The Design of Wizard-of-Oz Studies to Support Students' Learning to Learn Together

Manolis Mavrikis[1], Toby Dragon[2], and Bruce M. McLaren[2,3]

[1] London Knowledge Lab, Institute of Education, London, UK
[2] Center for E-Learning Technology (CeLTech), Saarland University, Germany
[3] Carnegie Mellon University, USA

Abstract. This paper presents the design of forthcoming wizard-of-Oz (WOZ) studies in the context of the EU-funded Metafora project that is developing a Computer-Supported Collaborative Learning (CSCL) system to scaffold 12 to 16-year-old students in learning to learn together (L2L2). In the Metafora system, students undertake lengthy collaborative challenges interacting in 3 different spaces: a planning tool, a discussion tool and a domain-specific microworld. In the meantime they are encouraged to reflect on the L2L2 process. We outline the difficulties behind the design of intelligent support for such a complex pedagogical scenario and the need for WOZ studies. We then present how we plan to ease the wizard's role by providing appropriate tools and templates with the main goal of evaluating potential analysis and feedback possibilities.

1 Introduction

In a world where advancements in communication technologies have enabled the emergence of computer-supported collaborative learning, learners require not only appropriate pedagogical domain-level support — an area where ITS research has focused its attention since its inception [1] — but also support in developing the collaborative skills that the so called 'knowledge age' demands. The Metafora project[1] is developing a Computer-Supported Collaborative Learning (CSCL) system to scaffold students, 12 to 16 years old, while they undertake collaborative and reflective activities to improve their skills in learning to learn together (L2L2).

Designing the appropriate intelligent support for L2L2 in a system like Metafora requires not only an operationalization of collaboration skills but also a context-dependent analysis of how L2L2 manifests under different circumstances. However, the ill-defined nature of collaborative learning makes the task of identifying how to automatically support collaboration more difficult than it would be for a well-defined domain. In particular, with the exception of a very few notable studies (e.g. [2]) there is a lack of research on how and when to provide relevant, timely and effective feedback particularly to the extent that it can be operationalised. A principled approach for the design of the intelligent support for collaboration is to rely on knowledge elicitation from experts. However, it is now well understood that it is very difficult to elicit precise, concise and operationalized knowledge from teachers or educators, especially when this requires an understanding of novel interaction with systems for which they may lack expertise (c.f. [3,4]). The difficulty is exacerbated in complex, collaborative, multi-tool learning

[1] http://www.metafora-project.org

scenarios because the bandwidth of information and communication modalities available to an intelligent system can be very different than those available to a human. In other words, it is hard to understand how the system can recognize or interact with students to offer feedback that is similar to examples of teacher interaction.

A well-known approach to deal with these difficulties is Wizard-of-Oz (WOZ) studies that employ a human to provide intelligent feedback through system functionality, based on the actions taken by students. The students working with the system do not know that a human is providing feedback; they assume it is the system at work. WOZ studies can have both exploratory and evaluative objectives and have been used mostly to study student-system interaction in well-defined domains. While few studies have been conducted for collaborative and/or exploratory environments [3,5,6] our goal is somewhat more ambitious and complex in nature. We seek to offer an appropriate setup to help the wizard provide adaptive support not only in reaction to student collaboration but also to support their L2L2 explicitly. The situation is also complicated by the multi-faceted context that emerges from the complex challenges and multiple tools that students interact with in the Metafora system.

The aim of this paper is to propose WOZ tools and a specific setup to support students in using the Metafora system. We also present general methodological lessons about the use of WOZ studies in collaborative, exploratory environments.

2 The Metafora pedagogy and platform

The pedagogy behind the Metafora project relies on scenarios around one of several lengthy or real-world challenges in mathematics or science that students are asked to undertake. The challenges encourage students to interact with microworlds (including simulators and games) where they either build digital artifacts (models) that allow them to engage in collaborative problem solving or simply test hypotheses or theories related to the challenges. The activities that students undertake have been described in detail in other publications [7–9]. One of these activities (relevant to the remainder of the paper) involves the eXpresser microworld [10] where students can construct animated figural patterns (like the one that appears embedded in a discussion contribution, Fig. 2**)**. These patterns are only colored correctly once students find their representative algebraic rules. Underlying this surface goal, the main objective is to enhance students' appreciation of the power and flexibility of algebra [9]. One of the related Metafora challenges is to find as many ways to make the pattern and corresponding algebraic rule as possible and to compare the correctness and equivalence of the rules with other teams.

The Metafora platform supports the L2L2 process in several ways. First, it provides a space for planning and reflecting using a multi-user, web-based application of visual cards that present different learning processes and stages with which students continuously make plans and reflect on their work (see Fig. 2). In addition, as students undertake the challenge, they are encouraged to discuss and argue about their ideas using a specialized discussion tool – LASAD [11] – that offers a structured approach to discussion through argumentation graphs (see a simple example in Fig. 2). The discussion space is customized to allow students to share artifacts created within other tools. These *referable objects* are thumbnails that can be included in the discussion, as illustrated by the "1 – Help Request" node in Fig. 2.

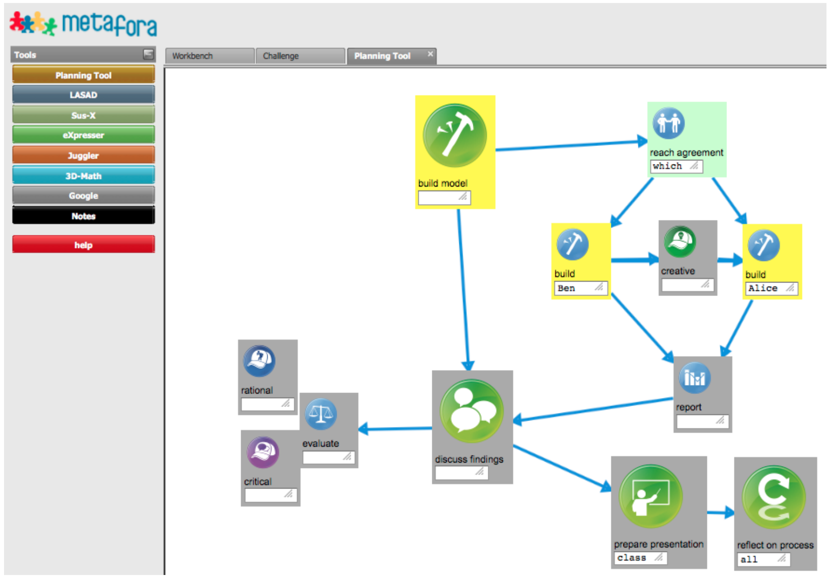


Figure 1. A screenshot of the Metafora Planning Tool. Cards refer to activity stages or processes . Started activities as marked as yellow and finished ones as green.

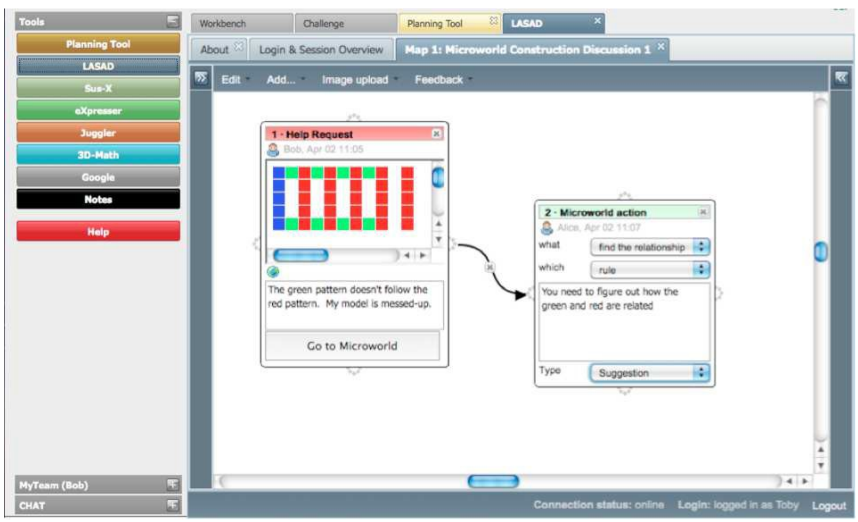


Figure 2: A sample discussion map in LASAD. A figural pattern made with the eXpresser microworld is embedded as a thumbnail within a Help Request (left), to which another student offers a suggestion (right).

A combination of literature reviews and early pilots with the Metafora platform has allowed us to identify high-level processes and lower-level behaviors that demonstrate L2L2 competences. While some competences are scenario- and domain-dependent we consider *distributed leadership*, *mutual engagement*, *help seeking/giving*, and *reflection on the group learning process* as general competences that are important when learning together and on a higher level, learning how to learn together. With this perspective in mind, the aim of the Metafora project is to provide both a cognitive tool that encourages L2L2 behaviors as well as to provide intelligent support to teachers and students during the learning process.

3 WOZ in Metafora

3.1 Objectives

As mentioned in the introduction, WOZ studies can have both exploratory and evaluative objectives. The WOZ studies proposed in this paper emphasize evaluation, but also allow for some exploration. Initial work to develop a coherent vision of learning to learn together (L2L2) has included a combination of literature review, meetings with educators, and observations from early pilot studies with students using the system. This L2L2 concept has manifested as a set of interesting and expected behaviors that arise, or should arise, when the Metafora system is used. For each of these behaviors, we have defined interaction indicators that could be used to detect the behavior, and corresponding feedback that we hypothesize would be useful to provide. Our WOZ studies aim to evaluate these indicators and their corresponding feedback while also allowing for some exploration around the timing and importance of the feedback that has been defined. Specifically, the studies have the following objectives:

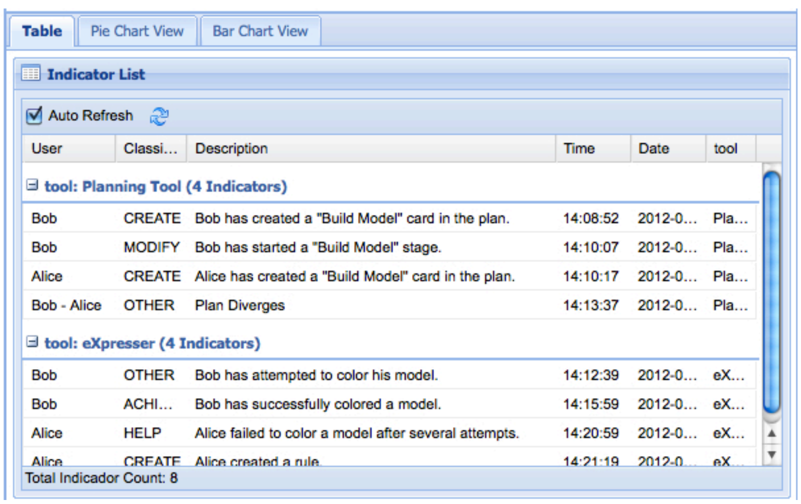
- Evaluate indicators (see section 3.2):
  - Is the current set of indicators (mappings between system activity and behaviors) appropriate? Do the indicators produced by the analysis components provide sufficient context to allow the wizard to recognize behaviors without observing the direct student interaction?
  - Which indicators are potentially useful to be seen directly by students and teachers?
  - Are there missing indicators needed to enable/enhance decision-making (for the wizard and / or the system)?
- Evaluate potential automated feedback (see section 3.3):
  - Is the current set of rules (mappings between behaviors and feedback) appropriate? Is the pre-specified feedback appropriate when a given behavior is recognized?
  - Is the feedback provided to the students relevant and/or helpful in promoting reflection on L2L2?
- Explore timing and methods of feedback (see section 3.3)
  - Should students be interrupted to offer feedback? If so, when should they be made aware that feedback is available?

Having established the objectives, a pressing concern is how to make the studies possible in the complex learning scenarios afforded by Metafora. We developed two sets of tools to ease the process. The first we term the Monitoring Tool, which offers the wizard different filtered views of the live indicators reported by the analysis components of the Metafora system. This visualization can provide insight into how the learning process is progressing for individuals and teams, and generally supports the wizard's ability to analyze student activity. In so far as the current set of indicators and filtering abilities are appropriate, the wizard should be able to exclusively monitor these filtered views in order to recognize behaviors. Second, the Feedback Tool offers a GUI interface from which the wizard can generate feedback for the students, simulating the feedback that we envisage the automated feedback system to provide in the future. These two tools are presented below.

The wizard also has access to shared artifacts created by the students, i.e. the plans created in the planning tool, any discussion within LASAD, as well as the models created in the microworld. This direct access to student artifacts is offered mainly to address the exploration side of the WOZ studies since it allows the wizards to observe and respond to behaviors that we had not anticipated in advance. This could lead to identification of new and interesting L2L2 behaviors and similarly allow for current or new feedback to be given even for those behaviors not captured by the current indicator set.

3.2 Analysis and the Monitoring Tool

The final automated analysis system will need to recognize behaviors from system interaction and map these behaviors to feedback. In order to complete the first step in this process (the recognition of behaviors), we use a system of indicators. As students are interacting with each tool, individual intelligent components of the tool analyze their work and post indicators to a centralized analysis communication channel. This automated analysis ranges from low-level indicators (such as indicating the creation or modification of artifacts) to high-level analyses (such as identification of whether a student is achieving an objective in a microworld or facing a particular difficulty). These analyses are collected by a centralized Metafora analysis component, and can then be visualized and analyzed further. The theory behind this work and first implementation steps is discussed in more detail in [9]. To aid the work of the wizard, this analysis information is used to provide a useful summary of information through visualization tools that filter and aggregate the indicators available in real-time (see an example in Fig 3).



User	Classi...	Description	Time	Date	tool
tool: Planning Tool (4 Indicators)					
Bob	CREATE	Bob has created a "Build Model" card in the plan.	14:08:52	2012-0...	Pla...
Bob	MODIFY	Bob has started a "Build Model" stage.	14:10:07	2012-0...	Pla...
Alice	CREATE	Alice has created a "Build Model" card in the plan.	14:10:17	2012-0...	Pla...
Bob - Alice	OTHER	Plan Diverges	14:13:37	2012-0...	Pla...
tool: eXpresser (4 Indicators)					
Bob	OTHER	Bob has attempted to color his model.	14:12:39	2012-0...	eX...
Bob	ACHI...	Bob has successfully colored a model.	14:15:59	2012-0...	eX...
Alice	HELP	Alice failed to color a model after several attempts.	14:20:59	2012-0...	eX...
Alice	CREATE	Alice created a rule.	14:21:19	2012-0...	eX...

Figure 3: The Monitoring tool displaying various indicators from different tools in a table format. The wizard can filter, sort, and see different views of these indicators, including pie or bar charts.

3.3 Feedback

The next step to simulating an intelligent system is to understand, from recognized behaviors, what feedback should be given and when. As mentioned previously, we have developed a tentative mapping of behaviors to feedback from both theory and classroom experience (examples are offered in section 3.4). From a theoretical standpoint, the method to implement these mappings is straightforward. When a behavior is recognized, the wizard offers the related feedback. However, an open question is deciding *when* to intervene. The assistance dilemma [12] — the difficult choice between providing or withholding guidance, common to most ITS — is even greater in collaborative, exploratory learning. Provision of support requires not only deciding whether to provide assistance or not but also when to interrupt the students' interaction. Due to the highly interactive nature of direct manipulation environments, students' work should be (for the most part) interrupted only when there is a clear opportunity for reflecting on their L2L2 or when they explicitly seek for help or advice (e.g. realizing they are stuck) [3,13]. As this requirement is not detailed enough to be directly operationalized, the goal of the WOZ studies is to specify a more precise definition of "for the most part", and to understand if the "L2L2 opportunities" we have identified justify interruptions. Accordingly, we define three different forms in which any given feedback can be offered:

- *High Interruption Feedback* (HIF): A message interrupts the interaction, which cannot continue until the message has been acknowledged (i.e. a modal prompt is opened in the middle of the screen, that the student must acknowledge).
- *No Interruption Feedback* (NIF): This type of feedback appears in a special area like a history of available feedback. A student can access the relevant area when desired (or prompted by teachers and/or the wizard).
- *Low Interruption Feedback* (LIF): This involves a notification of the availability of feedback and the action flow of the student is not explicitly interrupted. The student can access the (full) feedback by going to the NIF area.

The tool that the wizard uses to provide feedback is seen in Fig 4. It allows the choice of feedback type and the recipient (all students in a team or specific students). To speed the process of writing feedback but also to evaluate the specific messages that we have designed, the tool provides message templates that correspond to the various stages of the activity the students are undertaking and the different tools they may be using. It also keeps a history of messages that have been sent so they can be re-used.

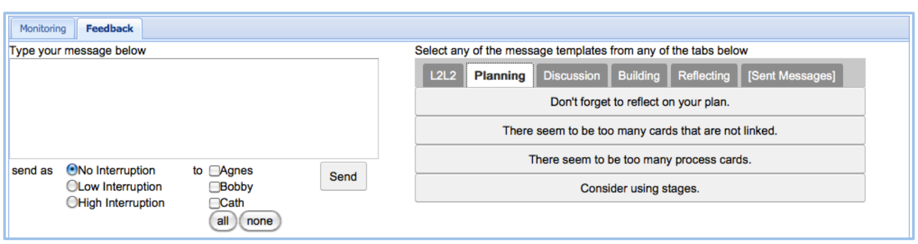


Figure 4: The feedback provision tool. The wizard can use or adapt existing 'template' feedback statements from different tabs (right) representing different learning stages (from the planning tool). The wizard is also able (but discouraged) to create new messages (left). Before sending the message, the wizard has to choose the level of interruption and the recipients (bottom left).

3.3 WOZ example scenario

The design of the WOZ studies in Metafora must be general enough to accommodate different learning scenarios; led by 4 different institutions, using different microworlds, and with different populations of students. While we remain as "scenario general" as possible in our analysis and WOZ setup, below we provide a brief scenario that focuses on a challenge involving eXpresser microworld (presented in Section 2) to offer a clear concept of what a wizard will see and do in one specific instance of the tool.

One situation that invites feedback from an L2L2 perspective is when students in a group create a plan that does not diverge as expected. As students are expected to construct different eXpresser models, lack of divergence in the plan could imply that they are working on the same model and will miss the opportunity to compare different approaches. This can be recognized directly by indicators from the planning tool that reports on divergence in the student plans. The role of the wizard in this case is to suggest that they create different models. More often, we have observed a plan diverging but not converging to a common discussion. Appropriate feedback in these situations (according to our pilot studies) would be to encourage students to discuss their separate results, as this can introduce each to different perspectives and allow them to reach generalisations. Watching for specific indicators, the wizard can recognize these indicators and prompt the students accordingly, see Table 1.

A more complex situation occurs when a student has several failed attempts at finding a rule in eXpresser, but does not seek for help. The wizard can detect this situation by observing indicators from eXpresser about failed attempts to construct a model. Here, one commonly-observed form of feedback from a teacher (again from our pilot studies) is to prompt the struggling student to converse with another student who has already successfully built a model. This would have been posted as an indicator from the planning tool (BUILD_DONE) and, if the model is correct, there would be a corresponding indicator from eXpresser (MODEL_COLORING _CORRECT). Noticing these indicators, the wizard can choose to provide the relevant feedback (Table 1).

Table 1: Sample rules for the wizard to follow and prompts (co-designed with teachers and educators in advance) to be evaluated during WOZ studies.

Stage: Planning	
Rule for wizard (ID: DIVERGE-CONVERGE)	**Feedback Template**
If a plan diverges (PLAN_DIVERGE indicator) and does not converge (PLAN_CONVERGE) in a reasonable amount of time.	Is your group's plan diverging? Consider comparing your separate work.
Stage: Building a model	
Rule for wizard (ID: STUDENT-IN-NEED)	**Feedback Template**
If a student X has finished building a correct model (BUILD_DONE & MODEL_CORRECT) and another student Y is struggling (continuous FAILED_ATTEMPTS), and student Y has not started a discussion (no discussion ACTIVITY indicators for x and y), then prompt for discussion.	Do you have difficulties with your eXpresser model? Consider asking for help from others. Student/Group X has finished the task.
Stage: Discussion	
If there are several contributions that are not attended (NO_LINKS_FOR_CONTRIBUTIONS), then draw students' attention to them.	There are several contributions in the discussion that demand your attention.

4 Conclusions & Future work

The paper presented our plans for WOZ studies in Metafora. The main contribution of this paper is the presentation of a design of both tools and a conceptual workflow that allows a wizard to take the place of the still-unimplemented portions of our analysis and feedback system. This requires the wizard to have the ability to recognize key behaviors within the system and to offer feedback to students in the same form that the system will eventually provide. The indicators created by the integrated tools enable monitoring of the behaviours. The Feedback Tool with the choice of different interruption types and the templates that correspond to different activity stages ease the wizard's role. In addition, by giving some freedom to the wizard, we will explore the difficult question of when to interrupt and accommodate for unanticipated situations. By providing feedback templates and rules for different situations (see table 1), we will iteratively evaluate and improve the design of the actual feedback provided.

As immediate future work we will conduct training sessions with wizards who will take part in the studies and who are experts in the different scenarios available in the Metafora platform. In each scenario the general L2L2 processes on which we focus (distributed leadership, mutual engagement, help seeking/giving, and reflection on the group learning process) are manifested in different ways. Therefore different rules are required to monitor and support them. To ease the wizard's decision-making process we are planning to co-design with wizards (and the pedagogical designers responsible for each scenario) high-level flowcharts. These flowcharts start with the activity stages (e.g. planning, interacting with microworlds, discussing reflecting) and, within each branch, they provide the possible behaviors that the wizard should monitor. Such flowcharts have been reported to be successful in previous WOZ studies [e.g. 6]. The branches in the flowchart will correspond to the wizard interface (e.g. the tabs on Figure 4) helping the wizard easily find the appropriate set of feedback templates for a given situation.

Given the difficulty of eliciting pedagogical knowledge in relation to providing explicit support to students, and the high cost associated with developing AI techniques, we aim for the WOZ approach and tools that we have described to be used beyond our specific project. They could support other cases where AI needs to be developed for integrated exploratory and/or collaborative environments and more generally, any cases where it is important to evaluate the intelligence of the system before implementing it. In particular we would like to extend the tools beyond the specifics of Metafora; thus, we welcome feedback on our approach.

Acknowledgements

Thanks to Stefan Kreitmayer and Ugur Kira for help in developing the wizard-of-oz monitoring and feedback tool. We would also like to acknowledge the rest of the members of the Metafora team (see http://www.metafora-project.org) for their contributions and stimulating discussions

The Metafora project is co-funded by the European Union under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (FP7), Contract No. 257872. This document does not represent the opinion of the EU, and the EU is not responsible for any use that might be made of its content.

References

1. VanLehn, K.: The behavior of tutoring systems. International Journal of Artificial Intelligence in Education **16**(3) (2006) 227–265
2. Saab, N., Van Joolingen, W. R., & Van Hout-Wolters, B. H. A. M. (2007). Supporting communication in a collaborative discovery learning environment. *Instructional Science*, 35 (1), 73-98.
3. Mavrikis, M., Gutierrez-Santos, S.: Not all wizards are from Oz: Iterative design of intelligent learning environments by communication capacity tapering. Computers & Education 54(3) (2010) 641–651
4. Porayska-Pomsta, K., Mavrikis, M., Pain, H.: Diagnosing and acting on student affect: the tutor's perspective. User Modeling and User-Adapted Interaction **18**(1) (2008) 125–173
5. Tsovaltzi, D., Rummel, N., Pinkwart, N., Scheuer, O., Harrer, A., Braun, I., McLaren, B.M.: CoChemEx: Supporting Conceptual Chemistry Learning via Computer-Mediated Collaboration Scripts. In: Proceedings of the Third European Conference on Technology Enhanced Learning (ECTEL-08). (2008)
6. Braun, I., Rummel, N.: Facilitating learning from computer-supported collaborative inquiry: the challenge of directing learners' interactions to useful ends. Research and Practice in Technology Enhanced Learning **5**(3) (2010) 205–244
7. Smyrnaiou, Z., Moustaki, F., Kynigos, C.: Students' constructionist game modelling activities as part of inquiry learning processes. Electronic Journal of e-Learning. Special issue on Games-Based Learning (2012)
8. Abdu, R., DeGroot, R., Drachman, R.: Teacher's Role in Computer Supported Collaborative Learning, CHAIS conference (2012) 1–6
9. Dragon, T., McLaren, B.M., Mavrikis, M., Geraniou, E.: Scaffolding Collaborative Learning Opportunities: Integrating Microworld Use and Argumentation. In Ardissono, ed.: Advances in User Modeling: UMAP 2011 Workshops (pp. 18-30), Girona, Spain, July 11-15, 2011, Revised Selected Papers. Volume 7138 of Lecture Notes in Computer Science., Girona, Spain (2012) 18–30
10. Mavrikis, M., Noss, R., Hoyles, C., Geraniou, E., (in press). Sowing the seeds of algebraic generalisation: designing epistemic affordances for an intelligent microworld. Special Issue on Knowledge Transformation, Design and Technology, Journal of Computer Assisted Learning. Available in early view online http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2729.2011.00469.x/abstract
11. Loll, F., Pinkwart, N., Scheuer, O., McLaren, B.M. (2012) How Tough Should It Be? Simplifying the Development of Argumentation Systems using a Configurable Platform. In Pinkwart, ed.: Educational Technologies for Teaching Argumentation Skills. Bentham Science Publishers
12. Koedinger, K.R., Aleven, V.: Exploring the Assistance Dilemma in Experiments with Cognitive Tutors. Educational Psychology Review 19(3) (2007) 239–264
13. Mavrikis, M., Gutierrez-Santos, S., Geraniou, E., Noss, R. (in press) Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. Journal of Personal and Ubiquitous Computing. Available in early view online http://www.springerlink.com/content/h714222602173300/