Figure 1a. Students construct and run simulations of a warehouse model built with tangible objects (shelves, docks, etc.)

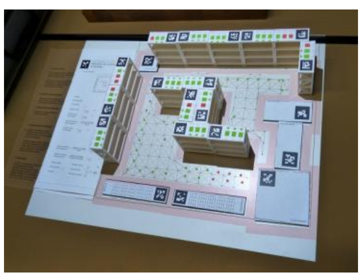

Figure 1b. A camera and a projector enable visual feedback to be augmented directly on top of the physical model.

Classroom-experience evaluation: Evaluating pervasive technologies in a classroom setting

Son Do-Lenh[1]
son.dolenh@epfl.ch

Patrick Jermann[1]
patrick.jermann@epfl.ch

Christoph Arn[2]
christoph.arn@ehb-schweiz.ch

Guillaume Zufferey[1]
guillaume.zufferey@epfl.ch

Pierre Dillenbourg[1]
pierre.dillenbourg@epfl.ch

[1] Swiss Federal Institute of Technology Lausanne (EPFL), Rolex Learning Center, Lausanne, Switzerland

[2] Swiss Federal Institute for Vocational Education and Training (SFIVET), Zollikofen, Switzerland

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
In this paper, we suggest the use of an ecological approach to measuring the effectiveness of pervasive technologies in a classroom setting. We report the lessons learned from evaluating the TinkerLamp, an interactive tabletop interface. We illustrate that due to the many factors involved in authentic settings, the technologies would better be evaluated based on how well they support and enhance the experience of the classroom ecology in addition to based on learning outcomes.

Motivation
Due to the novelty and the difficulty in deployment, most evaluations of pervasive user interfaces (UIs) (interactive tabletops, ubiquitous devices, tangible user interfaces, etc.) are done in the lab. However, when it comes to HCI for education, it is increasingly important to examine the roles of pervasive UIs in broader and authentic contexts, especially in classrooms, rather than only in lab settings.

What to collect in the classroom to measure the effectiveness of the UI is an interesting and challenging issue. Learner-centered ratings have dominated as the primary measures of traditional UIs effectiveness [1]. A common way is to use test scores as a measure of learning outcomes. An interface is deemed effective if it

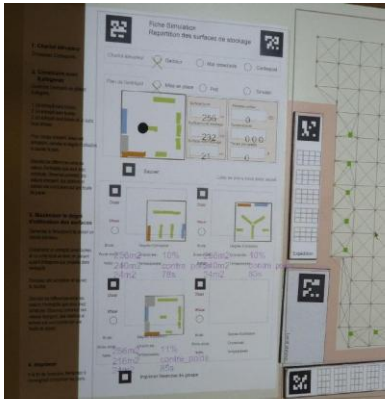

Figure 2. Students use a paper-based interface, the TinkerSheet, to change parameters of the system and to view the summarized outputs of a simulation run on top of the model.

Figure 3. The baseline condition with students studying using paper and pens. See more in [1].

enhances the post-test score or the relative difference between the pre- and post-test scores, namely learning gain.

In this paper, we present our experience of evaluating an interactive tabletop interface in a classroom setting to argue that pervasive UI technologies for education should be evaluated with the "classroom experience" in mind if they are to be fairly and more fully evaluated, in complement to learning outcomes.

TinkerLamp and its evaluations

We have developed a tabletop system, the TinkerLamp (Fig. 1a) for training apprentices in logistics. Group of apprentices can perform problem-solving activities using a tangible interface (Fig. 1b) and a paper-based interface (Fig. 2).

After running usability lab studies, the system has been deployed and used in several vocational schools for two years. We conducted field evaluations with nearly 300 students and 8 teachers in several separate studies from 2008 to 2010. The evaluations involved groups of students studying in a classroom setting, under two conditions: either using TinkerLamp or a baseline condition with traditional paper and pens (see e.g.[1], Fig. 3). At the end of each study, the students had to complete an individual post-test for us to measure the effectiveness of the tabletop UI on learning outcomes.

Our studies led to contradictory results. For example, in [1], we have shown that apprentices who worked collaboratively around the TinkerLamp did not gain statistically better learning outcomes (reflected by post-test scores) with respect to those who performed the same activity in the baseline condition. However, in another study with a different task, we had evidence that the tabletop helped students to have statistically better test scores.

Based on these findings, one could be confused in evaluating the interface, or even could mistakenly argue that the tabletop interface did not (or did) help students to learn more than the baseline condition with only paper and pens. However, our qualitative analyses of logs, field notes and video recordings suggest that several problems distorted the fairness of the learning outcomes-based evaluation approach in these studies.

First, they typically took place in only 1 session of 3 hours. It can be said that it is hard to observe any significant effects on learning about logistics concepts in such a short exposure time to the system. Another recent research involving students reviewed biological contents using a tabletop led to a similar result in terms of exam scores. Their study lasts for 4 sessions of 1 hour each [2].

Second, as the post-test is carried out with paper and pen, one can argue that there is a bias towards the baseline condition. Although doing tests on paper with pen has been traditional and legitimate in schools, we believe that the students working with the TinkerLamp and its interface may not develop the same skills as those in the baseline condition. Hence, perhaps it would have made more sense if different tests had been used, or another approach had been used to evaluate the effectiveness of the UI.

Third, various factors in a classroom (which could not and should not be controlled) could contribute and affect the learning outcomes. For example, our qualitative analysis revealed the teacher as an extremely important factor in the classroom despite the

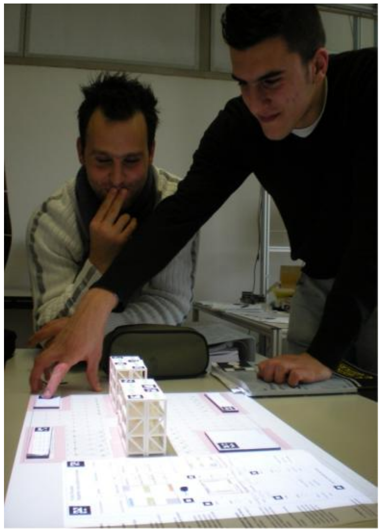

Figure 4. Pervasive technologies in general and tabletop interface in particular support more social processes and interactions in the classroom. Hence, these processes and interactions need to be paid more attention in the evaluation.

technology. Evidences showed that student's learning depends largely on the teacher teaching, his classroom orchestration, his interactions with the students, etc. Another example, students in these real settings tended to have much more off-task conversations which effectively reduced their time discussing about the lesson, which in turn greatly affected their own learning outcomes.

Towards measuring UIs effectiveness based on classroom experience

Given the inconsistent results given by the learning-outcomes approach presented above, we propose the use of classroom-experience evaluation, a unified conceptualization of measuring the effectiveness of pervasive technologies in classroom contexts.

Learning is also about the process in addition to the learning outcomes which are only its product/by-product. Sometimes the skills or the experience the students gain during the class (hence classroom-experience), e.g. discussing in group or with the teacher, is invaluable despite possibly not being related to the test.

The classroom experience is even more important when pervasive technologies come into play (Fig. 4). The focus of evaluation of pervasive UIs is clearly no longer only on the interactions between a single user and the technology as is the case in traditional UI. It is now also on the interactions and social processes between people in the classroom since they are not as constrained by obtrusive technologies.

Several works in HCI research have followed this trend to examine the pervasive UIs in a broader context, e.g.[3]. To better guide the classroom-experience evaluation process though, we propose the adoption of Distributed Cognition Theory and Information Ecologies perspectives [4,5]. From these perspectives, learning is situated in an ecology. While a traditional UI with a single user is too simple to be seen as an ecology, a classroom with pervasive UIs inside obviously forms one. This classroom-ecosystem has many species, processes and artifacts involved. The species consist of teacher, students, technicians, etc. The processes involve the pedagogical and technological workflows, the transfer of experience and knowledge between the teachers and students and among students, etc. The artifacts are educational scenarios, physical devices and its arrangement, learning artifacts, etc. A change in one element can be felt throughout the classroom. All elements in the classroom are inter-dependent and co-evolve.

Following [5], we define five crucial components of a classroom eco-system and the evaluation of the UI becomes a process of examining how well it supports those components. First, how well does the UI support the classroom to enforce its *system* status? Given the complexity of causal relations within an ecosystem, external interventions (the integration of the UI) have to be minimalist, both in terms of design and in terms of effects on existing elements such as the compatibility with text books, with other technologies, with current teaching and learning practices. Another question is how the UI facilitates the transfer of learning artifacts created before, during and after the class.

Second, how well does the UI support the *diversity* of the classroom? The diversity is crucial to the health of an ecology. The UI should not reduce the richness of experience and environment of the classroom. Besides that**,** we need to evaluate how well the system is

Figure 5. The teacher, his teaching activities, and his classroom orchestration need to be considered in evaluating the effectiveness of pervasive UIs on learning.

designed for flexibility (designed for different students and teachers with different skills and motivation, taking into account unexpected events, e.g. a student being sick leading to a change in group size, etc.).

Third, one should evaluate how the UI supports the *co-evolution* of all elements inside the classroom, including the UI. How well it supports the co-experience (creating meaning and emotion together through technology use) shared by the students and teacher? An example is how the UI allows the teacher and students to share a history of what has been done in the classroom.

Fourth, inside the classroom, the teacher can be considered as a *keystone species.* He/she is the decisive factor in a classroom context and has influence over the whole class. Hence, how the UI facilitates his teaching and his classroom orchestration is essential in measuring its effectiveness (Fig.5). Some other examples include: how the teacher can use it unaided in front of the class, how it helps the teacher to track what is going on in the classroom and in the groups.

Fifth, the *habitation* of the UI in the classroom also plays a role in the evaluation, e.g. how smooth it is adapted in the learning scenario, how well it is connected to other elements in the class, how it fits into the physical classroom arrangement, whether it has a permanent position inside the room, etc.

Given the complexity of the classroom eco-system, it is reasonable to expect that multiple sources can provide a more reliable and comprehensive picture of the effectiveness of the interface to learning than just one source. One practical consideration arguing against the use of this approach is the time and effort required. In the end, the purpose of the research and the nature of the classroom (student population, teaching subject, etc.) are likely to dictate how many sources the researchers collect and analyze.

The idea of using multiple data sources in evaluation and focusing rather on the experience than the outcomes is not particularly novel (e.g.[6]). However, we believe that contextualizing it in a classroom setting and looking at the evaluation process from an ecological perspective bring some benefits and are key in understanding the full effectiveness of pervasive UIs on learning. Our future work include exploring a framework for evaluating the classroom experience, define what metrics could be used to evaluate which aspects and apply it in our studies.

Acknowledgements

This research is part of Dual-T project. We thank the teachers and students from the vocational schools and those who helped with our experiments.

References

[1] S. Do-Lenh, P. Jermann, S. Cuendet, G. Zufferey, P. Dillenbourg., Task Performance vs. Learning Outcomes: A Study of a Tangible User Interface in the Classroom. In Proc. of EC-TEL 2010: 78-92

[2] Piper, A.M., Hollan, J. Tabletop Displays for Small Group Study: Affordances of Paper and Digital Materials. Proc. of CHI 2009, 1227-1236.

[3] X. Cao, S. Lindley, J. Helmes, A. Sellen. Telling the whole story: Anticipation, inspiration and reputation in a field deployment of TellTable. CSCW 2010, 251-260

[4] Rogers, Y., and Ellis, J. Distributed cognition: an alternative framework for analysing and explaining collaborative working. J. of Info. Tech 9 (1994), 119–128.

[5] Nardi, B., and O'Day, V. Information Ecologies: Using Technology with Heart. MIT Press, 1999.

[6] T. Daradoumis, A. Martinez-Mones, and F. Xhafa, A layered framework for evaluating on-line collaborative learning interactions. Int. J. Hum.-Comp. Stud. 64, 7 (2006), 622-635.