences being significant in both cases ($p = 0.02$ and $p = 0.04$, respectively, both values lower than 0.05). The use of specific icons for representing notifications facilitates finding the answer more quickly in CIAN diagrams.

As regards the *recognizability* measures (*number of visits to target* and *time from first fixation on target to answer*), they are also better in the CIAN notation ($M = 7.17$, $SD = 4.68$ and $M = 22.70$, $SD = 17.62$) although in these cases the difference is not significant ($p = 0.11$ and $p = 0.16$, respectively). These values indicate that once participants have located the answer to the exercise in the CIAN model, they have fewer doubts about it before answering. Once the participants have found the notification icon (thanks to its *visibility*) they can also recognize this element quickly, thanks to the *closeness of mapping* that characterizes most of the graphical elements provided by CIAN.

Finally, we are going to comment on the results obtained in the last type of exercise, which tried to assess the "***ease of understanding the flow or order between tasks***". Participants were asked about this issue in the *post-test*, with the CTT notation obtaining the better score in this case (58.70% of the respondents thought that in CTT it was easier to interpret and understand the flow between tasks). We want to contrast this subjective perception with the data provided by the *eye tracker* (Table 14).

Table 14
Eye tracking metrics generated in the *task to understand the flow or order between tasks*.

Eye Tracking measures	CIAN notation[a]	CTT notation[a]	F*	p*
Time To Answer (TTA)	39.78 (22.54)	**32.78** (**10.43**)	1.83	0.18
Total number of fixations (in the entire diagram) (TNF)	66.83 (42.91)	**45.78** (**24.56**)	4.17	0.05
Average fixation duration (AFD)	**0.31** (**0.06**)	0.37 (0.07)	**8.15**	**0.01**
Time to first fixation on the target (TTFFT)	**8.86** (**5.20**)	9.69 (6.11)	0.25	0.62
Number of fixations prior to first fixation on the target (NFPFFT)	**32.74** (**18.74**)	36.00 (20.37)	0.32	0.57
Number of visits to target (area that contains the solution) before answer (NVT)	8.70 (4.30)	**6.13** (**3.47**)	**4.96**	**0.03**
Time from first fixation on target to answer (TFFFTA)	30.92 (22.73)	**23.09** (**10.94**)	2.22	0.14

[a] We show the *mean* scores and the *standard deviations* (in parentheses).
* Results of the *analysis of variance* for the data with a confidence level of 95% ($\alpha = 0.05$).

The *time to answer* metric, which can be interpreted as an *efficiency* measure, indicates that the time spent by the subjects before answering was lower in the case of analyzing CTT diagrams ($M = 32.78$, $SD = 10.43$), although the difference was not significant ($p = 0.18 > 0.05$). This value is consistent with the perception of the majority of subjects that rated CTT models as better than CIAN for this kind of understandability task.

The values obtained in the measures related to *cognitive load* are consistent with those obtained in the two previous cases. The *total number of fixations* in the entire diagram is lower in the case of the CTT notation ($M = 45.78$, $SD = 24.56$), while the better value of *average fixation duration* is obtained by the CIAN notation ($M = 0.31$, $SD = 0.06$), with the difference in the last case being significant ($p = 0.01 < 0.05$).

The two metrics related to the *findability* of the answer (TTFFT and NFPFFT), and contrary to expectations (and the subjective perception of the majority of participants), indicates that the answers were located more quickly in the CIAN version of the exercise although the differences between the two metrics are not statistically significant ($p = 0.62$ and $p = 0.57$, both values higher than 0.05).

Finally, with regard to *recognizability* of the solution (*number of visits to target* and *time to first fixation on target to answer*), the better scores were obtained in this case by the CTT notation ($M = 6.31$, $SD = 3.47$; $M = 23.09$, $SD = 10.94$), with the difference in the first metric being statistically significant (NVT) ($p = 0.03 < 0.05$).

As a result of the data obtained in the three exercises, we can extract the following ***conclusions***:

- In relation to the values of TTA (*time to answer*), which can be considered a *performance* and *efficiency* measure, the obtained data is consistent in the three exercises with the preferences of participants. They rated CIAN as a better option in order to understand the relationship between tasks and roles and to locate notifications, while they preferred the CTT notation for understanding the flow or order between tasks. The objective data provided by the *eye tracker* seems to confirm the subjective perception of participants in terms of this issue.
- Regarding the *cognitive processing* measures, in all the cases the *total number of fixations* (TNF) is higher in the case of CIAN, which can be interpreted as the diagrams specified in this notation causing more confusion and, therefore, less effective scanning patterns. We believe that the fact that the participants were already familiar with CTT may have influenced these results. However, in all cases the value of AFD (*average fixation duration*) is lower in the case of CIAN, which can be interpreted as this notation facilitating the extraction of information. That is, the participants need less time to understand the individual objects.
- Finally, in the case of CIAN, the participants more quickly *found* the solution in the three exercises (lower values of the *findability* metrics: TTFFT and NFPFFT), although CTT has better data in the solution *recognition* in the first and third types of comprehension exercises (NVT and TFFFTA). Thus, in these cases, although CIAN facilitates *finding* particular contents, the participants need to analyze the diagram for more time to be sure before answering.

5.4. Discussion and lesson learned

Our objective in this work was to assess a notation (CIAN) for modeling some aspects of collaborative learning systems. CIAN, like most requirements notations, is highly visual. Visual notations play a critical role in communicating with end user and customers, mainly to convey information to non-technical people. Empirical research in requirements engineering contexts confirms that the graphical form of notations significantly impacts their effectiveness for both problem solving and end user communication [97]. In designing notations, decisions about graphic representations tend to be made in most cases in a subjective way (i.e., they are based on the decisions of researchers and notation designers), without reference to theory (for example, principles of visual perception and cognition) or empirical evidence that justify these decisions [65].

As we mentioned in the introduction of this paper, one drawback of most requirements modeling methods is that they lack proper evaluations, mainly of an empirical nature. There are several methods and techniques for assessing graphical notations *quality* (as we have commented on in Section 4). Some of them are more objective while others are more subjective. Some authors also propose *guidelines* and *design rules* for graphical notations. Features such as simplicity, aesthetics, expressiveness, and naturalness are often mentioned in the literature, but they are vaguely defined and highly subjective.

In this work, we began characterizing our proposed notation for modeling collaborative learning tasks (CIAN) using one of the most popular among the existing theoretical frameworks: the *Cognitive Dimensions* framework. We selected and applied only the dimensions that we consider have influence on the *understanding* of models (the main quality aspect considered in this work). But, we also wanted to contrast our own particular opinion regarding the dimensions provided by this framework with the opinions of potential users of our proposal. For this purpose, we ran three empirical and controlled experiments. The results in most of the cases were aligned with our particular opinion.

One of the dimensions included in this framework is that of ***hidden dependencies***. In the introduction section, we enumerated the aspects we consider that a graphical notation for modeling CSCL systems must include. One of them was "support to *express graphically and jointly* the relationship between elements specified (*activities*, *roles* and *resources*)". This feature is related to the aforementioned cognitive dimension. CIAN allows showing, in the context of a task, the roles and data objects manipulated in the context of such a task. The relationship between these concepts can be inferred based on their perceptual (visual) configuration, due to the fact that these elements appear graphically together in the diagrams. We consider that it is one of the strong points of the *process model* in the CIAN notation, and considering the results obtained in the three experiments this is one of its positive features. From this result, we can state the following guidelines that researchers who want to define a visual notation for collaborative modeling can consider if they want to improve the *understandability* of the diagrams created using their notation:

Guideline 1: In order to reduce the *hidden dependencies* in the notation, the relationship between the main concepts to be expressed must be explicitly and visually shown in the diagram. In the case of CSCL systems, these main elements are: *roles*, *learning tasks* and *resources* (*data* and support *tools*).
Guideline 2: For graphically expressing this relationship, consider applying the *Gestalt laws* of Perceptual Organization (proximity, continuity, etc.) [98] to configure the visual elements.

The second cognitive dimension that we have assessed in CIAN and CTT in the empirical studies is the ***closeness of mapping***. This feature facilitates (as we have contrasted using the *findability* metrics provided by the *eye tracker* device) the *location* of particular elements in a diagram during *understandability* tasks. In addition, this feature allows reducing the *cognitive load* (*mental effort*) of users during comprehension tasks. So:

Guideline 3: Where possible, when selecting visual representations for the main concepts in the visual language to be created, it is recommendable to use visual representations whose appearance suggests the meaning of the symbol.

Another cognitive dimension which we consider related, to a certain extent, with the previous one is the called ***role expressiveness*** (related to the role that each particular graphical element plays in the overall diagram). This feature facilitates the *recognition* and interpretation of individual elements, and it is related to the ability of a particular symbol to be easily *discriminable*. In this sense, the chosen icons must be visually distinguishable from other symbols.

In this regard, we think that CIAN promotes the perceptual discrimination between graphic symbols. For example, in the case of the icons proposed to specify several types of tasks (individual, cooperative and collaborative), CIAN proposes three different and easily distinguishable icons. In the case of CTT, the icons proposed to graphically represent individual (interaction) and cooperative tasks are visually difficult to distinguish. Discriminability is determined by the *visual distance* between symbols, which can be measured by the number of visual variables (shape, size, color, brightness, orientation and texture) on which they differ [65]. The greater the visual distance between symbols, the faster and more accurately they will be recognized [99]. The icons proposed by CTT for individual and cooperative tasks are very similar, due to the fact that they are the same size and shape. Considering these issues we can formulate the following guideline:

Guideline 4: When selecting icons in a visual language, in order to reduce their physical complexity and improve their *usability* and *effectiveness*, they must be characterized by their *clarity* (legibility), *simplicity* (without unnecessary embellishments, colors or details) and *discriminability*.

The next dimension is related with the ***visibility*** of the elements. We consider that the main concepts to be considered in the specification of collaborative learning processes must be supported graphically and be visible in the created models. For example, in the case of CTT, the data objects used in the context of a particular task can be specified (using the CTTE editor) but have no visual representation in the diagrams (they are not visible). The elements that we consider it necessary to express graphically in the specification of a CSCL system are enumerated in the following guideline:

Guideline 5: A visual language for modeling collaborative learning systems must propose a visual representation (a graphical symbol) for each of the main concepts to be specified in this kind of system. We propose to graphically represent at least the following: *organizational members*; *learning tasks* (distinguishing between individual, cooperative and collaborative tasks) and the *resources* to be used in the context of such tasks (supporting tools and information resources); *temporal relationships*; *notifications*; *events* and *information passing*.

Taking this into account, we have to ensure that we fulfill the so-called *principle of semiotic clarity*, that states that there must be a one-to-one correspondence between semantic constructs and graphical symbols, avoiding *symbol redundancy* (multiple icons representing the same concept), *symbol overload* (different constructs represented by the same graphical icon), *symbol excess* (when there are symbols that do not correspond to any semantic construct) and *symbol deficit* (when there are concepts that are not represented by any symbol) [65]. This last anomaly (symbol deficit) occurs in the case of specification of notifications or data objects in the case of CTT. These elements are not graphically represented in CTT models. In the case of CTT, the specification of notifications must be expressed by a task labeled, for example, as "*Notification to* ...". Users must read the labels of the task to identify it as a "notification task". The differentiation between a regular task and a notification task is textually expressed, but not graphically. This fact reduces the discriminability of these two types of tasks. To maximize discriminability, it is recommendable to use visual icons or variables (different colors, for example) so that the differences can be detected automatically using perceptual and visual processes. As we have tested in the third experiment, *findability* and *recognizability* of notifications is higher in the case of CIAN, due to the fact that this visual language provides a specific icon to represent it.

Guideline 6: In order to state the *visual vocabulary* in the proposed notations and improve the *semiotic clarity*, it is highly recommendable to define an explicit *metamodel* expressed, if possible, in standard format (e.g. using MOF [100]).

Guideline 7: Where possible, concepts should be encoded *graphically* to obtain maximum advantage of the power of human visual processing.

We consider that CIAN does not present *symbol deficit*, making it more *expressive* than CTT. However, we have detected a possible deficiency in the case of CIAN that does not occur in the case of CTT. This fault is related to the number of icons proposed by the CIAN notation. As we have contrasted in the described experiments (experiment 3), the majority of the participants had the subjective perception that CIAN is more complex. This perception can be related to the number of different symbols included in its vocabulary, that can be measured as the number of legend entries required (Fig. 1 and Fig. 3). The human ability to discriminate between perceptually distinct alternatives is around six categories [101]. CIAN exceeds this limit. In this sense, the graphic complexity of CIAN is higher than that of CTT. In the future, we are going to investigate a way to reduce the semantic complexity of our notation. Bearing this in mind, we can ask ourselves if, for example, it is really necessary in *process model* to distinguish between the various types of organizational members (role, software agent, etc.). Another point that we have to analyze is if we need to textually express some of the LOTOS temporal operators in the flows of this diagram. For example "$\gg$" operator (enabling or sequence temporal operator) can be omitted because the graphical representation of flow already expresses the sequentiality between tasks. Also "[ ]" (choice) operator is implicitly and visually expressed in the graph of the process model. We have to revise the use of all the temporal operators included in the language whose visual representations make sense in the case of the hierarchical representation proposed by CTT, but can be redundant in the case of the graphs specified using CIAN.

Guideline 8: When creating the *visual vocabulary*, try to propose a manageable number of graphic symbols; that is, limit the *size of the visual vocabulary* of the notation. The number of symbols can affect the ability to discriminate them and the perceived complexity of the notation by users of the notation (mainly for novice users).

Another interesting cognitive dimension is related to the use of ***secondary notations***. This dimension suggests that sometimes we can use cues (for example colors) to clarify, emphasize or complement information. The inclusion of these additional cues (colors, brightness, texture, etc.) is not supported in the current version of CIAN. We plan to consider and assess these features in the future, including *eye tracking* analysis [102].

In relation to the final cognitive dimension (the ***hard mental effort***), we can consider the results related to this issue obtained in the three experiments. The *subjective perception* about *ease of use* can be considered as the degree to which a person believes that using a particular method would be *effort-free*. This variable represents a perceptual judgment of the *effort* required to learn and use a method. In this regard, the results obtained indicate that the majority of the participants found CIAN to be quite easy to use in supporting the collaborative learning systems modeling. The performance of the participants (*time*, *score*, *understandability effectiveness* and *understandability efficiency*) in the experiments is also related to their positive perceptions. We consider that the subjective perception of the users about this aspect is influenced by the support for the cognitive dimensions mentioned. As Gemino et al. [67] suggest, the *usefulness* of a modeling method is related to its ability to represent, communicate and understand the domain. Hence, we consider that all these features are directly related to the dimensions provided by the CD framework.

Finally, it is interesting to comment on the ability of both notations for ***managing complexity***. In this sense, CTT and CIAN provide *hierarchy* and *modularization* mechanisms, respectively. However, in the case of CTT, all the levels of abstraction appear in a monolithic diagram, making it more difficult to manage. We believe that CIAN proposes a more effective complexity management mechanism, allowing the creation of separated models to specify more detailed levels of abstraction. However, additional empirical evaluations are necessary to contrast this opinion.

To conclude, we want to mention some ***methodological issues*** of the family of experiments performed:

- ***Sample nature***. In order to guarantee the external validation of empirical studies, it is recommendable to recruit representative participants of the final population that will use our specification method. Bearing this in mind, we have used "convenience samples" (undergraduate students) in the first and third studies. This fact threatens the validity of statistical conclusions and external validity. However, if we consider that the subjects had enough maturity to participate and the task to be performed did not require a high level of industrial experience then, according to [81,80], these experimental subjects can be considered appropriate.

- ***Laboratory vs. real case scenarios***. Although real world data provides higher *ecological validity*, if we want to perform an *eye tracking* study (as we have realized in the third experiment) it is recommendable to perform *controlled experiments*. In this context, we have more control over the factors affecting the experiment (environmental lighting conditions, possible interruptions, etc.).
- ***Pilot test***. When doing experiments in general, and when we want to perform an *eye tracking* study in particular, it is necessary and highly recommendable to run a *pilot test* to verify the experimental conditions.
- ***Measurement instrument***. In second and third experiments, we have used a method and measurement instrument contrasted and based on theoretical foundations [13]. With regard to *eye tracking* metrics, we have considered the interpretations provided by most of the expert authors in this field [87].
- ***Subjective vs. objective results***. We consider as a main contribution of this work the fact that it combines not only subjective opinion of participants, as the majority of existing studies do for assessing notations, but also objective (physical) sources of information. The use of subjective surveys (perception-based) has the inconvenience that answers may be biased. Using *eye tracking* techniques, we obtain objective evidence that allows us to perform a more complete analysis and contrast results. However, as we commented in Section 5.3.2 sometimes a same *eye tracking* metric can be interpreted in different ways. Hence, it is important to attempt to supplement *eye tracking* data with additional information gained from the participants about their experiences during the *eye tracking* session. Therefore, the mixing of both sources of information (*eye tracking* metrics and perception-based questionnaires) is necessary in this kind of experiments.

6. Concluding remarks and future works

The use of graphical notations can significantly facilitate the specification and design of learning systems. In particular, our interest is centered on the design of CSCL systems and the modeling of group learning activities for which the current standards and notations present some limitations. We have studied the existing proposals in this field and identified the main features to be considered when specifying collaborative tasks. We propose the use of the CIAN notation for modeling this type of learning activity.

One important aspect to take into consideration when a new modeling notation is proposed is to assess its quality. In this paper, we describe three empirical studies that attempt to measure the adequacy of that notation to model collaboration, contrasting its use with another notation which allows modeling these aspects (the CTT notation). In these experiments, we use several sources of information: subjective perception of the designers, their profiles, their performance on a set of understandability activities of models as well as the physical evidence provided by an *eye tracker* device. The use of *eye tracking* techniques provides objective evidence (complementary to other sources of information) that allows us to perform a more complete analysis and comparison of the two notations.

The results provided in these experimental studies denote positive perceptions about the use of the CIAN notation for modeling collaborative learning activities. We are aware that there are many other factors which could affect people's decisions when using requirement modeling methods (e.g., tool infrastructure adoption and standardization). We propose to analyze these additional factors in the future. As future work, we also plan to replicate the last experiment and to increase the sample size as well as analyzing the relationship and consistency between the various sources of information used in it. For this purpose we plan to apply not only statistical methods, but also more complete and rich analysis methods such as Bayesian networks.

Acknowledgements

This work has been partially supported by the INTEGroup, GITE-LEARN and iColab projects, funded by the Regional Government of Castilla-La Mancha, and the EDUCA-Prog project funded by the Ministry of Science and Innovation. The authors want to thank all the participants in the studies performed for their collaboration.

References

[1] L.S. Botturi, S.T. Todd Stubbs, Handbook of Visual Languages for Instructional Design: Theories and Practices, Idea Group (IGI), 2008.
[2] A. Rawlings, P. van Rosmalen, R. Koper, M. Artacho-Rodriguez, P. Lefrere, Survey of educational modelling languages (EML), in: CEN/ISSS WS/LT Learning Technologies Workshop, September 19, 2002.
[3] I. Martínez-Ortiz, P. Moreno-Ger, J.L. Sierra, B. Fernández-Manjón, A flow-oriented visual language for learning designs, in: 7th International Conference on Web-based Learning (ICWL), Jinhua, China, in: LNCS, vol. 5145, 2008.
[4] M. Caeiro Rodríguez, M.J. Marcelino, M. Llamas-Nistal, L.E. Anido Rifón, A.J. Mendes, Supporting the modeling of flexible educational units PoEML: a separation of concerns approach, J. Univers. Comput. Sci. 13 (7) (2007).
[5] L. Botturi, E2ML: A visual language for the design of instruction, Educ. Technol. Res. Dev. 54 (3) (2006).
[6] T. Koschmann, CSCL: Theory and Practice of an Emerging Paradigm, 1996.
[7] A.I. Molina, F. Jurado, W.J. Giraldo, M.A. Redondo, M. Ortega, Specifying scripts and collaborative tasks in CSCL environment using IMS-LD and CIAN, in: IEEE International Conference on Advanced Learning Technologies (ICALT), 2008.
[8] X. Lacaze, P. Palanque, Comprehensive handling of temporal issues in tasks models: what is needed and how to support it?, in: Workshop the Temporal Aspects of Work for HCI (CHI 2004), Vienna, Austria, 2004.

[9] C.A. Ellis, S.J. Gibbs, G.L. Rein, Groupware: Some issues and experiences, Commun. ACM 34 (1) (1991).
[10] A.I. Molina, M.A. Redondo, M. Ortega, A methodological approach for user interface development of collaborative applications: A case study, Sci. Comput. Program. 74 (2009).
[11] M. Paredes, A.I. Molina, M.A. Redondo, M. Ortega, Developing collaborative and ubiquitous applications using CIAM, J. Univers. Comput. Sci. 14 (16) (2008).
[12] W.J. Giraldo, A.I. Molina, C.A. Collazos, M. Ortega, M.A. Redondo, CIAT, A model-based tool for designing groupware user interfaces using CIAM, in: Computer Aided-Design of User Interfaces (CADUI), Springer-Verlag, 2008.
[13] S. Abrahao, E. Insfran, J.A. Carsí, M. Genero, Evaluating requirements modeling methods based on user perceptions: A familiy of experiments, Inf. Sci. 181 (2011).
[14] J. Nielsen, K. Pernice, Técnicas de Eye Tracking para usabilidad Web, ANAYA Multimedia, New Riders, 2010.
[15] T.R. Green, M. Petre, Usability analysis of visual programming environments: a “cognitive dimensions” framework, J. Vis. Lang. Comput. 7 (1996) 131–174.
[16] F. Paternò, ConcurTaskTrees: An engineered notation for task models, in: D. Diaper, N.A. Stanton (Eds.), The Handbook of Task Analysis for HCI, LEA, Mahwah, NJ, 2004.
[17] G. Mori, F. Paternò, C. Santoro, CTTE: Support for developing and analyzing task models for interactive system design, IEEE Trans. Softw. Eng. 28 (9) (2002).
[18] R. Koper, J. Manderveld, Educational modelling language: modelling reusable, interoperable, rich and personalised units of learning, Br. J. Educ. Technol. 35 (5) (2004) 537–551.
[19] R. Koper, Modeling units of study from a pedagogical perspective: the pedagogical meta-model behind EML, Technical Report, Open Univ. of the Netherlands, 2001.
[20] F. De Vries, C. Tattersall, R. Koper, Future developments of IMS learning design tooling, Educ. Technol. Soc. 9 (1) (2006) 9–12.
[21] D. Sampson, P. Karampiperis, P. Zervas, ASKLDT: A web-based learning scenarios authoring environment based on IMS learning design, Adv. Technol. Learn. 2 (4) (2005) 207–215.
[22] J.R. Dalziel, Implementing learning design: the learning activity management system (LAMS), in: Proc. of the 20th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education, 2003, pp. 593–596.
[23] G. Paquette, I. Teja, M. Leonard, K. Lundgren-Cayrol, O. Marino, An instructional engineering method and design tool, in: R. Koper, C. Tattersall (Eds.), Learning Design: A Handbook on Modelling and Delivering Networked, Education and Training, Springer-Verlag, 2005, pp. 161–184.
[24] M. Derntl, S. Neumann, P. Oberhuemer, Community support for authoring, sharing, and reusing instructional models: the open graphical learning modeler (OpenGLM), in: Proceedings of IEEE ICALT 2011, Athens, Georgia, USA, IEEE, 2011.
[25] I. Martínez-Ortiz, J.L. Sierra, B. Fernández-Manjón, Translating e-learning flow-oriented activity sequencing descriptions into rule-based designs, in: 6th International Conference on Information Technology: New Generations ITNG 2009, April 27–29, Las Vegas, Nevada, USA, 2009.
[26] M. Derntl, Patterns for person-centered e-learning, PhD thesis, Faculty of Computer Science, University of Vienna, Austria, 2005.
[27] M. Derntl, Patterns for Person-Centered e-Learning, Dissertations in Database and Information Systems – Infix, vol. 96, IOS Press, Amsterdam, 2007.
[28] M. Caeiro-Rodriguez, L. Andino-Rifon, M. Llamas-Nistal, POEML: A perspective-oriented educational modeling language meta-model for engineering e-learning practices, in: 15th International World Wide Web Conference (WWW2006), May 23–26, Edinburgh, Scotland, 2006.
[29] IMS Global, IMS learning design specification, http://www.imsglobal.org/learningdesign/, last visited on 1 November 2013.
[30] R. Koper, C. Tattersall (Eds.), Learning Design – A Handbook on Modelling and Delivering Networked Education and Training, Springer, Heidelberg, 2005.
[31] Business Process Management Initiative, Business Process Modeling Notation Specification, Available online at: http://www.bpmn.org/, last visited on 1 November 2013.
[32] P. Karampiperis, D. Sampson, Towards a common graphical language for learning flows: transforming BPEL to IMS learning design level A representations, in: Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), 2007.
[33] Business process execution language, http://www.oracle.com/, last visited on 1 November 2013.
[34] G. Booch, J. Rumbaugh, I. Jacobson, The Unified Modeling Language User Guide, Addison–Wesley, Reading, MA, 1999.
[35] F. Jurado, M.A. Redondo, M. Ortega, Specifying collaborative tasks of a CSCL environment with IMS-LD, in: The Third International Conference on Cooperative Design, Visualization and Engineering (CDVE2006), Mallorca, 2006, in: LNCS, Springer, 2006.
[36] P. Laforcade, T. Nodenot, C. Sallaberry, Un langage de modélisation pédagogique basé sur UML, in: Numéro spécial: Conceptions et usages des plates-formes de formation, Rev. STICEF Sci. Technol. Inf. Commun. Éduc. Form. 12 (2005).
[37] P. Laforcade, Méta-modélisation UML pour la conception et la mise en oeuvre de situations-problèmes coopératives, Ph.D. thesis, Université de Pau et des Pays de l'Adour, France, 2004.
[38] M. van Welie, G.C. van der Veer, Groupware task analysis, in: E. Hollnagel (Ed.), Handbook of Cognitive Task Design, LEA, NJ, 2003, pp. 447–476.
[39] D. Pinelle, Improving groupware design for loosely coupled groups, Ph.D. thesis, Department of Computer Science, University of Saskatchewan, 2004.
[40] Y.K. Lim, Multiple aspect based task analysis (MABTA) for user requirements gathering in highly-contextualized interactive system design, in: 3rd Annual Conference on Task Models and Diagrams (TAMODIA 2004), Prague, Czech Republic, in: ACM International Conference Proceeding Series, 2004.
[41] T. Bolognesi, H. Brinskma, Introduction to the ISO specification language LOTOS, Comput. Netw. ISDN Syst. 14 (1987) 25–59.
[42] S. Carlsen, Action port model: a mixed paradigm conceptual workflow modeling language, in: Proceedings of the 3rd IFCIS International Conference on Cooperative Information Systems, IEEE Computer Society, 1998.
[43] H. Trætteberg, Model-based user interface design, Ph.D. thesis, Dept. of Computer and Information Sciences, Norwegian University of Science and Technology, 2002.
[44] W.M.P. van der Aalst, P. Berthelmess, C.A. Ellis, J. Wainer, Proclets: a framework for lightweight interacting workflow processes, Int. J. Coop. Inf. Syst. 10 (4) (2001) 443–482.
[45] J.L. Garrido, M. Gea, M.L. Rodríguez, Requirements Engineering in Cooperative Systems. Requirements Engineering for Sociotechnical Systems, Idea Group Inc., USA, 2005, pp. 226–244.
[46] J. Rubart, P. Dawabi, Shared data modeling with UML-G, Int. J. Comput. Appl. Technol. 19 (2004).
[47] A.I. Molina, M.A. Redondo, M. Ortega, A review of notations for conceptual modeling of groupware systems, in: J.A. Macías, T. Granollers, P.M. Latorre (Eds.), New Trends on Human–Computer Interaction: Research, Development, New Tools and Methods, Springer-Verlag, 2009, pp. 75–86.
[48] C. Martel, L. Vignollet, Using the Learning Design Language to model activities supported by services, Int. J. Learn. Technol. 3 (4) (2008) 368–387.
[49] D. Mota, C. Vaz de Carvalho, L.P. Reis, A conceptual model for collaborative learning activities design, in: 2011 IEEE Global Engineering Education Conference (EDUCON)—Learning Environments and Ecosystems in Engineering Education, 2011.
[50] L. Kobbe, A. Weinberger, P. Dillenbourg, A. Harrer, R. Hämäläinen, P. Häkkinen, F. Fischer, Specifying computer-supported collaboration scripts, Comput.-Support. Collab. Learn. 2 (2007) 211–224.
[51] F. Fischer, I. Kollar, H. Mandl, J. Haake (Eds.), Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives, Springer, New York, 2007.

[52] C.A. Collazos, L.A. Guerrero, J.A. Pino, S.F. Ochoa, Collaborative scenarios to promote positive interdependence among group members, in: J. Favela, D. Decouchant (Eds.), CRIWG 2003, in: LNCS, vol. 2806, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 356–370.
[53] P. Dillenbourg, M. Baker, A. Blaye, C. O'Malley, The Evolution of Research on Collaborative Learning, London, 1995.
[54] M.A. Redondo, C. Bravo, M. Ortega, M.F. Verdejo, Providing adaptation and guidance for design learning by problem solving. The DomoSim-TPC approach, Comput. Educ. 48 (2007).
[55] C. Bravo, M.A. Redondo, M. Ortega, M.F. Verdejo, Collaborative distributed environments for learning design tasks by means of modeling and simulation, J. Netw. Comput. Appl. 29 (2006) 321–342.
[56] D.L. Moody, Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions, Data Knowl. Eng. 55 (3) (2005).
[57] J. Nelson, G. Poels, M. Genero, M. Piattini, Quality in conceptual modeling – five examples of the state of art, Data Knowl. Eng. 55 (3) (2005).
[58] N. Fenton, S. Pfleeger, Software Metrics: A Rigorous Approach, second ed., Chapman & Hall, London, 1997.
[59] R.E. Mayer, Models for understanding, Rev. Educ. Res. 59 (1) (1989).
[60] D.S. Kolovos, R.F. Paige, T. Kelly, F.A.C. Polack, Requirements for domain-specific languages, in: First ECOOP Workshop on Domain-Specific Program Development (ECOOP), 2006.
[61] ISO/IEC 9126-1, Software Engineering—Software Product Quality—Part 1: Quality Model, International Organization for Standardization, Geneva, Switzerland, 2001.
[62] H.J. Nelson, G. Poels, M. Genero, M. Piattini, A conceptual modeling quality framework, Softw. Qual. J. 20 (2012).
[63] M. Genero, M. Piattini, C. Calero (Eds.), Metrics for Software Conceptual Models, Imperial College Press, United Kingdom, 2005.
[64] G. Karsai, H. Krahn, C. Pinkernell, B. Rumpe, M. Schindler, S. Volkel, Design guidelines for domain specific languages, in: The 9th OOPSLA Workshop on Domain-Specific Modeling, 2009.
[65] D.L. Moody, The "Physics" of notations: towards a scientific basis for constructing visual notations in software engineering, IEEE Trans. Softw. Eng. 35 (6) (2009).
[66] S. Kelly, R. Pohjonen, Worst practices for domain-specific modeling, IEEE Softw. 26 (4) (2009).
[67] A. Gemino, Y. Wand, Complexity and clarity in conceptual modeling: comparison of mandatory and optional properties, Data Knowl. Eng. 55 (2005).
[68] A. Maes, G. Poels, Evaluating quality of conceptual modelling scripts based on user perceptions, Data Knowl. Eng. 63 (2007).
[69] S. Yusuf, H. Kagdi, J.I. Maletic, Assessing the comprehension of UML class diagrams via eye tracking, in: 15th IEEE International Conference on Program Comprehension (ICPC), 2007.
[70] B. Sharif, H. Kagdi, On the use of eye tracking in software traceability, in: Proceedings of the 6th International Workshop on Traceability in Emerging Forms of Software Engineering (TEFSE 2011), 2011.
[71] B. Sharif, J.I. Maletic, An eye tracking study on the effects of layout in understanding the role of design patterns, in: 26th IEEE International Conference on Software Maintenance in Timisoara, Romania, 2010.
[72] T.R.G. Green, A.F. Blackwell, Cognitive Dimensions of Information Artefacts: A Tutorial. Version 1.2, October 1998.
[73] A.F. Blackwell, T.R.G. Green, Notational systems: the cognitive dimensions of notations framework, in: J.M. Carroll (Ed.), HCI Models Theories and Frameworks, Morgan Kaufmann, San Francisco, 2003, pp. 103–134.
[74] K. Sousa, J. Vanderdonckt, B. Henderson-Sellers, C. Gonzalez-Perez, Evaluating a graphical notation for modelling software development methodologies, J. Vis. Lang. Comput. 23 (2012) 195–212.
[75] J. Gallardo, C. Bravo, M.A. Redondo, J. De Lara, Modeling collaboration protocols for collaborative modeling tools: Experiences and applications, J. Vis. Lang. Comput. 24 (2013) 10–23.
[76] E. Guerra, J. De Lara, A. Malizia, P. Díaz, Supporting user-oriented analysis for multi-view domain-specific visual languages, Inf. Softw. Technol. 51 (2009) 769–784.
[77] B. Mora, F. García, F. Ruiz, M. Piattini, Graphical versus textual software measurement modelling: an empirical study, Softw. Qual. J. 19 (2011) 201–233.
[78] T.R.G. Green, M. Petre, R.K.E. Bellamy, Comprehensibility of visual and textual programs: A test of superlativism against the match–mismatch conjecture, in: 4th Workshop on Empirical Studies of Programmers, 1991, pp. 121–146.
[79] A. Kitchenham, S. Pfleeger, D.C. Hoaglin, K. El Emam, J. Rosenberg, Preliminary guidelines for empirical research in software engineering, IEEE Trans. Softw. Eng. 28 (8) (2002) 721–734.
[80] M. Höst, B. Regnell, C. Wholin, Using students as subjects – a comparative study of students and professionals in lead-time impact assessment, in: Proceedings of the 4th Conference on Empirical Assessment and Evaluation in Software Engineering, EASE'00, Keele, UK, 2000, pp. 201–214.
[81] V. Basili, F. Shull, F. Lanubile, Building knowledge through families of experiments, IEEE Trans. Softw. Eng. 25 (4) (1999) 435–437.
[82] B.H.C. Cheng, J.M. Atlee, Research directions in requirements engineering, in: L.C. Briand, A.L. Wolf (Eds.), Workshop on the Future of Software Engineering at ISCE 2007, Minneapolis, USA, 2007, pp. 285–303.
[83] M. Ciolkowski, F. Shull, S. Biffl, A family of experiments to investigate the influence of context on the effect of inspection techniques, in: Proceedings of the 6th International Conference on Empirical Assessment in Software Engineering, EASE'02, Keele, UK, 2002, pp. 48–60.
[84] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, Experimentation in Software Engineering: An Introduction, Kluwer Academic Publishers, 2000.
[85] V. Basili, H.D. Rombach, The TAME project: towards improvement-oriented software environments, IEEE Trans. Softw. Eng. 14 (6) (1988) 758–773.
[86] H. Larkin, H.A. Simon, Why a diagram is (sometimes) worth ten thousand words, Cogn. Sci. 11 (1) (1987) 65–100.
[87] A. Poole, J.B. Linden, Eye Tracking in Human–Computer Interaction and Usability Research: Current Status and Future Prospects, 2004.
[88] J. Nielsen, K. Pernice, Eyetracking Methodology. How to Conduct and Evaluate Usability Studies Using Eyetracking, Nielsen Norman Group, 2009.
[89] A. Rösler, Using the Tobii mobile device stand in usability testing on mobile devices, Whitepaper, 2012.
[90] Z. Guan, S. Lee, E. Cuddihy, J. Ramey, The validity of the stimulated retrospective think-aloud method as measured by eye tracking, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006), April 22–27, 2006, Montréal, Québec, Canada, 2006.
[91] Tobii, Retrospective think aloud and eye tracking. Comparing the value of different cues when using the retrospective think aloud method in web usability testing, Whitepaper by Tobii Technology, 2009.
[92] A. Hyrskykari, S. Ovaska, P. Majaranta, K.J. Räihä, M. Lehtinen, Gaze path stimulation in retrospective think aloud, J. Eye Mov. Res. 2 (4) (2008) 1–18.
[93] M.J. Van den Haak, M.D.T. de Jong, J. Schellens, Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue, Behav. Inf. Technol. 22 (5) (2003) 339–351.
[94] A. Paivio, Mental Representations: A Dual Coding Approach, Oxford Univ. Press, 1986.
[95] L. Lin, R.K. Atkinson, Using animations and visual cueing to support learning of scientific concepts and processes, Comput. Educ. 56 (2011) 650–658.
[96] Tobii, Accuracy and precision, Test report, Tobii T60 XL Eye tracker, http://www.tobii.com/, 2011.
[97] K. Masri, D. Parker, A. Gemino, Using iconic graphics in entity relationship diagrams: The impact on understanding, J. Database Manag. 19 (3) (2008) 22–41.
[98] M. Wertheimer, Laws of organization in perceptual forms, in: W. Ellis (Ed.), A Source Book of Gestalt Psychology, 1938.
[99] W.D. Winn, An account of how readers search for information in diagrams, Contemp. Educ. Psychol. 33 (3) (1993) 162–185.
[100] OMG, MOF Core Specification, Object Management Group, 2006.
[101] G.A. Miller, The magical number seven, plus or minus two: some limists on our capacity for processing information, Psychol. Rev. 63 (1956) 81–97.
[102] E. Ozcelik, T. Karakus, E. Kursun, K. Cagiltay, An eye-tracking study of how color coding affects multimedia learning, Comput. Educ. 53 (2009) 445–453.