QRE
Qualitative Research in Education

Hipatia Press
www.hipatiapress.com

Instructions for authors, subscriptions and further details:

http://qre.hipatiapress.com

An Orchestrating Evaluation of Complex Educational Technologies: a Case Study of a CSCL System

Luis P. Prieto[1]; Yannis Dimitriadis[2] & Juan I. Asensio-Pérez[2]

1) CHILI Lab, École Polytechnique Fédérale de Lausanne, Switzerland
2) Department of Signal Theory, Universidad de Valladolid, Spain

Date of publication: June 28th, 2014
Edition period: June 2014 - October 2014

To cite this article: Prieto, L. P. Dimitriadis, Y., & Asensio-Pérez, J. I. (2014). An Orchestrating Evaluation of Complex Educational Technologies: a Case Study of a CSCL System. *Qualitative Research in Education, 3*(2), 175-205. doi: 10.4771/qre.2014.44

To link this article: http://dx.doi.org/10.4471/qre.2014.44

Qualitative Research in Education Vol.3 No.2 June 2014 pp. 175-205

Orchestrating Evaluation of Complex Educational Technologies: a Case Study of a CSCL System

Luis P. Prieto
École Polytechnique Fédérale de Lausanne

Yannis Dimitriadis
Universidad de Valladolid

Juan I. Asensio -Pérez
Universidad de Valladolid

(Received: 23 February 2014; Accepted: 3 June 2014; Published; 28 June 2014)

Abstract

As digital technologies permeate every aspect of our lives, the complexity of the educational settings, and of the technological support we use within them, unceasingly rises. This increased complexity, along with the need for educational practitioners to apply such technologies within multi-constraint authentic settings, has given rise to the notion of technology-enhanced learning practice as "orchestration of learning". However, at the same time, the complexity involved in evaluating the benefits of such educational technologies has also increased, prompting questions about the way evaluators can cope with the different places, technologies, informants and issues involved in their evaluation activity. By proposing the notion of "orchestrating evaluation", this paper tries to reconcile the often disparate "front office accounts" of research publications and the "shop floor practice" of evaluation of educational technology, through the case study of evaluating a system to help teachers in coordinating computer-supported collaborative learning (CSCL) scenarios. We reuse an internationally-evaluated conceptual framework of "orchestration aspects" (design, management, adaptation, pragmatism, etc.) to structure the case's narrative, showing how the original evaluation questions and methods were modulated in the face of the multiple (authentic) evaluation setting constraints.

Keywords: evaluation, mixed methods, hybrid methodologies, educational technology, orchestration.

ISSN: 2014-6418
DOI: 10.4471/qre.2014.44

As in every other aspect of our lives, information and communication technologies (ICT) are slowly permeating educational practice. Our classrooms (no longer restricted to a physical space and face-to-face, synchronous interaction) are becoming messy, complex socio-technical ecosystems of resources (Luckin, 2008).

This increased complexity of technology-enhanced learning innovations, and the difficulties of implementing them while complying with the multiple constraints of authentic formal educational practice (curriculum, time available, etc.) have lately come into the foreground of attention in educational research, through the notion of "orchestrating learning" (Dillenbourg, Järvelä, & Fischer, 2009; Prieto, Holenko-Dlab, Abdulwahed, Gutiérrez, & Balid, 2011; Sutherland & Joubert, 2009). Although the international research community interested in this topic does not unanimously agree on its exact nature or its definition (see, for example, the special section of Computers & Education, 69, 2013 for a recent compilation of contrasting perspectives on the subject of orchestrating learning), there seems to be a common emphasis on proposing innovations that take into account the multiple restrictions of authentic educational settings, as opposed to, e.g., experiments in controlled conditions (Roschelle, Dimitriadis, & Hoppe, 2013).

In parallel with authentic educational settings' growing technological complexity, the research evaluation of such technological innovations is also becoming more intricate (Jorrín-Abellán, Stake, & Martínez-Monés, 2009; Treleaven, 2004). These evaluations made by researchers or teacher-researchers (e.g., involved in action-research) have to consider pedagogical and technological issues, the effects and interactions of multiple technological and legacy learning tools, and the point of view of multiple actors and informants (e.g., teachers, students, parents, other staff). Moreover, since learning itself may happen in different times and physical contexts (in the classroom, at home, in a field trip, on the way home), very often evaluation of the learning technologies has to follow the learning process across these contexts as well.

However, most evaluations of technological innovations for learning, including those that occur in authentic settings, still follow the same evaluation approaches and ways of presenting research that we used when that complexity was absent. If practice of evaluation is becoming increasingly complex, but such complexity is not reflected in how research

is presented, we might be facing a "shop floor problem" (Garfinkel, 2002), in which real field practice (what evaluators do as they go about their research) and the "front office accounts" of such practice (how such research is reported, e.g., in articles or in project reports) are increasingly disconnected.

In this paper we propose the notion of "orchestrating evaluation", a transposition of the concept of "orchestrating learning" explained above to the practice of evaluating educational technologies. Thus, in this context, orchestrating evaluation can be defined as the *coordination of the* (increasingly complex) *practice of evaluating learning technologies, within the multiple constraints of authentic educational settings*. In order to explore this notion, we apply a conceptual framework on orchestrating learning (proposed and evaluated at the level of the international research community on technology-enhanced learning, see (Prieto, Holenko-Dlab, et al., 2011; Prieto, 2012)) to organize the "shop floor account" of the evaluation of one concrete educational technology. This technology (GLUE!-PS) is a system to help teachers coordinate computer-supported collaborative learning (CSCL) situations (Prieto et al., 2013, 2014). We hope that this kind of account helps future evaluation practitioners (e.g., researchers, teacher-researchers) in articulating their evaluation practice (especially for those less experienced researcher-evaluators - as opposed to external/specialist evaluators), sparking up the debate of what evaluation practices are methodologically sound, but at the same time feasible within today's authentic educational settings.

The structure of the paper is as follows: in the next section, we briefly review basic notions of evaluation, especially in the field of educational technologies; then, the notion of "orchestrating learning" is explained, along with an existing conceptual framework to understand orchestration, and how it could be transposed to the practice of evaluation. Afterwards, we describe the context and methodology of the case study that will illustrate such transposition. The results of analyzing the evaluation of GLUE!-PS from an orchestration perspective are detailed in the following section. Finally, a brief discussion is included and conclusions are drawn for further research along this line of work.

Evaluation of Educational Technology

The field of evaluation in educational research has a long and rich history. (Oliver, 2000) defines evaluation as "the process by which people make value judgments about things". (R. E. Stake, 2004) rather sees it as improving understanding of the quality of what we want to evaluate in its particular setting. Along its history, several "paradigms" (quantitative, qualitative, pragmatic…, see (Oliver, 2000)) and "generations" (measurement, description, judgment, response – see (Guba & Lincoln, 1989)) have been proposed, and are still hotly debated within the evaluation community, with no unanimous answers to how evaluation should be done.

In this paper, we focus on the evaluation of learning technologies, in the context of educational innovation projects, carried out by researchers or teachers acting as such (like in action-research). In this narrower context, evaluation judgments "concern the educational value of innovations, or the pragmatics of introducing novel teaching techniques and resources" (Oliver, 2000). As in the wider field of evaluation in general, in learning technologies this paradigms' and generations' debate remains unresolved, and some authors conclude that there is no "silver bullet" in evaluation (Oliver, 2000). This has led to a proliferation of methods and frameworks for evaluating learning technologies (examples of this proliferation can be seen in the Journal of Educational Technology & Society, 3(4) and 5(3)), ranging from more traditional professional/external evaluation to education practitioners' action-research (Schön, 1983).

One important issue identified by researchers on learning technologies evaluation is that of authenticity, that is, the notion of how closely an evaluation captures the context of an existing course (Oliver & Conole, 2004). This issue is closely related with the well-known problems of conducting evaluation through controlled experiments (Draper, 1997). Although the issue of authenticity is not at all new, there has been a recent revival of the interest in it from different voices within the technology-enhanced learning research community (McKenney, 2013; Roschelle et al., 2013). This increased interest in proposing technological innovations that address authentic educational settings will undoubtedly lead to a greater need of evaluations that occur in authentic contexts - our focus in this paper.

When introducing this article, we have argued that the increasing (technological as well as pedagogical) complexity of current educational settings necessarily implies a more complex evaluation process. In order to illustrate this, and to frame our evaluation case study later on, let us look at one concrete field within educational technologies: computer-supported collaborative learning (CSCL).

(Stahl, Koschmann, & Suthers, 2006) define CSCL as the branch of research that studies "how people can learn together with the help of computers". Indeed, these authors already anticipate that "the interplay of learning with technology turns out to be quite intricate". In evaluating CSCL, the social component of collaboration adds new difficulties to those typical of learning technologies evaluation (Treleaven, 2004).

As in the general learning technologies field, in CSCL we can also find a proliferation of approaches and frameworks to evaluation: (Economides, 2005; Ewing & Miller, 2002; Jorrín-Abellán et al., 2009; Martínez, Dimitriadis, Rubia, Gómez, & de la Fuente, 2003; Pozzi, Manca, Persico, & Sarti, 2007; Tsiatsos, Andreas, & Pomportsis, 2010; Vatrapu, Suthers, & Medina, 2008). More recently, (Lonchamp, 2012) highlighted the inherent difficulty of analyzing and evaluating CSCL systems, using Rabardel's instrumental theory to explain the different moments that have to be taken into account (preparation phase vs. use phase of the system) when analyzing them. Moreover, certain authors have suggested that recent horizontal trends in computer-supported learning, such as the possibility of having "ubiquitous learning" (Bruce, 2008) may further complicate the evaluation of learning scenarios and technologies across different moments and settings (Jorrín-Abellán & Stake, 2009).

However, for the evaluation practitioner (e.g., a researcher aiming at evaluating a CSCL innovation), most of these approaches and frameworks pose a common problem: they are very often expressed in general, rather abstract terms. Although this is completely understandable (since they are purposefully de-contextualized as they aim to be useful in multiple TEL/CSCL contexts), it nonetheless poses an "abstraction gap" that is not easy to bridge for the unexperienced evaluator. This gap could be compared to the one facing teachers when they have to apply de-contextualized researcher-proposed principles in the concrete context of their own classrooms (Prieto, Villagrá-Sobrino, et al., 2011). Although there exist efforts that try to guide non-expert evaluators with question itineraries,

graphical representations and illustrative examples (e.g., (Jorrín-Abellán et al., 2009)), for most evaluation approaches only a few reported research examples are available. However, similar to (Garfinkel, 2002)'s "front-office accounts", these research reports often center on showing the effectiveness of one innovation/technology for learning, and not in the practice of evaluation itself (in Garfinkel's terms, the "shop floor practice" of evaluation).

In order to help TEL and CSCL researcher-evaluators bridge this "abstraction gap", in the following section we will posit the notion of "orchestrating evaluation". This notion highlights aspects of the evaluation process which often are not described in enough detail in reported research, and which can help evaluators (especially non-experts) understand how the evaluators of learning technologies go about their practice (especially when operating inside the constraints of authentic educational settings).

Practice within the Multiple Constraints of an Authentic Setting: Orchestrating Learning and Orchestrating Evaluation

In an English dictionary, 'orchestrate' is defined as "to arrange or combine so as to achieve a desired or maximum effect". In educational research literature, the word orchestration has been frequently used as a metaphor for teacher practice (e.g., (Kovalainen, Kumpulainen, & Satu, 2001)), given the fact that teachers often have to arrange different elements to achieve a maximum learning effect. However, in learning technologies research this term has gained special relevance in the past few years (Sutherland & Joubert, 2009).

Particularly in the field of CSCL, (Fischer & Dillenbourg, 2006) defined orchestration as the process of "productively coordinating supportive interventions across multiple learning activities occurring at multiple social levels" (cited in (Dillenbourg et al., 2009)). However, as noted by (Prieto, Holenko-Dlab, et al., 2011), there is a disparity of opinions and emphases around this term in the research community. Trying to synthesize these differing points of view, (Roschelle et al., 2013) highlight the common emphasis on paying attention, when proposing learning technology innovations, to the multiple constraints (curriculum, time, discipline,… i.e., not only the learning process) that characterize educational practice in authentic settings. (Dillenbourg, 2013) posits that orchestration can be

brought into attention by looking at the different activities that conform the educational practice with technologies in an authentic classroom, even if they are not directly related to the learning process itself (e.g., the time taken to log into the system that students will use for learning). (Perrotta & Evans, 2013), on the other hand, remind us of the implicit assumptions of these notions of orchestration (teaching as neutral, rational practice towards maximizing learning), and highlight the complex interplay of social pressures and expectations that surround the use of technology in the classroom.

After a literature review on the use of the term 'orchestration' in the field of technology-enhanced learning, (Prieto, Holenko-Dlab, et al., 2011) propose eight different aspects that make up the complex notion of orchestration. Five of these aspects are descriptive of the orchestration process itself: *Design* (the preparation, planning of the learning activities), *Management* (including multiple aspects of the coordination during the activities: time management, group management, maintaining discipline, etc.), *Awareness* (the perceptual processes involved in the coordination, assessment of the learning progress, etc.), *Adaptation* (planned or unplanned modifications to the learning activities, to address unexpected events or learning opportunities), and the respective *Roles of the actors* involved in this process (who performs the aforementioned processes: the teacher, a researcher team, technical staff, students themselves, etc.). They also propose three additional aspects that relate with the reasons upon which the coordination is performed: *Theories* (the explicit or implicit models upon which actors construct the coordination), *Pragmatism* (the contextual constraints that define what is possible or mandatory in the authentic setting, e.g., the adherence to a curriculum or the fixed time duration of a session) and *Alignment* (the combination of different contextual features, tools and elements into synergies to achieve an effective learning experience). This framework tries to reflect the points of view of a multi-disciplinary international research community, and has been indeed evaluated by a considerable portion of such international community (see (Prieto, 2012)). This consensus-based validation highlights the completeness of the framework to address (often conflicting) perspectives on the subject, as well as its value for novice researchers, to help them frame and place their research within this field.

In this paper, we posit the notion of "orchestrating evaluation" as *the process of coordinating the practice of evaluating learning technologies, within the multiple constraints of an authentic educational setting*. By similitude with the notion of orchestrating learning, we can think of the abstract term "practice" as standing for *the processes and tools (often used in multiple contexts)* that evaluators use to achieve such evaluation. As in Garfinkel's "shop floor problem", we propose that a detailed account of the multi-constrained, complex process followed (beyond the methodology and results often provided when reporting research) can help in understanding evaluation practice (especially for novice evaluation practitioners). In order to operationalize this "orchestrating evaluation" concept, we "transpose" Prieto et al.'s framework presented above (which tries to characterize the complexity of educational practice in authentic settings) to the activity of evaluating learning technologies in authentic educational settings (a related but different complex practice). We hypothesize that this framework can be especially suited for this purpose, as it was developed in trying to widen researchers' focus of attention on a complex practice while encompassing conflicting schools of thought and perspectives (as often happens in the field of evaluation), and because of its pedagogical value for novice researchers (one of our main target audiences in this paper). In this new context of evaluating educational technologies within the multiple constraints of authentic settings, the framework aspects can be interpreted in the following way:

- **Design**: Encompasses the original planning of the evaluation (evaluation design), including the selection of techniques, informants, etc. This is the aspect that most evaluation frameworks (e.g., the CSCL-EREM described in (Jorrín-Abellán et al., 2009)) focus on.
- **Management**: The multiple activities involved in the evaluation enactment, both explicit in the evaluation design (data gathering events, data analysis, etc.) and implicit/logistical (entering the field, social coordination of informants, setup of physical/virtual infrastructures for evaluation, data conversions/pre-processing, etc.).
- **Awareness**: The ongoing perceptual processes (i.e., monitoring) of the evaluation process, normally aimed at assessing whether the evaluation objectives will be met. This includes meetings of the

evaluation team, journals or reflections during the evaluation process, pre-assessment of the gathered data, etc.

- **Adaptation**: Includes any modifications to the original evaluation design, as evaluators try to meet the evaluation objectives within the setting constraints (as perceived through the awareness mechanisms above). These adaptations can be either due to unexpected occurrences, unacknowledged constraints, failures to get data in the quantity/quality needed, etc.
- **Role of actors**: Covers who is involved in the evaluation, including the evaluator team, who/what is the evaluand (the object of evaluation), who are the main stakeholders, their respective roles, how it affects the labor of evaluation, and how the evaluation will be reported to each of them.
- **Theory**: Describing what are the theories and models that shape the evaluation, at the different levels – from evaluator's ontological stance (positivist, interpretive, pragmatic) to concrete theories of learning and evaluation, evaluation frameworks, etc. that will shape how the evaluation is conducted.
- **Pragmatism**: The myriad of authentic setting constraints that have to be respected during the evaluation (curriculum, time restrictions, available resources), as well as unexpected opportunities that may rise in the authentic context during evaluation (e.g., for gathering further data, etc.).
- **Alignment**: The efforts of evaluators in trying to find new opportunities and avenues of exploration as the different elements above interact with each other (e.g., incorporating unexpected evaluation adaptations as designed features in further research iterations, using unexpected but available actors as new sources of information, using uncovered setting constraints as emerging or future research challenges).

In the following section, we illustrate the application of this framework to analyze one case of evaluation of an educational technology (thus, in a sense, we perform a meta-evaluation): a CSCL system to support teachers in orchestrating CSCL scenarios. Please note that the word "orchestration" is also part of the research goal of the evaluated technology. To avoid confusion, we will refer to "orchestrating learning" (the goal of the

technology evaluated) and "orchestrating evaluation" (the goal of the meta-evaluation performed in this article) throughout the text.

Context (and Methodology): a Technological System for Teachers Doing CSCL

The evaluation that we analyze in this study took place in the context of the GSIC-EMIC research group at the University of Valladolid (Spain). For over a decade, this multi-disciplinary group has been doing research in the field of CSCL (after years of research in the fields of artificial intelligence and cooperative work - CSCW). The group, formed by engineers, computer scientists and pedagogists, has made great emphasis in supporting the labor of teachers that wish to put CSCL scenarios in practice, both through innovative technologies (e.g., (Bote-Lorenzo et al., 2008; Villasclaras-Fernández, Hernández-Leo, Asensio-Pérez, & Dimitriadis, 2013)) and conceptual tools (Gómez-Sánchez et al., 2009; Hernández-Leo, Asensio-Pérez, & Dimitriadis, 2005). Methodologically, the group has employed a variety of approaches, both quantitative and qualitative, with an emphasis in interpretive perspectives (e.g., (Martínez-Monés et al., 2005)) and mixed-method approaches (see, e.g., (Martínez et al., 2003)).

More concretely, the technological innovation whose evaluation we will be studying is a system called GLUE!-PS. This system is mainly composed by a software architecture and an associated data model (first presented in (Prieto, Asensio-Pérez, Dimitriadis, Gómez-Sánchez, & Muñoz-Cristóbal, 2011)), which aim at helping teachers manage CSCL scenarios that use distributed (web) learning environments (DLEs) as their main technological support. DLEs are learning environments composed by a heterogeneous array of web 2.0 tools (blogs, wikis, shared office applications, etc.) and Virtual Learning Environments (VLEs, e.g., Moodle), as coined by (MacNeill & Kraan, 2010).

As reported in (Prieto et al., 2013), this kind of environments is difficult to manage for non-technology experts, and it is not trivial to create a technological support composed of such an heterogeneous array of web applications, that is coherent with the teacher's pedagogical intentions. The GSIC-EMIC research team developed a prototype implementing the GLUE!-PS proposal (available at http://gsic.uva.es/glueps, last visit: January 2014). This prototype currently supports deploying teachers'

activity ideas (expressed in one of three learning design formats), transforming them into multiple different DLEs made up of combinations of the Moodle and MediaWiki learning environments, as well as more than 15 other "Web 2.0" tools. The user interface of GLUE!-PS, as the teacher would see it, is shown in Figure 1. Although initially conceived as an aid for the teacher in the process of preparation of the learning activities' technological support, further features were added in the process of trying the system in authentic CSCL situations (e.g. the ability to perform run-time changes in the DLE according to unexpected events). This led the research team to conceive GLUE!-PS as supporting the teachers' practice in a wider sense, within the constraints of authentic CSCL settings, i.e., as a tool supporting teachers' "orchestration of learning". However, such "orchestration learning" support had to be validated empirically, by its use in real courses, and by a wide variety of teachers from different disciplines. Such validation, and especially its results, are described in (Prieto, 2012; Prieto et al., 2014). In the following section, we rather focus on describing how the *process* of evaluating GLUE!-PS was performed, how we "orchestrated the evaluation".

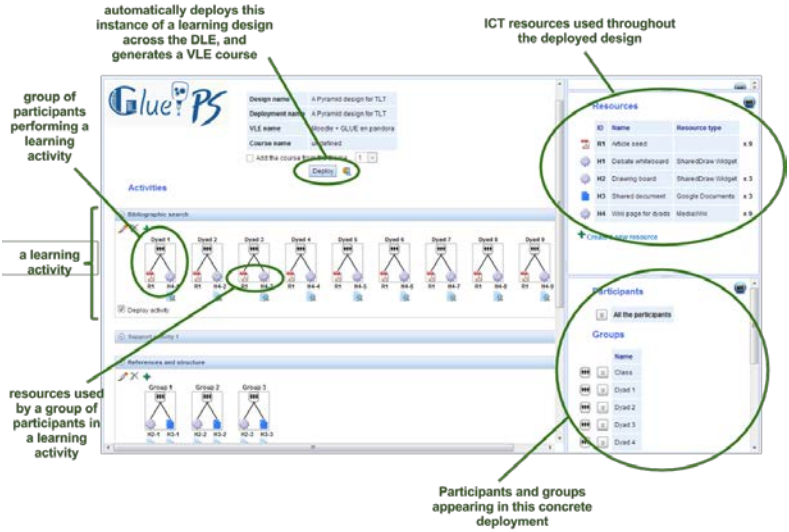


Figure 1. Graphical user interface of the GLUE!-PS prototype. Taken from (Prieto et al., 2014)

For this meta-evaluation study, we have followed (R. E. Stake, 1983)'s responsive approach to evaluation (or meta-evaluation in this case), paying close attention to the activity of evaluating the system, and trying to respond to the information needs of the "people on site", that is, the researcher team that is evaluating the technological system. In this case study, the main research question (and the main meta-evaluative issue used to explore it) has been: 'How did researchers orchestrate the evaluation of GLUE!-PS?'. In order to focus our analysis, we have used an anticipatory data reduction to illuminate this main issue, through eight topics that follow the eight aspects of "orchestrating evaluation" framework presented in the previous section. The data sources used for the study include publications related to the evaluation of GLUE!-PS (including the main proponent of the system's Ph.D. thesis, Prieto, 2012), internal research reports, personal research notebook/notes, team emails and other internal documentation generated during the evaluation.

Orchestrating the Evaluation of GLUE!-PS

As discussed in ((Prieto, 2012) - Chapter 5) and (Prieto et al., 2014), the GLUE!-PS system was evaluated with regard to the orchestration support it provided to teachers in their CSCL practice. This evaluation was done through several studies, in real university courses and in teacher workshops with non-technical teachers from a variety of disciplines. The evidence gathered supports a number of findings, which are summarized graphically in Figure 2:

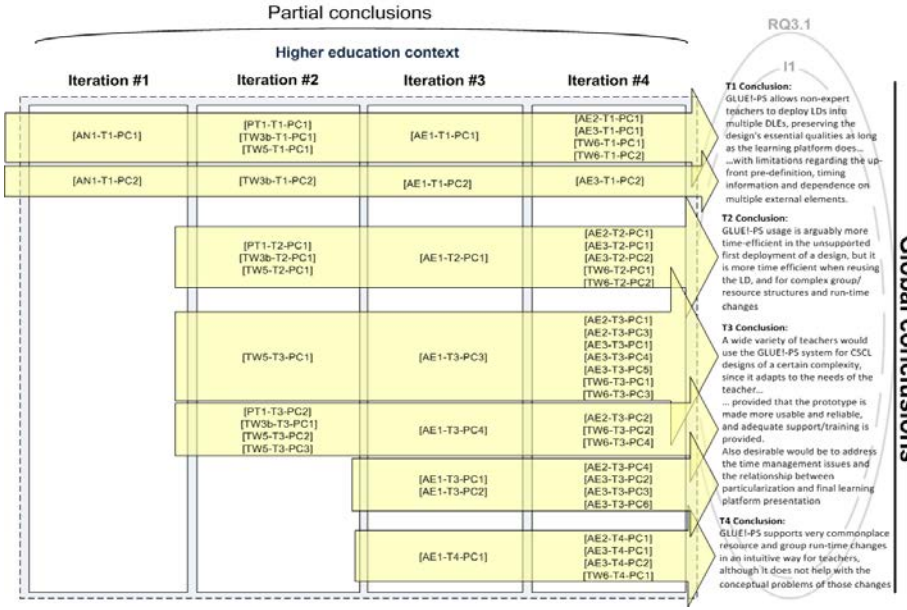


Figure 2. Representation of the results of the evaluation of the orchestration support provided by GLUE!-PS, taken from (Prieto, 2012). The labels between brackets represent partial conclusions extracted from different evaluation happenings (e.g. TW5 for a teacher workshop, AE1 for an authentic course experience, etc.)

These are the *results* of the evaluation of GLUE!-PS. But, how were those evaluation results achieved? What was the *evaluation process* that led to these findings? In the following paragraphs we summarize this meta-evaluation following the "orchestrating evaluation" framework proposed above. The order chosen for the portrayal of each topic (different from the one used in the framework description above) intends to provide a more understandable argument line (as the "orchestrating evaluation" framework does not mandate a concrete order in the analysis of the eight aspects).

Theory

From the point of view of the "paradigm debate" of evaluation, our stance is more in line with a *pragmatic*, post-modern approach that "acknowledges that different underpinnings exist, and adopts each when required" (Oliver, 2000). Within this general worldview, our research team chose an "engineering method" approach (typical in software engineering, see (Glass, 1995; Orlikowski & Baroudi, 1991)) to the research around GLUE!-PS. This method, like many others, contemplates an "evaluative" phase, without prescription of a concrete evaluation method. However, it is important to acknowledge that this kind of methods by definition see the evaluation as an *iterative* endeavor, with our findings and understanding of the learning technology and its impact on the authentic setting being expanded and triangulated with every new evaluation iteration.

Aside from this iterativeness, our evaluation approach was mediated by the CSCL-EREM framework (Jorrín-Abellán et al., 2009), an instrument aimed at helping researchers design their evaluations, following a *responsive* approach to it (see the 'Design' section below for further details). Following the recommendations of this framework, our aforementioned pragmatic stance, and the recommendations of many CSCL researchers (Stahl et al., 2006; Strijbos & Fischer, 2007), mixed methods (Creswell, 2009) were considered the best option for data gathering and analysis within our evaluations. Since the phenomenon of "orchestrating (technology-enhanced) learning" is relatively new and still ill-defined, with little or no clear research constructs/instruments that can be used in a deductive or quantitative way, we considered the evaluation of GLUE!-PS as rather exploratory, thus slanting our methods and techniques more to the qualitative side. Finally, it is interesting to note that the "orchestrating learning" framework by (Prieto, Holenko-Dlab, et al., 2011) was elaborated in parallel by a partially-overlapping researcher team, during the course of this evaluation. This lead to the inclusion of such a framework to operationalize the evaluation rather late within the evaluation process (see the 'Adaptation' section below).

Role of the Actors

As it has been mentioned, the evaluation of GLUE!-PS was performed by a researcher team from the same GSIC-EMIC research group that proposed the system (as opposed to having an external evaluation team). As mentioned earlier, the system proposed and its evaluation were part of a Ph.D. thesis, whose central theme was the support of "orchestration of learning" in CSCL scenarios using DLEs (Prieto, 2012). This implied that the main evaluator was a relatively inexperienced researcher with engineering background, even if supported closely by a core team of two very experienced CSCL researchers (the Ph.D. advisors). The evaluation process was also supported by a varying, multi-disciplinary set of researchers from the same group (up to four researchers, including both Ph.D. students and doctors from pedagogy or engineering), who performed different roles throughout the process, as needed: methodology and engineering consultancy, aiding in data gathering and analysis, etc.

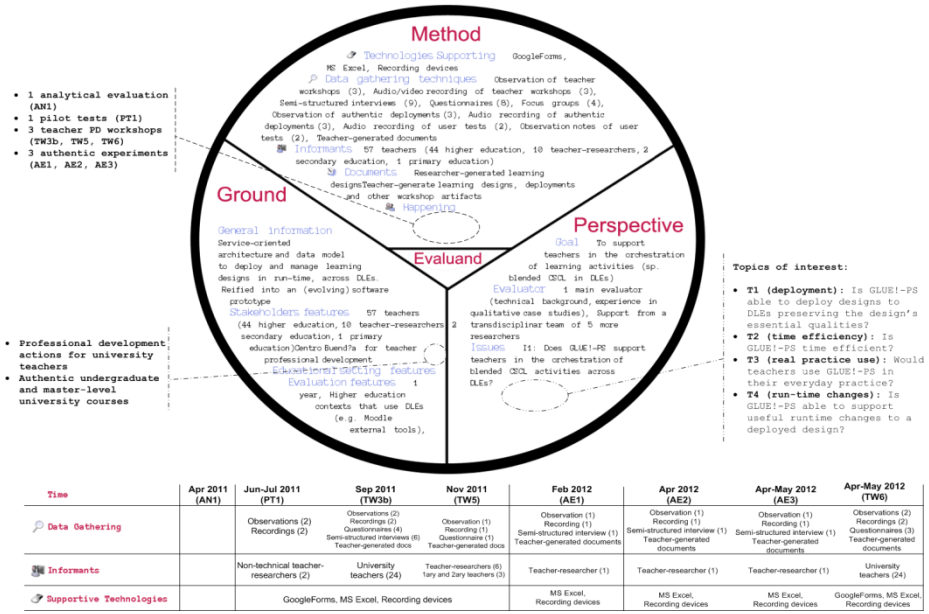
Other important stakeholders in the evaluation process were the informants, most of them university teachers. In this regard, two main groups of teachers can be distinguished: a) teachers who used the GLUE!-PS system to orchestrate CSCL activities in authentic university courses; and b) teachers who used and assessed GLUE!-PS in semi-authentic professional development workshops. The first group of teachers was formed by teacher-researchers (with varying degrees of teaching experience, but who knew about CSCL principles) from the same research group that proposed the system, while the second group was formed by a wider group of university teachers from the same University of Valladolid, with little or no prior knowledge about CSCL. These two sets of informants (especially the first one) can introduce different biases in the data gathered from the evaluation, and cannot be considered (statistically) representative of the teacher population to whom the GLUE!-PS system was aimed. However, the decision of structuring the evaluation around these two groups was taken in trying to find a balance between informants that could afford for *deeper data gathering* (teachers that trusted the innovation enough to dedicate the time needed for learning and using the system in authentic conditions, and to provide extensive data to be gathered by evaluators), and less biased informants with a *wider variety of perspectives*,

backgrounds and attitudes towards ICT and CSCL (but with the common trait of wanting to know more about CSCL).

Finally, although the technological tools used for the different aspects of the evaluation could be considered a non-human actor of the evaluation, in this description we have chosen to mention those within its closest related aspect, for increased clarity.

Design

In order to plan and organize the evaluation, the research team used the CSCL-EREM framework (Jorrín-Abellán et al., 2009). This framework was considered especially adequate for this purpose, as it specifically addresses innovations (technological or otherwise) in the field of CSCL, and it was especially devised with an "inexperienced evaluator" in mind. The framework is structured along different "question paths" (depending on the nature of the 'evaluand', the thing to be evaluated), that help define the evaluation's contextual information (Ground), the goals, important issues and evaluator team (Perspective), as well as the techniques, tools, informants that can help evaluators reach those goals (Method). The framework also provides other aids to the evaluation design, such as graphical representations of the design (see Figure 3) and recommendations about writing the research report. It is interesting to note that such graphical representations and the different question paths have also been implemented technologically through a web application that, e.g., generates automatically CSCL-EREM's graphical representations and research reports (see http://pandora.tel.uva.es/cscl-erem/, last visit: January 2014).



Time	Apr 2011 (AN1)	Jun-Jul 2011 (PT1)	Sep 2011 (TW3b)	Nov 2011 (TW5)	Feb 2012 (AE1)	Apr 2012 (AE2)	Apr-May 2012 (AE3)	Apr-May 2012 (TW6)
Data Gathering		Observations (2) Recordings (2)	Observations (2) Recordings (2) Questionnaires (4) Semi-structured interviews (6) Teacher-generated docs	Observation (1) Recording (1) Questionnaire (1) Teacher-generated docs	Observation (1) Recording (1) Semi-structured interview (1) Teacher-generated documents	Observation (1) Recording (1) Semi-structured interview (1) Teacher-generated documents	Observation (1) Recording (1) Semi-structured interview (1) Teacher-generated documents	Observations (2) Recordings (2) Questionnaires (3) Teacher-generated documents
Informants		Non-technical teacher-researchers (2)	University teachers (24)	Teacher-researchers (6) 1ary and 2ary teachers (3)	Teacher-researcher (1)	Teacher-researcher (1)	Teacher-researcher (1)	University teachers (24)
Supportive Technologies		GoogleForms, MS Excel, Recording devices			MS Excel, Recording devices	MS Excel, Recording devices	MS Excel, Recording devices	GoogleForms, MS Excel, Recording devices

Figure 3. Graphical representation of the GLUE!-PS evaluation design, based on the CSCL-EREM framework. Taken from (Prieto, 2012).

However, this "evaluation design" was not a punctual process that happened only at the beginning of the evaluation process. As we have mentioned, the research around GLUE!-PS, and its evaluation, were done in an iterative fashion. Indeed, the evaluation design as it appears on Figure 3 is only the final state of the evaluation design, after several re-conceptualizations (see 'Adaptation' below). For instance, in this last incarnation of the evaluation design, the conceptual framework for "orchestrating learning" (Prieto, Holenko-Dlab, et al., 2011) was used to operationalize the issues and topics that the evaluation should focus on, within the complex and multifarious notion of "orchestration" (thus complementing well CSCL-EREM's advice, which does not go into the specifics of how to choose the issues and topics to focus a concrete evaluation effort). In this case, the four topics of interest in the right hand side of Figure 3 represent the four aspects of orchestration that GLUE!-PS was designed to support.

Management

The activity of managing the evaluation activity, aside from the general methodological guidelines outlined above (which often appear in the reporting of results, such as (Prieto, 2012; Prieto et al., 2014)), is seldom described in learning technology evaluations. Due to the pragmatic stance of the researcher team (see 'Theory' above), the *multiplicity of 'happenings'* (data gathering events, such as the intervention in a real course or a teacher workshop) and of data gathering techniques within each happening (recordings, interviews, observations, document analyses, … as suggested by the CSCL-EREM framework), were essential features in the evaluation of GLUE!-PS.

This multiplicity required a considerable management effort, which implied the *coordination of data gathering* (e.g., by having one or more preparatory meetings with the 'data gathering team', preparing the necessary infrastructure like recording devices or gathering of logs from involved systems, etc.), the *preparation and running of the events* themselves (preparing the workshop materials for a workshop, ensuring that the ICT infrastructures work as expected, preparation of questionnaires, interview guides and other instruments, etc.), and the coordination of the *data analysis and synthesis process* (e.g., transcription of audio and video sources, meetings among the evaluation team to review available evidence, etc.). It is seldom acknowledged (but it is our experience after these and other evaluation efforts) that this myriad of activities, and the multitude of little *logistic details* that they imply (having every member of the team briefed on the goals of the happening, reviewing and piloting the research instruments beforehand, testing the technologies involved in the happening just before the happening itself, having contingency plans for the failure of the different human and technological elements involved), can have a critical impact on the quality of the data gathered and the findings to be extracted from them. In this sense, having the support of a numerous and varied researcher team proved invaluable.

Indeed, even with this support, the evaluation process entailed a considerable effort, which called for *pragmatic compromises* between the available data and the analyses performed on them (e.g., semi-transcribing the audio for an interview and coding that semi-transcription, instead of doing a full transcription and coding of that data source). The timing of the

different happenings, which was often dictated by extrinsic contextual constraints (see 'Pragmatism' below), and included several overlapping or simultaneous happenings, also contributed to this need of calibrating evaluation efforts. Despite the negative impact that these compromises in data gathering and analysis may have in the studies' credibility, we consider the value of the multiplicity of informants and data gathering techniques (given their potential for triangulation of findings and detection of emergent issues, i.e., for learning about the impact of the learning technology under study) as outweighing the differential added value of a more exhaustive analysis.

Awareness

Following directly from the multiplicity and complexity of the evaluation activities mentioned above, it was crucial during the evaluation to have a clear awareness of how the process was unfolding and whether the evaluation goals were being achieved. Although (Prieto, 2012) portrays the research and evaluation of GLUE!-PS as happening in four clearly-marked iterations (which seem to imply a phase of reflection on the findings and planning for the next iteration), the process was in reality much less linear and compartmentalized, with evaluation happenings following (or overlapping) one another in rapid succession.

In this context, different awareness mechanisms were implemented by the evaluation team, at different levels: a) several "evaluation reports" were produced by the main researcher, detailing (at a certain point in time) the overall evaluation approach and proposed evaluation happenings along with their more detailed design; b) periodic "core researcher team" (normally, the main researcher and his two advisors) meetings in which the goals of the research and the needed evaluation strategy were reviewed; c) for each happening, "extended researcher team" meetings (including the core researcher team plus other members involved in the happening at hand), held before, during and after a happening, in which the tactical details and findings of the happening were discussed, and adaptation measures were discussed; d) the (often collaborative) preparation of happening materials, data gathering instruments, etc. was performed using collaborative tools (such as Google Docs, see https://drive.google.com, last visit: January

2014), which enabled agile and fast preparation and reviewing of materials, coordination of pending tasks, etc.

Adaptation

The awareness processes mentioned above allowed the researcher team to rapidly adapt the evaluation strategy in the face of recently-acquired findings, or to modify the concrete data gathering of a happening in the face of unexpected events of a happening. To illustrate these adaptations, let us look at a few examples which occurred during the evaluation of GLUE!-PS:

- In several of the evaluation's happenings, especially in teacher workshops, the technology under study (or other technologies upon which the happening relied - e.g. the network access in the room) failed unexpectedly (an event that is nevertheless quite common when dealing with prototypes developed for research purposes). These events often decreased the amount and quality of the data gathered, as participants could not experience in full the support that the GLUE!-PS system provided. This, in turn, led to the happening providing insufficient findings about the evaluation issues, and prompted for the realization of further happenings to gather more data.
- Another common adaptation was derived from the fact that teacher workshops often did not follow too closely its original plan (e.g., if participant teachers, or if facilitators spent more time than expected explaining a crucial part of the workshop). The consequent adjustments in the schedule often had an impact in the evaluation's data gathering (e.g., a questionnaire could not be answered, or had to be answered online after the workshop, etc.). In these cases, the dual nature of the teacher workshops as evaluation happenings and as authentic professional development actions forced the researcher team to strike a careful balance between addressing the learning needs of participants, and collecting data for the evaluation (with the former taking precedence over the latter, for ethical reasons).
- Opportunities for emergent happenings (not originally planned in the overall evaluation design) also occurred during the evaluation process, and served to offset the negative impact of the unexpected adaptations mentioned above (see also Figure 4). In this regard,

having a numerous and varied number of teachers as members of the research group, as well as having a track record of professional development actions within the university, proved invaluable for the researcher team. The fact that the GLUE!-PS system was intended to solve existing problems of the teacher community also helped, as it potentially transformed the participation in the evaluation into a win-win situation for participants.

- Another important adaptation that occurred during the evaluation was the modification (or rather, the increased focus) of the different notions that guided the evaluation. As we can see in Figure 4, the research question behind the evaluation was adapted as the features of the GLUE!-PS system evolved (prompted in part by the findings of the different evaluation happenings). The way in which the research question was explored (e.g., through evaluative topics in an anticipatory data reduction method, see (Miles & Huberman, 1994)) also evolved as the researcher team gained an understanding of what the notion of orchestration entailed (prompted in turn by the development of the conceptual framework in (Prieto, Holenko-Dlab, et al., 2011)). The number and nature of happenings, as it has been mentioned, also evolved: as initial evaluations turned out insufficient evidence, new ones were planned, and additional ones emerged as new opportunities to provide further evidence about new system features, or to explore recently-added evaluation topics.

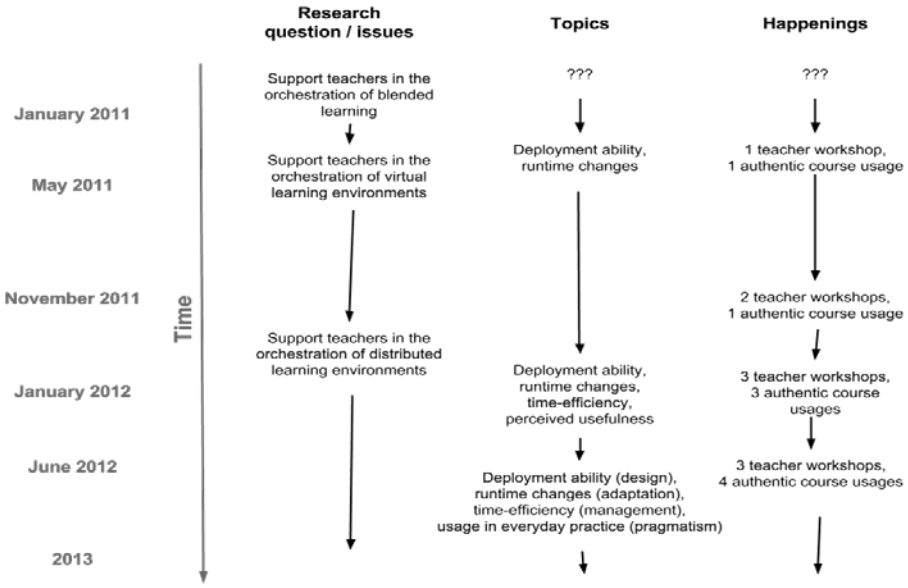


Figure 4. Chronological evolution of selected main concepts in the evaluation of the GLUE!-PS system.

As it can be seen from these adaptations, the evaluation process, which was described in an orderly manner (to be understandable by the readers) in the "front office accounts" of publications, is in reality a much more fluid and malleable process, in which the goals, the analytical lens and the methods used are adapted to the pragmatic constraints and unexpected events of the setting. This can be considered a form of the "progressive in-focus" that characterizes responsive evaluation (Stake, 2010).

Pragmatism

In the previous sections, the impact of several setting constraints have been mentioned, and many others also had to be dealt with by the researcher team: having to adhere with the academic course calendar (both for the inclusion of interventions in authentic course usages, and for programming the teacher workshops in times of lower teacher workload), the (limited) availability of specific people (e.g., teacher researchers and other

informants), the necessity to adapt data gathering to what was feasible to be done by volunteer teachers in the limited time allotted to a teacher workshop, etc. The pragmatic adherence of the researcher team's evaluation activities to what was possible in a certain moment in the setting is also clearly represented by the in-happening adaptations and "damage control" in the face of unexpected occurrences, which had to balance the need for data gathering and the response to the informants' needs in terms of professional development (see 'Adaptation' above).

Synergies

As it can be seen from all of the above, the researcher team tried to make the most of the contextual elements at their disposal: both in terms of human resources (e.g., militant teachers willing to try out the GLUE!-PS system in their courses, workshop participants that agreed to providing information as they learned about CSCL, etc.), as well as technological and material resources (the usage of publicly available tools for coordination and management of the researcher team, specific evaluation tools like the CSCL-EREM platform, university facilities suitable for the kind of collaborative work that the happenings required, etc.).

Conclusion

In this article, we have presented the notion of "orchestrating learning", used in the field of TEL to address the increased complexity of educational practice in authentic settings, and we have applied it to the evaluation of learning technologies in such complex authentic TEL settings, which also has become more intricate. Moreover, we have operationalized this new notion of "orchestrating evaluation" by reusing a conceptual framework for research in TEL orchestration, which aims at helping identify evaluative tensions towards a more holistic view of such orchestration. This transposition can be intuitively justified, for example, if we consider evaluation of learning technologies as a learning process about the impact such technologies in an authentic setting. This evaluation learning process is often collaborative (within a research/evaluator team), supported by computers (hence, CSCL), and bound to the multiple constraints of an authentic educational setting (in which the evaluation occurs). Thus, it has

to be somehow orchestrated. Other frameworks for orchestrating learning have also been proposed, such as (Dillenbourg, 2013)'s "kernel and rings" model. Considering the application of these other models to orchestrating evaluation is left for future efforts along this line.

One of this paper's main contributions is to provide a meta-evaluation of one example evaluation of learning technologies. This structured account illustrates, through a concrete example, many issues commonly mentioned in research methodology manuals (adapting to emergent questions, the evolution of the research questions and their focus, etc.), but whose contextualized operationalization in the field is seldom described. Our "shop floor description" can be related to general evaluation issues such as (Guba, 1981)'s criteria for quality in research, or (R. Stake, 2010)'s progressive in-focus. However, fully exploring these relationships exceeds the scope of this publication, and will have to be addressed in the future.

In this paper we have offered a post-hoc analysis of an existing evaluation of learning technologies, to gain insights into how it was orchestrated. However, the notion of orchestrating learning and the operationalization in different aspects that we have done here could also be applied in other moments of the evaluation process. For example, we could envision applying this notion while designing the evaluation of a learning technology, e.g., by integrating this transposed orchestration framework with existing frameworks for evaluation design, such as the CSCL-EREM (Jorrín-Abellán et al., 2009). Again, this is left for future research, as is also left the potential generalization of this "orchestrating evaluation" framework beyond the evaluation of learning technologies, to evaluation of educational innovations in general, and even beyond that, to a general evaluation approach. The fact that most evaluations today are becoming cross-contextual, require teamwork and the use of multiple technologies, point to an increasing need in the researcher and evaluation communities of support in understanding how we can go from the abstract evaluation manual to the contextualized practice of evaluation within a multiplicity of constraints.

Acknowledgements

The research portrayed in this article has been partially supported by the Spanish Research Project TIN2011-28308-C03-02 and the European Projects METIS (531262-LLP-2012-ES-

KA3-KA3MP) and PREATY (526965-LLP-1-2012-1-GRCOMENIUS-CMP). The authors would like to thank the rest of the GSIC-EMIC research group for their collaboration and ideas, which have helped shape the proposals in this article.

References

Bote-Lorenzo, M. L., Gómez-Sánchez, E., Vega-Gorgojo, G., Dimitriadis, Y., Asensio-Pérez, J. I., & Jorrín-Abellán, I. M. (2008). Gridcole: a tailorable grid service based system that supports scripted collaborative learning. *Computers & Education*, *51*(1), 155–172. doi: 10.1016/j.compedu.2007.05.004

Bruce, B. C. (2008). Ubiquitous Learning, Ubiquitous Computing, and life experience. In *Proceedings of the 6th International Conference on Networked Learning (NLC2008)* (pp. 583–590).

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage..

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492. doi:10.1016/j.compedu.2013.04.013

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The Evolution of Research in Computer-Supported Collaborative Learning: from design to orchestration. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-Enhanced Learning: Principles and Products* (pp. 3–19). New York, NY: Springer.

Draper, S. W. (1997). Prospects for summative evaluation of CAL in higher education. *Research in Learning Technology*, *5*(1), 33-39.

Economides, A. A. (2005). Collaborative Learning Evaluation (CLE) framework. *WSEAS Transactions on Advances in Engineering Education*, *2*(4), 339–346.

Ewing, J., & Miller, D. (2002). A framework for evaluating computer supported collaborative learning. *Educational Technology & Society*, *5*(1), 112–118.

Fischer, F., & Dillenbourg, P. (2006). Challenges of orchestrating computer-supported collaborative learning. In *Paper presented at the 87th Annual Meeting of the American Educational Research Association (AERA)*.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism* (Vol. 6). Lanham, MR: Rowman & Littlefield.

Glass, R. L. (1995). A structure-based critique of contemporary computing research. *Journal of Systems and Software*, *28*(1), 3–7. doi: 10.1016/0164-1212(94)00077-Z

Gómez-Sánchez, E., Bote-Lorenzo, M. L., Jorrín-Abellán, I. M., Vega-Gorgojo, G., Asensio-Pérez, J. I., & Dimitriadis, Y. (2009). Conceptual framework for design, technological support and evaluation of collaborative learning. *International Journal of Engineering Education*, *25*(3), 557–568.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *A Journal of Theory, Research, and Development on Educational Communication and Technology*, *29*(2), 75–91.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: Sage.

Hernández-Leo, D., Asensio-Pérez, J. I., & Dimitriadis, Y. (2005). Computational Representation of Collaborative Learning Flow Patterns Using IMS Learning Desing. *Educational Technology & Society*, *8*(4), 75–89.

Jorrín-Abellán, I. M., & Stake, R. E. (2009). Does Ubiquitous Learning Call for Ubiquitous Forms of Formal Evaluation?: An Evaluand oriented Responsive Evaluation Model. *Ubiquitous Learning: An International Journal*, *1*(3), 71–82.

Jorrín-Abellán, I. M., Stake, R. E., & Martínez-Monés, A. (2009). The Needlework in evaluating a CSCL system: The Evaluand-oriented Responsive Evaluation Model. In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 68–72).

Kovalainen, M., Kumpulainen, K., & Satu, V. (2001). Orchestrating classroom interaction in a community of inquiry: Modes of teacher participation. *Journal of Classroom Interaction*, *36*(2), 17–28.

Lonchamp, J. (2012). An instrumental perspective on CSCL systems. *International Journal of Computer-Supported Collaborative Learning*, *7*(2), 211–237. doi: 10.1007/s11412-012-9141-4

Luckin, R. (2008). The Learner Centric Ecology of Resources: a Framework for using Technology to Scaffold Learning. *Computers & Education*, *50*, 449–462. doi: 10.1016/j.compedu.2007.09.018

MacNeill, S., & Kraan, W. (2010). *Distributed Learning Environments: A Briefing Paper*. Retrieved from http://publications.cetis.ac.uk/wp-content/uploads/2011/02/Distributed_Learning.pdf

Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, *41*(4), 353–368. doi: 10.1016/j.compedu.2003.06.001

Martínez-Monés, A., Gómez-Sánchez, E., Dimitriadis, Y., Jorrín-Abellán, I. M., Rubia-Avi, B., & Vega-Gorgojo, G. (2005). Multiple Case Studies to Enhance Project-Based Learning in a Computer Architecture Course. *IEEE Transactions on Education*, *48*(3), 482–489.

McKenney, S. (2013). Designing and researching technology-enhanced learning for the zone of proximal implementation. *Research in Learning Technology*, *21*. doi: 10.3402/rlt.v21i0.17374

Miles, M. B., & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Newbury Park, CA: Sage.

Oliver, M. (2000). An introduction to the Evaluation of Learning Technology. *Educational Technology & Society*, *3*(4), 20–30.

Oliver, M., & Conole, G. (2004). Evaluating information and communications technology: a tool kit for practitioners. In R. Holliman & E. Scanlon (Eds.), *Mediating Science Learning through Information and Communications Technology* (pp. 139–150). Routledge.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, *2*(1), 1–28.

Perrotta, C., & Evans, M. A. (2013). Orchestration, power, and educational technology: A response to Dillenbourg. *Computers & Education*, *69*, 520–522. doi: 10.1016/j.compedu.2013.04.007

Pozzi, F., Manca, S., Persico, D., & Sarti, L. (2007). A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments. *Innovations in Education and Teaching International*, *44*(2), 169–179. doi: 10.1080/14703290701240929

Prieto, L. P. (2012). *Supporting orchestration of blended CSCL scenarios in Distributed Learning Environments*. School of Telecommunications Engineering, University of Valladolid, Spain.

Prieto, L. P., Asensio-Pérez, J. I., Dimitriadis, Y., Gómez-Sánchez, E., & Muñoz-Cristóbal, J. A. (2011). GLUE!-PS: A multi-language architecture and data model to deploy TEL designs to multiple learning environments. In *Proceedings of the European Conference on Technology-Enhanced Learning (EC-TEL 2011)* (pp. 285–298).

Prieto, L. P., Asensio-Pérez, J. I., Muñoz-Cristóbal, J. A., Dimitriadis, Y., Jorrín-Abellán, I. M., & Gómez-Sánchez, E. (2013). Enabling Teachers to Deploy CSCL Designs Across Distributed Learning Environments. *IEEE Transactions on Learning Technologies*, *6*(4), 324–336. doi: 10.1109/TLT.2013.22

Prieto, L. P., Asensio-Pérez, J. I., Muñoz-Cristóbal, J. A., Jorrín-Abellán, I. M., Dimitriadis, Y., & Gómez-Sánchez, E. (2014). Supporting orchestration of CSCL scenarios in web-based Distributed Learning Environments. *Computers & Education*, *73*, 9–25. doi: 10.1016/j.compedu.2013.12.008

Prieto, L. P., Holenko-Dlab, M., Abdulwahed, M., Gutiérrez, I., & Balid, W. (2011). Orchestrating Technology Enhanced Learning: a literature review and a conceptual framework. *International Journal of Technology-Enhanced Learning (IJTEL)*, *3*(6), 583–598.

Prieto, L. P., Villagrá-Sobrino, S., Dimitriadis, Y., Schank, P., Penuel, W., & H., D. A. (2011). Mind the Gaps: Using Patterns to Change Everyday Classroom Practice Towards Contingent CSCL Teaching. In *Proceedings of the International Conference of Computer-Supported Collaborative Learning (CSCL2011)* (Vol. 1, pp. 518–525).

Roschelle, J., Dimitriadis, Y., & Hoppe, U. (2013). Classroom orchestration: Synthesis. *Computers & Education*, *69*, 523–526. doi: 10.1016/j.compedu.2013.04.010

Schön, D. A. (1983). *The Reflective Practitioner - How Professionals Think in Action*. New York, NY: Basic Books.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*. Cambridge: Cambridge University Press.

Stake, R. E. (2010). *Qualitative Research: Studying How Things Work*. The Guilford Press.

Stake, R. E. (1983). Program Evaluation, Particularly Responsive Evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.) ,Evaluation Models (pp. 287–310). Dordrech: Springer. doi: 10.1007/978-94-009-6669-7_17

Stake, R. E. (2004). *Standards-based and responsive evaluation*. Sage.

Strijbos, J. W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction*, *17*(4), 389–393.

Sutherland, R., & Joubert, M. (2009). *D1.1: The STELLAR vision and strategy statement*. Retrieved from http://www.stellarnet.eu/kmi/deliverables/20090929_d1-1___vision-and-strategy.pdf

Treleaven, L. (2004). A new taxonomy for evaluation studies of online collaborative learning. In *Online collaborative learning: Theory and practice* (pp. 160–179). Idea Group Inc. (IGI).

Tsiatsos, T., Andreas, K., & Pomportsis, A. (2010). Evaluation Framework for Collaborative Educational Virtual Environments. *Educational Technology & Society*, *13*(2), 65–77.

Vatrapu, R., Suthers, D. D., & Medina, R. (2008). Usability, sociability, and learnability: A CSCL design evaluation framework. In *Proceedings of the 16th International Conference on Computers in Education*. Taipei, Taiwan.

Villasclaras-Fernández, E., Hernández-Leo, D., Asensio-Pérez, J. I., & Dimitriadis, Y. (2013). Web Collage: An implementation of support for assessment design in CSCL macro-scripts. *Computers & Education*, *67*, 79–97. doi: 10.1016/j.compedu.2013.03.002

Luis P. Prieto is Research Scientistat the CHILI Lab, École Polytechnique Fédérale de Lausanne, Switzerland. ORCID id: 0000-0002-0057-0682

Yannis Diamitriadis is Full Professor in the Department of Signal Theory of University of Valladolid, Spain. ORCID id: 0000-0001-7275-2242

Juan I. Asensio is Professor in the Department of Signal Theory of University of Valladolid, Spain. ORCID id: 000-0002-1114-2819

Contact Address: EPFL IC ISIM CHILI, RLC D1 740 (Rolex Learning Center), Station 20, CH-1015 Lausanne, Switzerland. E-mail: luis.prieto@epfl.ch