Learning from Failures in Designing and Evaluating Full-Body Interaction Learning Environments

Laura Malinverni
Universitat Pompeu Fabra
C/ Roc Boronat 138, 08018
Barcelona, Spain
laura.malinverni@upf.edu

Narcis Pares
Universitat Pompeu Fabra
C/ Roc Boronat 138, 08018
Barcelona, Spain
narcis.pares@upf.edu

Abstract

Full-Body Interaction Learning Environments (FUBILEs) are interactive experiences aimed at facilitating the use of the body and physical space as sources to help in children’s construction of knowledge. The increasing importance of the field demands going beyond the initial novelty factor to better understand the specificities of this medium. In particular, little literature has debated around the definition of appropriate design and research methods. To contribute to this debate, we present two case studies from our past research, on the design and evaluation of two FUBILEs. We use these studies to set the stage for a critical reflection on methodological pitfalls and weaknesses of the approach we undertook at that moment. This post-mortem reflection offers a critical standpoint to discuss on appropriate design and research methods for FUBILEs. Finally, we discuss design guidelines and paths for research.

Author Keywords

Full-Body Interaction; Embodied Learning; Research Methods; Failures.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous
Introduction
The embodied interaction paradigm [10] has spurred researchers’ interest in exploring the pedagogical affordances of post WIMP technologies [4,9]. This research is grounded on the embodied cognition claims on the role of the body in shaping thoughts [5,12] and the importance of concrete experience in grounding knowledge [14]. Starting from this perspective, several Full-Body Interaction Learning Environments (FUBILEs) have been developed over the last years [1,2,7,18], becoming an increasingly important research field. These environments aim at taking advantage of full-body interaction with digital technologies to support the learning of topics (e.g. music [3], mathematics [8], language [18], etc.). Hence, they offer novel ways of learning by directly addressing the body and physical space as sources for knowledge construction.

Even when FUBILEs open promising possibilities, given their capacity to involve the users at different levels (i.e. sensorimotor experience, cognitive aspects and affective factors), research in this field offers a fragmented panorama of its specific benefits [21]. Potential shortcomings have been identified in the tendency of shaping their design and evaluation around methods and techniques derived from research on traditional WIMP interfaces [22,28]. Only recently an emerging debate is addressing the need for complementary and specifically suited methodological approaches [24,25].

To contribute to this debate, we present two case studies related to the design and evaluation of two multiuser FUBILEs named Archimedes [20] and PhyGame. Specifically, we shortly introduce their design, evaluation and results to set the stage for a post-mortem critical reflection on our methodological pitfalls and on the weaknesses of the proposed approach. Focusing on the gaps in the reported studies we delineate possible design guidelines and paths for future research. Lessons learned will, therefore, offer a critical standpoint to discuss design and research methods for FUBILEs.

THE CASE STUDIES
Archimedes and PhyGame are FUBILEs designed for the Interactive Slide (IS) exertion interface, a large inflatable slide augmented with interactive technology [30]. Both games were targeted to 11-12 years old children and attempted to take advantage of the physical properties of the full-body interactive platform to help explain contents related to the Physics subject syllabus. Their design aimed at exploring users’ sensorimotor experiences as mediators in the learning experience. Specifically, we focused on physics based on our hypothesis that bodily experience of forces and gravity could help children understand these concepts. In the design and evaluation process, users were involved only in the last stage as testers. The evaluation focused on assessing the impact of full-body interaction on learning gains. For this purpose, the experimental procedure was based on a between-subjects comparison of learning gains between the application designed for the IS and an equivalent version adapted for a desktop computer (DC) configuration (i.e. mouse-based control with a mouse for each user allowing a multi-user format). The learning gains were evaluated through pre- and post-multiple-choice questionnaires.

Archimedes
Archimedes (Figure 1) aimed at supporting the learning of buoyancy and Archimedes principle in 11 years old children. Given that the most significant experience on
the IS is related to "sliding down", we concentrated the definition of learning goals around phenomena linked to gravity. Buoyancy was therefore chosen as the main learning goal. After operationalizing the learning goals, we developed a simple problem-solving game based on existing syllabus material. To solve the game children have to help a fish and a cat to reach their goals. To help the fish, children must raise the level of a water pond on the right. To help the cat, children must build a bridge over the pond on the left (Figure 1). To achieve this, a set of virtual objects are made available to the children: beach-balls, rocks and wood logs. The virtual objects slide left to right over the top of the environment, appearing in a random order. Children must select objects and throw them into one of the two ponds of water. They can do this by clicking over the object in the IS version, or by double-clicking over the object in the DC version. The design of the gameplay implies that children had to understand the physical properties of the objects and strategically use them to solve the game’s challenge.

Archimedes: Experimental Setting
The game was evaluated with a population of 331 children (mean age: 11 years old ±1) recruited from different primary schools in Barcelona. During the experiment, participants were randomly assigned to either the IS or the DC condition. We administered pre- and post-test multiple-choice questionnaires to assess the learning gains in each condition. Questionnaires were developed through the consultation of curricular evaluation material for primary and secondary schools and were administered in a dedicated room and supervised by a researcher and a teacher. Pre-tests were administered 40 minutes before the experience, while post-tests were administered 10 minutes after the game. In both cases, children were allowed to play until they reached the end of the game, regardless of time taken.

Archimedes: Findings
Since data were not normally distributed, we used a Wilcoxon Signed test to compare pre- and post-test scores for IS and DC conditions. Results showed a statistically significant difference between pre-test and post-test for the IS condition (Z = -1.983; p = 0.047) with an improvement in post-test scores (positive rank = 40.09; negative rank = 37.30). Instead in the DC condition, no significant difference was found between pre- and post-test. However a modest improvement was also reported (Z = -1.621, sig = 0.105; positive rank = 37.79; negative rank = 31.37). Finally, a Mann-Whitney U test showed no significant difference in the comparison of post-test scores between the two conditions. For a summary of the results see Table 1.

Archimedes: From shortcomings to requirements
Results from Archimedes evaluation suggested that the IS could be more effective than the DC in facilitating the learning of physics-related concepts. However, the lack of significant differences between results of the two conditions motivated further exploration on how full-body interaction can be designed for learning processes. Starting from the analysis of the shortcomings in Archimedes, we defined novel requirements for the design of our second game: PhyGame. Specifically, we considered the following issues:

a) The weaknesses of the mappings between learning goals and user's sensorimotor experience: The results of Archimedes may suggest that the affordances of the sensorimotor experience were only partially suited to

<table>
<thead>
<tr>
<th>Interactive Slide (pre-post)</th>
<th>p = 0.047 (post)</th>
<th>Z = -1.983</th>
</tr>
</thead>
</table>

| Desktop Computer (pre-post) | p = 0.105 (post) | Z = -1.621 |

| Interactive Slide – Desktop Computer (post-post) | p = 0.291(IS) | Z = -1.056 |

Table 1: Summary of results of Archimedes

Figure 1: Children Playing with Archimedes on the Interactive Slide (top) general view and (bottom) detail.
support the alleged learning goals. At the same time, it may be possible that the defined content (Archimedes’ principle) was not adequate for the physical interaction provided by the IS. A possible path to explore this hypothesis was operationalized in the following requirements: 1) The necessity of identifying a topic that better fits the specificities of the sensorimotor experience that users can have on the IS; 2) The necessity of designing sensorimotor experiences capable of offering a stronger consistency between the defined learning goals and the actions performed by the users on the physical platform.

b) **Evaluating the assessment instrument**: During the administration of the post-test questionnaires, several children complained about having to do the test “again”. At the same time, issues were raised around the potential effect of the pre-test in making children “test-wise”. To address these issues, we decided to include a meta-assessment instrument in the evaluation of the second game, PhyGame. With this aim, we inserted a cluster of questions completely unrelated to the learning goals, both in the pre-test and post-test questionnaires. The purpose of these questions (meta-assessment instrument) was to evaluate children’s commitment to complete the post-test. Our hypothesis was that an improvement in those unrelated questions could indicate a “test-wise” effect, while a downgrade could indicate an effect related to a decrease in attention while doing the test.

**PhyGame**

PhyGame (Figure 2) was developed with the main objective of providing a coherent mapping between the learning goals and the sensorimotor experience. For this purpose, we analyzed the possible actions that can be performed on the IS [16] and we defined a content that could coherently accommodate them. Through this analysis, we decided to focus on the physics of motion and chose Newton’s laws of motion as the main learning goals. We then designed specific game mechanics with the chosen mappings between sensorimotor experience and learning goals (see Table 2).

**Table 2**: PhyGame: Mapping between learning goals, actions, kinesthetic experiences and game mechanics

<table>
<thead>
<tr>
<th>Learning Goals</th>
<th>Sensorimotor Experience</th>
<th>Game Mechanics</th>
</tr>
</thead>
<tbody>
<tr>
<td>An object at rest stays at rest until an outside force affects it</td>
<td>Experience how forces change your state of motion</td>
<td>Put an object in motion</td>
</tr>
<tr>
<td>An object in motion stays in motion until an outside force affects it</td>
<td>Experience the effort in trying to stop oneself and the tendency to continue in motion</td>
<td>Make an object stop / Stop your own motion</td>
</tr>
<tr>
<td>The greater a force exerted on an object, the more it accelerates</td>
<td>Experience the effect of different forces</td>
<td>Depending on the acceleration required, one or more children have to slide down</td>
</tr>
<tr>
<td>The greater the mass of an object, the more it resists acceleration.</td>
<td>Experience the need of greater effort to move objects with greater mass</td>
<td>Move objects with different masses</td>
</tr>
</tbody>
</table>

**Figure 2**: Children playing with Phygame on the Interactive Slide (top) general view and (bottom) detail

**Table 3**: Structure of the questionnaires for PhyGame

<table>
<thead>
<tr>
<th>Question Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conceptual knowledge</strong>: to assess the learning goals stated in Table 2.</td>
</tr>
<tr>
<td><strong>Procedural knowledge</strong>: to assess the transfer of knowledge in other contexts</td>
</tr>
<tr>
<td><strong>Meta-assessment instrument</strong>: unrelated questions to evaluate the reliability of the assessment procedure.</td>
</tr>
</tbody>
</table>
Table 4: PhyGame: Summary of post-test comparative results

<table>
<thead>
<tr>
<th></th>
<th>IS vs. DC Post-test Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual knowledge</td>
<td>p = 0.001 (IS) Z = -3.215</td>
</tr>
<tr>
<td>Procedural knowledge</td>
<td>p = 0.001 (IS) Z = -3.996</td>
</tr>
<tr>
<td>Meta-assessment</td>
<td>p = 0.007 (IS) Z = -2.688</td>
</tr>
</tbody>
</table>

Table 5: PhyGame: Summary of pre- and post-test results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>IS</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-post</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual knowledge</td>
<td>p = 0.036 (post) Z = 2.098</td>
<td>p = 0.334 (pre) Z = -0.967</td>
<td></td>
</tr>
<tr>
<td>Procedural knowledge</td>
<td>p = 0.473 (post) Z = -0.718</td>
<td>p = 0.004 (pre) Z = -2.847</td>
<td></td>
</tr>
<tr>
<td>Meta-assessment</td>
<td>p = 0.001 (pre) Z = 3.576</td>
<td>p = 0.001 (pre) Z = 3.919</td>
<td></td>
</tr>
</tbody>
</table>

(Condition 1), or manipulated virtual avatars with a mouse in the DC (Condition 2), depending on the experimental group they were assigned to.

PhyGame: Experimental Setting

The procedure of the experiment was similar to that of Archimedes. We recruited 209 children from 6th grade (mean age: 11) who were assigned to the IS condition or the DC condition. Pre- and post-test multiple-choice questionnaires were used to assess learning gains. As in the previous study, the tests were administered in a dedicated space and the same time interval was used between pre-test, game, and post-test. Questionnaires were structured into three main clusters (conceptual knowledge, procedural knowledge and meta-assessment instrument), as described in Table 3.

PhyGame: Findings

Since data were not normally distributed we used a Mann-Whitney U test to compare post-test scores between the IS and DC conditions. Results showed an overall significant difference in all question clusters (including the meta-assessment instrument) between the two conditions, in favor of the IS (see Table 3).

We then applied a Wilcoxon Signed test to compare pre- and post-test scores for the different question clusters for each condition (Table 5). A significant difference between pre- and post-test scores was found in the cluster of conceptual knowledge questions for the IS condition (p = 0.036, Z = -2.098) but not for DC. However, no significant learning gain was reported in procedural knowledge in the IS condition, while actually a significant downgrading occurred for DC.

Interestingly, results from the meta-assessment instrument showed a significant downgrading in both conditions. This decrease may confirm the hypothesis on to the effect of administration of pre- and post-test on children’s attitude in filling out the questionnaires. Furthermore, the significant difference in the meta-assessment questions between IS and DC highlights relevant issues related to the content validity of the questionnaires.

DISCUSSION

The results show that Archimedes favored a significant learning gain in the IS condition. However, we found no significant difference between playing on the IS or with the DC. Conversely, PhyGame produced an overall significant difference between the two conditions but learning gains were restricted to only one cluster of questions in the IS condition. The differences in these results call for a reflection on their design strategies and for a critical exam of the research approach and the used assessment instruments.

Designing FUBILEs

To design Archimedes we employed existing learning activities to develop a problem-solving game where children need to grasp a set of concepts related to buoyancy in order to solve it. However, little attention was paid to defining an appropriate sensorimotor experience. This approach led to the development of a game that was useful to help the understanding of buoyancy but that failed in taking advantage of the specificities of Full-Body Interaction, i.e. use the body to construct knowledge.

To address this issue, in Phygame, we defined the learning goals accordingly to the specificities of the IS and we focused on designing sensorimotor experiences that can provide a consistent grounding for the addressed topic. However, in the effort of properly mapping learning goals to sensorimotor experiences,
we lacked the necessary awareness and the reflexivity to design an effective game. Specifically, if we critically examine PhyGame, we can see that being able to solve the game does not necessarily imply the understanding of the defined learning goals. This caused the game to fail in making children question their own mental models of physics, since they can succeed in the game even when relying on a naive physics conception; i.e. the child can successfully apply the rules of the game, and still not need to understand Newton’s Law of motion. The difference in the employed design strategies and their shortcomings require a critical reflection on the relation between the defined learning goals, the interaction design, the game structure and the design approach.

Defining learning goals: resources and communication

Designing for FUBILEs requires a careful understanding of the available resources and of the concepts, contents and experiences that can be particularly suitable for this specific medium. In the context of Full-Body Interaction, resources such as sensorimotor experience, spatial and material configuration, and digital augmentation constitute some of the core features of this medium. Hence, in order to set appropriate learning goals, researchers should carefully understand the affordances and the potential that these resources have to express meaning [15].

In our case, we chose to employ Full-Body interaction to express concepts related to very specific physics concepts (e.g. Archimedes’ principle and Newton’s Law of motion). Nonetheless, we poorly took into account that spatiality and physicality express meaning in a way that is fundamentally different from language. This differentiation must be well understood since spatiality and physicality are intrinsically polysemic. Therefore, since more than one meaning can be associated with the same action or spatial configuration [6], making them particularly ambiguous, it becomes necessary to ask ourselves about the consistency of the learning goals that we had defined. For instance, in PhyGame, how effective can we be when trying to communicate a univocal concept (Newton’s Laws of motion) through a polysemic resource? And, consequently, which are the meanings that can best fit this kind of experiences? Dealing with these questions can be crucial for an adequate development of FUBILEs since they may provide insight in creating relevant distinctions between “what can be easily expressed through this medium”, and “what is less straightforward or even impossible”.

Interaction design and mapping

The analysis of the different mapping strategies used in PhyGame and Archimedes can provide initial guidelines for future works. Within the context of interactive environments, actions can be interpreted at least on two levels: (1) the simulated action, which is the action signified through changes in the game [23], and (2) the performed action, which is the action that is physically enacted by the user.

In Archimedes, the learning content was mainly embedded in the simulated action; e.g. when throwing a virtual rock in the water pond the water level rises more than when throwing a beach-ball. Therefore, even when the different effects of user’s actions were clearly visible in the simulation, at a practical level there was no differentiation present in the user’s performed actions; i.e. the action of the user was always “throw”. Instead, PhyGame focused on embedding the content into the performed action; e.g. two children had to slide together over one end of a catapult in order to exert sufficient “force” to send the penguin egg at a sufficient
height to make it fall into a hanging bucket. This mapping strategy, based on creating tight analogies between the learning goals and the sensorimotor experiences may be partially effective in fostering embodied learning. Thus, we believe that, in the context of FUBILEs, interaction design should go beyond usability, in the sense of designing an efficient performance of a task through our body. Instead, the interaction should evoke the ideas that the designed experience is meant to communicate and the performed actions should be contingent to the phenomenon we seek to communicate.

**Game structure, physical activity and learning process**

A common shortcoming of both Archimedes and PhyGame may lie in the design of the relation between the game structure and the physical activity. Both games were highly immersive and required children to maintain a constant attention on the events occurring on the projected surface. At the same time, both games fostered fast paced actions, which probably did not allow sufficient time for reflection and strategy formation. This drawback should make us reflect on the necessity of designing learning environments that are capable of supporting the learner in moving through different phases of knowledge construction, i.e. experimentation, reflection-in-action, reflection-on action [14,29]. Relevant workarounds can be: 1) the design of conditions to balance the physical workload with the cognitive load [13]; and 2) the modulation of the level of immersion to provide conditions that allow both engagement and reflection.

**Design approach**

Both projects were developed following a theory-informed, designer-driven approach. This means that the designers took all decisions, and users were not involved in any stage of the design process, except for its final evaluation. This methodological approach led to relevant shortcomings. On the one hand, as exposed in PhyGame, we failed to properly incorporate children’s mental models in the design. Therefore, we designed a game that children can solve even when relying only on a naïve physics conception. On the other hand, in both studies, we defined a set of actions to which we attributed a pre-given meaning, without properly considering how users actually interpreted them or used them to construct knowledge. This led to adult-centered and poorly useful design solutions.

To tackle these issues, a possible strategy would be to apply Participatory Design (PD) techniques [11] to include children’s contributions and understanding at different stages of the design process. However, despite PD methods have gained considerable importance in HCI during the last few years, specific efforts should address the research on PD approaches tailored to the specificities of Full-Body Interaction. Therefore, research needs to focus on defining techniques to design for and with the body [17,28].

**Research methods and assessment instruments**

In the two studies, the experimental procedure was based on a between subjects comparison of learning gains through pre- and post multiple-choice questionnaires. We chose this approach because of its widespread application in the evaluation of FUBILEs [21]. However, in our case, this method posed some important limitations on informing research to allow it to move forward. The main shortcomings were related to the poor reliability of the employed assessment instrument, its lack of sensitivity to properly grasp phenomena related to embodied learning, and its inadequacy to inform future design iterations.
Reliability issues
In PhyGame, we used a cluster of unrelated questions (meta-assessment instrument) to evaluate the reliability of the procedure. In both conditions, the significant downgrading result in this cluster can indicate a possible bias related to children’s attention while filling out the questionnaires. Furthermore, the significant difference between the IS and the DC conditions in the scores of the meta-assessment questions should open questions about the content validity of the proposed instruments. In other words, were we actually assessing learning gains or did we end up evaluating children’s willingness to answer the questions? Often, instruments such as multiple-choice questionnaires are suggested because of their suitability to offer quantitative data upon which statistical inferences can be easily made. However, as our study suggests, their reliability cannot be taken for granted and should be carefully scrutinized.

Sensitivity issues
Multiple-choice questionnaires have been criticized for their lack of sensitivity to grasp the reasoning processes behind the answers [27] and for their focus on cognitive skills only. These features render them poorly sensitive instruments to understand interpretative processes and to analyze how children construct meaning during the unfolding of an activity. These aspects may be particularly detrimental for research in FUBILEs since they do not offer a proper understanding of how the body and space actually contribute to shape the child’s experience and construction of knowledge. As a consequence, it becomes necessary to define adequate research methods which are capable of going beyond the current limitations and provide a deeper understanding of how meaning and learning are constructed in these environments. Relevant examples of this effort can be found in [19,24,26] where qualitative and multimodal analysis are employed to understand children’s meaning-making processes in Full-Body Interaction.

Issues in informing future design iterations
The employed instruments did not allow us to observe empirical data upon which design improvements can be justified, hence offering weak contributions to guide design iterations. This issue is particularly challenging in the context of FUBILEs since these environments are characterized by a plethora of stimuli (e.g. visual, aural, kinesthetic, proprioceptive, spatial, etc.). As a consequence, employing instruments that are sufficiently sensitive to offer a better understanding of the phenomenon can also contribute to inform design refinements.

Conclusions
We have presented the design and evaluation of two Full-body Interaction Learning Environments (FUBILEs) to set the stage for a critical reflection on appropriate design and research methods for Full-Body Interaction and learning. We, hence, offer a reflexive standpoint to look at this research field and shed light on the knowledge that researchers can construct by assuming a critical viewpoint on their own research.

Acknowledgements
We thank all children who have participated in the study. Supported by the Spanish Ministry of Economy and Competitiveness (Grant TIN2014-60599-P).

References


20. Laura Malinverni, Brenda Lopez Silva, and Narcis Pares. Impact of embodied interaction on learning processes: design and analysis of an educational application based on physical activity. *IDC ’12*


