EVALUATION OF EDUCATIONAL SOFTWARE FROM AN INTEGRAL PERSPECTIVE

Fernando J. Lage\textsuperscript{1,2}, Zulma Cataldi\textsuperscript{1,2}, Raúl Pessacq\textsuperscript{3} and Ramón García–Martínez\textsuperscript{2}

1. Laboratory of Educational Computer Science. informat@mara.fi.uba.ar
2. Laboratory of Intelligent Systems. rgm@mara.fi.uba.ar Department of Computing.
Faculty of Engineering. Universidad de Buenos Aires. ARGENTINA.
3. Faculty of Engineering, Universidad Nacional de La Plata.

ABSTRACT

This finding is part of a papers series of papers considering the most appropriate methodological aspects for the development of the educational programs. We have developed a methodology for the design of quality didactic programs has been developed and in this paper the evaluation process for the development of a program is presented.

This is not a simple task, since -as is usual with technology and technological applications- it takes into account regional or local aspects that should be keep in mind when developing the proposal. Here we basically present the process we used to carry out the evaluations of the prototypes, both external and internal, leaving the evaluation in the context for later publications.

Keywords: educational software, software evaluation

1. INTRODUCTION

In this paper we present a proposal for evaluation of a software that has been developed with a methodology that considers in their development didactic and pedagogic aspects, that is to say, that emphasizes in the educational aspects.

This work is based on a methodological development based in turn on the construction of a matrix of activities and in the definition of the cycle of life.

In this case we have used evolutionary prototypes, and we have begun to elaborate the prototypes according to the methodology proposed in previous papers [1]

This methodology was then presented with all relevant evaluations, beginning with the presentation of the prototypes developed, and then including the increments in their functionality.

The main idea was to present the prototypes to a group of students, with similar characteristics to those for who the final program would be developed.

To do so, an evaluation group of students from a graduate degree career in Technology was selected. This group was provided with the prototype along with an evaluation schedule, containing closed and open questions.
The main objective was also to detect problems in an early stage of the development in order to make any necessary changes before developing the final application.

Later on, both internal and external evaluations, were carried out, these were performed by a group of university professors.

2. DEVELOPMENT

Our starting point was that was requested to us by professors and students of graduate university course on non Computer Science, for the subject Computer. This software should take into account the parts of the computer, in particular their internal operation. The grounds for the students to request this software were that they considered that a traditional class was not enough to fulfill their expectations.

We consider that the program should be developed using Delphi 3, whereas several prototypes with increments in the functionalities should be used for its evaluation. With this idea in mind, a programmer developed the program according to a methodology developed by us [1], with the help of specialist in the contents of the thematic area.

Thus, a map of contents like those developed by Novak [2], was the starting point, in an hierarchical or tree order.

Each prototype was presented to the students of a degree career in Technology, for their evaluation and the suggested modifications were then incorporate by means of a plan.

The questions basically considered aspects of the communication interface and of the contents developed, students used an evaluation scale between 1 and 5. (5: excellent 4: very good 3: good 2: regular 1: bad or 5: very appropriate 4: appropriate enough 3: little appropriate 2: very little appropriate 1: not appropriate in all, depending on the type of question). Then, an average value was obtained.

This value allows to obtain a punctuation of the aspects kept in mind, to be able to modify those for those whose punctuation was lower than 2.5.

For every evaluation, there was an open space for the suggestions for changes or to include comments about the program or the interaction situation.

Evaluations were carried out for two prototypes (I and II) and the final product. The final product was evaluated both internally and externally.

PROTOTYPE I

To evaluate the first prototype, a group of 20 students was selected. The presentation of a screen previous design, the menu and the tree of contents were considered. Images, videotapes, and sound were not loaded yet. First, a standard Windows interface was tried, but the program was not very exposed. The questions included for prototype I are shown in Table 1.

1. Do you consider the general design of the screen appropriate?
2. Do you consider the use of those Windows/
   Buttons/
Colors/
Font type appropriate?
3. Do you consider that the program is interactive?
4. Do you consider the interface to be user-friendly?
5. Does the program give you good information about the course?
6. Do you consider the sequence of the screens to be logic?
7. Is the program easy to use?
8. Do you consider the use of the icons to be correct?
9. Do you find the use of quick keys useful?
10. Has the program interested you?
Suggestions for changes Yes - No

Table 1: Outline of the Evaluation for prototype 1

The results can be qualitatively summarized: screen design was found to be appropriate, the same as the windows and buttons used. However, the use of the colors and font type were not found to be appropriate. The interface was found to be easy to browse, and the sequence of screens was considered to be very good and easy to use in general.

There were no problems as regards interactivity, and the evaluators were interested in and curious about the second prototype, which would include more functions. Another issue to point out is the fact that many evaluators ignored the existence of the quick keys, they were not even interested in them.

As regards the open question, students were asked if they thought that any changes should be carried out in order to have suggestions for changes before passing to the next stage of the development.

Basically, their suggestions were:

- To use a larger font size in order to be able to read the program in an notebook.
- To change the colors so that there was more contrast.
- To change the mouse pointer when an object of the screen was activated.

PROTOTYPE II

According to the question presented and the suggestions carried out, the following incremental prototype was developed with the corresponding improvements. A glossary, images, some videotapes and information about each of the parts of the computer were loaded. Now the students could have a much nearer idea about how the final program would look like.

In this case, the questions about interface communication were fewer, the evaluators pointed out the necessity of adjusting aspects related to the contents and their relevancy a lot of emphasis was put on their presentation, as well as on structuring and adaptation to the necessities of the group.

1. Do you consider the selection of contents to be appropriate?
2. Would you consider the use of the finished program to be appropriate in other levels?
3. Were the changes carried out pertinent?
4. Did you want the tutorial?
5. Did it help to understand different topic?
6. Did you want sound in the videotapes?
   Suggestions for change Yes - No

Table 2: Outline of Evaluation for prototype II

It should be noted that most of the students were not interested in running the tutorial, so the conclusions was that they were not interested in using the program to replace their theoretical classes, but merely as a training oriented, support material for the course.

As regards, the suggestions made, the most interesting was perhaps the comment that the finished program would allow them "to see things that they had not imagined."

**EVALUATION OF FINAL PRODUCT**

A schedule with relevant questions to different approaches was made, taking as a starting point utility, pedagogic and didactic aspects, and technicians. The questions are shown in table 2, and they are dealt with as in the previous cases.

**UTILITY**
- Easiness of Use
- Degree of adaptation to other levels of users.

**PEDAGOGIC AND DIDACTIC**
- Level of updating
- Clarity of contents
- Browsing interface
- Level of Motivation
- Is it suitable to understand the topic?
- Is it suitable to learn the topic?

**TECHNICIANS**
- Documentation and help
- Are they adapted the resources that you need adapted?

Table 3: Outline of Evaluation of the Final Product

Finally, users were request to make suggestions by means of an open question: these suggestions could be either for the use of the program or to carry out some changes that was considered to be relevant.

In general, at this point the students did not request any changes, so it may be concluded that there was a generalized level of acceptance and agreement as regards the changes carried out from previous stages.

**INTERNAL EVALUATION**

The group that worked in the development of the program, agreed on the changes proposed by the students. The relevancy of the suggestions was also considered by our team.
It is should de also noted that a program of this nature should de regularly updated, which implies a great deal of time used for this purpose.

The proposal made by the students of using the program in parallel with theoretical explanations and using it as a practical support for the classes was also considered.

EXTERNAL EVALUATION

The program was presented to the professors of a non Computer Sciences course of studies, and they considered it to be an interesting tool to teach the topic. They were asked to fill in a questionnaire similar to the previous ones, with close and open questions with information about the finished product.

The professors considered the proposal to be interesting, and they stressed the fact sometimes students, not being computer specialist to understand how the machine works, found this very complex.

Other aspect to take into account it literacy in computer sciences; it would be desirable that students have at least some basic knowledge of the area from primary school.

SOFTWARE PROGRAMS QUALITY: AN INTERDISCIPLINARY PROBLEM:
WHAT TO MEASURE IN THE EDUCATIONAL SOFTWARE?

Table 4 was elaborated considering the fact that educational software, according to the requirements of the application and educational objectives, requires special characteristics, and also that these characteristics should be appropriate as regards quality and relevancy. The usability (easiness of use) approach (in sense of friendliness) was not included, and some sub criterions [3] were analyzed.

These can be qualified in three proposal levels: very good, good or bad. Each level has a punctuation. At the end of the evaluation the punctuation obtained will be between those shown in table 5. It can be seen there that the programs with a punctuation between 21 and 30 points will be within an acceptable quality level.

A group of 20 students was used to evaluate the program, from which the data in table 4 were gathered, and an average of 22,1 for the 20 evaluating professors.

THE QUALITY FROM THE PEDAGOGIC PERSPECTIVE

Rivera Quijano [4] considers the new educational paradigm Learner Centered Approach (LCA) or student centered learning, and considers them in many even as consumers or clients.

As Educational Technologist, he affirms: "the quality of the technological projects is measured in terms of the behavior observed at the end of the formation”. It points out that the formation program should denote measurable attributes and observational characteristics in the student, otherwise It would be impossible to determine if the program achieves the objective ones or not.”

The Dimarche Scientifique or scientific method in education, considers that in education there are two essential components: humility to question our practices and our experiences, and team sprit [5]. These are positions we strongly agree with.
<table>
<thead>
<tr>
<th>Usability (easiness &amp; friendliness)</th>
<th>Sub criterions</th>
<th>Punctuation</th>
<th>Punctuation obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>External Usability</strong> (Fenton, 1991)</td>
<td>Speed to of learning (learnability)</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Easiness of use (operability)</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Addiction level</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Internal Usability</strong> (Fenton, 1991)</td>
<td>Level of legibility (Lecturibility)</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Degree of understanding</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structures of the manuals</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use of menus, graphics and images</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error and information messages</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Online Help</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Definition to adaptation of the interface</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Criterions y sub criterions

<table>
<thead>
<tr>
<th>Punctuation</th>
<th>Evaluation of the Proposal</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>Bad</td>
<td>UNACCEPTABLE</td>
</tr>
<tr>
<td>11-20</td>
<td>Regular</td>
<td>DOUBTFUL</td>
</tr>
<tr>
<td>21-30</td>
<td>Good</td>
<td>ACCEPTABLE</td>
</tr>
</tbody>
</table>

Table 5: Punctuation table.
MORE CONSIDERATIONS ABOUT THE EVALUATION OF EDUCATIONAL PROGRAMS

The evaluation of educational programs is a process that considers the adaptation degree of this programs to the educational context. When professors receive a program, they assume that it has been analyzed and evaluated both as regards its pedagogic and didactic aspects, and also as regards technicians, who affect the quality of the product developed by following certain rules to guarantee its quality.

Basically, we carry out both internal and external software evaluations in order to detect problems that could generate changes in the product. This was done during the early stages of the development in order to reduce costs and latter efforts. These evaluations were carried out considering eventual modifications suggested by the development team and end users, and also keeping in mind both professors and students in the learning context.

When professors get a product of the educational commercial type, it means that it has gone through internal and external evaluations stages. Also, to obtain the degree of effectiveness and of efficiency of the product, an evaluation will be carried out in the context of use.

It is necessary to define certain "approaches to select a program as suitable for the needs of the professors", and the use of the words evaluation and assessment should be considered. These words are generally indistinctly used in many of the consulted works to determine if a given program fulfills the objectives, both technical and pedagogic and/or didactic.

THE INTEGRATION OF PERSPECTIVES

Rivera Quijano [4] considers the carry out an analysis of necessities, between the desired, ideal situation and the current one, defined in clear and precise terms, emphasizing the didactic design of the learning activities, that is to say, taking the project and not of the necessities as the starting point.

He considers that the most important point in the planning and the design of the educational products is the correct definition of the general and particular objectives, as well as the determination of the contents, methodology to use and evaluation to make.

By means of a didactic design -or didactic, rigorous and detailed programming- the best technology to suit to each particular problem situation can be chosen. But there are many pedagogic models, and they depend on the characteristics of each institution.

"Evaluating is observing and measuring achievements. As long as a new project allows the incorporation of new knowledge or of abilities on the part of the students, and as long as that knowledge can be observed and measured, this assures that the project suits the needs. This pedagogic design determines whether a program is educational or not" [4].

Lastly, Guitert [6] affirms that "the existence of innovative practices can be enhanced by the use of technology, but these practices are not as related with its introduction in the classroom, as the previous conception professors have from own pedagogic experience. "

4. CONCLUSIONS

The evaluations carried out allow to study several issues:

- Students are interested in standardized interface, similar to commercial ones.
- Products that allow users browse without getting lost are more reliable because users always know where they are.
- Structured contents are a key issue in order to know where learning is aimed at.
- Students are motivated to take part of the changes during the development stages of the program.
- The evaluation of the successive prototypes, allows the developers to make quick changes, according to users or potential users suggestions.
  - The program should be well documented in order to make these changes, as well as subsequent updatings, as efficiently as possible.
  - Non-computer specialist users need in addition to the software, an instructor or professor to guide them and help them interpret the processes shown with videotapes and images. Otherwise, they would not be able to keep up with the machine, even if they were dealing with a very interactive program.
  - Evaluations are a guide to keep in mind, considering a certain group of variables.

When technological applications are designed, the development methodology is one of the pillars that will allow to be able to come closer to the proposed objectives.

On the other hand, it is still very hard to measure the quality of educational programs, indirect measures are partial and to the point, but they can only be taken as that a partial indication based on which a particular proposal can be elaborated.

This is the reason why we consider long lists of questions about a particular proposition of little use, from the point of view of education, students and given context.

One interesting and at the same time practical aspect is the measurement of some indicators that allow establish an acceptable quality from approaches like those of usability and "friendliness", which can be considered both from an internal and external viewpoint.

At the moment, there is a great necessity to optimize the processes, the educational one in particular, by considering students as "clients". Bearing this in mind, not only should we get a satisfactory result. But also an approach, to an ideal profile for students to acquire after a "training".

Experience tell us that the educational program, can be carried out using the better and novel tools, but, in fact, the best indicator is obtained after students finish their training period, since it is only then that it can be seen if they acquired the skills and knowledge they were expected to acquire. These are the indicators that will give an idea of the quality or suitability of the method proposed.

5. REFERENCES


