

A children, teachers and designers as evaluators of usability of educational software

Luciana Freire^a, Marcelo Soares^b and Stephania Padovani^c

^{ab}*Department of Design, Federal University of Pernambuco, Cidade Universitária, Recife, PE 50.670-420, Brazil,*

^c*Department of Design, Federal University of Paraná, Centro, Curitiba, PR 80.060-000, Brazil -*

Abstract: This article sets out to discuss the results of evaluations of usability made by children, teachers and designers through methods targeted on their profiles and contexts of use. Based on a survey of methods for evaluating usability, a field study was conducted in which such users - stakeholders - explored the educational software called "Mundo da Criança" ("Child's World"). The focus of this study was the analysis of this system and principally of the appropriateness of methods as to the avenues they suggest such users and contexts might explore.

Keywords : information, design, educational software, usability, technology.

1. Introduction

Over the last fifteen years we have been witnessing an increasing surge of various educational software packages aimed at the teaching-learning process, at different levels and covering different target groups, especially, children [1], [2], [3].

During this process, remarkable progress in scientific-technological fields has been achieved, in which the generation of dozens of methods, techniques and tools for the concept, development and evaluation of these kinds of sites / software packages can be seen. According [4], Ergonomics researchers need to focus on aspects such as usability and navigation in graphical interfaces in their studies, bearing in mind the contribution they can offer as to understanding the structure of the information presented to users of these information systems.

From this perspective, problematic situations arise where educational software packages are selected as tools for teaching after having been evaluated by certain user profiles, even though they do not, necessarily, have an "appropriate" repertoire for evaluating them - a repertoire which might consider, for ex-

ample, the characteristics and limitations of the users in their contexts of use, like discuss by [5] and [6].

In this context, an interdisciplinary assessment was made that reflected the need to investigate forms of collaborative assessment, with, children, teachers and designers. For this, users were recruited with the profiles mentioned, taking into consideration their previous knowledge and capacities for interaction in order to see them as "originators" of project requirements and not only as "receivers" of data from a system.

This is the core of this research: to gather opinions from users and to propose solutions focused on them in order to obtain reliable assessments and more suitable methods.

2. Conceptualization and characterization of educational software

According [7], educational software programs specifically target school education activities developed in classrooms, such as interactive story programs,

*Corresponding authors lucianafreire@gmail.com
soaresmm@me.com stephania.padovani@gmail.com

encyclopedias, tutorials, dictionaries, practical exercises, and authoring and several other actions like, for example, software programs that are simulations and educational games. However, for [8], educational software is everything that might be used in contexts of teaching-learning for different kinds of content.

For the purposes of this study, it was accepted that a piece of software is educational when its focus is directed towards contexts of teaching-learning, and founded on pedagogical principles from which arise its educational themes. According to [9], there is a requirement for a piece of educational software to satisfy the following conditions which are:

- To explore creativity and interactivity, while providing the student with an active posture;
- To arouse curiosity and foster collaborative and interdisciplinary work;
- To stimulate reflection, reasoning and understanding of its concepts;
- To stress the importance of the process rather than the result obtained from it;
- To consider aspects of language (e.g., age range, environment of use);
- To consider the socio-cultural, ethical, pedagogical, ecological etc aspects;

Having knowledge of these characteristics allowed us to acquire grounds on which to assess more accurately the software used and to understand that the direct users, in general, may not recognize all of these characteristics, but all of them do influence them in the choice of software which will meet their general and specific needs. As to indirect users, they understand these characteristics as being essential to guiding them in how to use the system

3. Children, teachers and designers as Stakeholders

For [10], stakeholders are users who have the "power of decision" in a system, that is, people who deal with the system (directly or indirectly) and whose actions and views take the consolidated shape of becoming the source of practical knowledge about the interface and how it should be used.

In developing a study on evaluating educational software, it becomes necessary, beforehand, to reflect on the context in which these tools will be used and thereby to define the stakeholders of the research study.

As regards children as users of educational software, integrating them with the teacher and pieces of

educational software (technologies) can be seen in the form of a more interactive network through the direct relationship "teacher-student-technology-content", as described by [11]. As for the teacher as a user of educational software, an idea which is strongly championed, the relationship is between the use of educational software by teachers and their students, where the computer network is held to facilitate the teaching-learning network.

As to designers as actors in this scenario, they are required to have an immense ability to conduct ergonomic-educational analysis when they evaluate educational software, so that both the criteria of functionality and the criteria of learning and communication can be analyzed. As stakeholders, their activity begins in the development cycle, but when the system is already in the phase of being used, they rarely interact with the children and teachers and therefore do not have precise feedback on the product created.

4. Methods and criteria for evaluating of educational software

According to [12], the expression "to evaluate pieces of educational software," means analyzing how such software can have an educational use and help learners to build knowledge and raise their ability to participate in the reality that they are experiencing.

Researchers, like [13], [14], [15] and [16] argue how great the need is "to bring into the spotlight" the views of stakeholders which have the same "evaluation weight" as the opinions of other developers, in order to produce artifacts consistent with their aspirations. The question is: what methods could be used to assess the educational software based on this premise?

Thus, instead of our presenting at this point the survey conducted in the first phase of the study, which considered 35 general methods for evaluating usability in computerized systems - and another 17 methods, specifically directed at evaluating educational software - we consider it more pertinent to present some of the main ergonomic and pedagogical criteria underlying the procedures for evaluation which recur in the methods cataloged.

Of the 17 targeted methods in the survey conducted in the phase of setting the theoretical basis of the study, three were selected for applying in the field study, due to the scope of their criteria and their specificity in indicating the profiles of users for which

they were developed: the method put forward by [13] for children; the [17] for teachers and the [18] aimed at designers.

5. Presentation of the research: a field study

In the first phase of the field study, analytical research was carried out, based on a review of the literature which identified the various methods for evaluating usability targeted on the evaluation of educational software. Based on this survey, we selected three evaluation methods that were identified by their authors as the most appropriate for use with stakeholders having the profiles already mentioned.

In the second phase of the research, we based ourselves on the guidelines for participatory design with the focus on field studies, using research studies such as those by [19] and [5].

This study was divided into three distinct stages (with the participation of children, educators and designers). This phase aimed to apply the methods of assessment targeting the stakeholders in order to gather and discuss the results generated for each profile

Holding the 1st meeting, at which a quick presentation was made of the software and evaluation method to be used thereafter by the users. There was little interaction between the participants, but observations were made on how to use the system as were various audio-visual records.

Holding a 2nd meeting: where each participant was interviewed, individually, about the software and the method directed towards him/her. This time there was direct interaction during interviews and audio-visual records were made.

The software used in the study, called "Child's World", was selected as it was considered that, in addition to the features mentioned in topic 2 of this article, there is the fact that that the system meets the National Curriculum Parameters – [20] of the Ministry of Education, for the first grades of elementary school

6. Evaluation of the child's world software: analysis and discussion

About of results obtained from the Hanna et al method [13], the study researches 44 children, aged between 6 and 8 years old, enrolled in classrooms of 1st grade of elementary school in two private schools in the city of Recife participated in the study. In the first

meeting, the participants used the software "Child's World", having as a context their own school computer laboratory, where they were observed and filmed.

We chose to use the children's precise study environment in order to obtain a narrower approximation of the reality experienced in the children's routine, as recommended by the Hanna et al method [13]. The results were recorded on observation cards and by video recording and taking photos of the context of use, in which their teachers were present.

In the second encounter, children were interviewed using a language appropriate to their level of understanding. The script of the interview - also filmed - was developed with support from a child psychologist and was used while the child continued to navigate the educational software.

However, some children forgot (or lost) the term of consent to participate. Thus, only 30 of them could have their data included as part of the study, with the approval of the ethics and research committee of the Federal University of Pernambuco. It is worth pointing out that the term of consent was signed by the parents and guardians in the schools.

For the analysis of the usability test, conducted using the [13] method, observation cards were used that served as a script so that the researcher could observe the strict compliance of the system with the main criteria of usability, in accordance with [21]. The most significant results on the software will be presented by criterion in the following topics:

- Criterion of Consistency - The system met the criterion of consistency as the children included in this study understood that the elements of the main menu, such as the menu of "options for activities ", would always appear in the same way in all its sections.

- Criterion of Compatibility - Through the interactions observed, the compatibility of the system was clear as was respect for the repertoire of this profile of users, as the children developed a relationship of empathy with the characters to the point of their "talking" with them during the tasks.

- Criterion of feedback (Return given by the system, stating that whatever transaction had been performed and how long it would take for users to have a response to their actions.) – In general, the system used its own characters as indicators of feedback, and it was they who guided users about what was going on and how long they should wait until the activity was fully available. It was identified that these elements were understood by most of the children. In only a few cases did we notice that the children un-

derstood that there are forms of feedback, but they nevertheless preferred to ask their teacher if the response of the system was really about the action performed.

- Criterion of Prevention of errors (The system should allow for easy identification and for user errors to be corrected easily, thus avoiding greater constraints) - What could be regarded as an error was the fact that some children tried to print certain activities, by following the instructions provided by the system, but no printer had been installed. Therefore, instead of a software warning appearing on the screen to say that there was no printer, what appeared was a "standard" warning of the Windows operating system and the children opted to click on "OK" and move on, without reading what was written, thus demonstrating they understood that this warning was not part of the "Child's World" software.

- Criterion of Control of the System by the user (Possibility given to users so that they can control actions in the system, by making any necessary changes in the way it is configured) - Most children showed no interest in controlling the system beyond the options provided to them by the interface. However, some of the problems related to equipment such as headphones, ended up encouraging them to be curious about how to modify certain settings. Thus, some children tried to change the volume, not only through the volume icon of the software, but through Windows and thus "restarted" the software in search of a solution.

About of results obtained from Pedactice method [17], we researched seven teachers from the two schools where the usability tests were conducted with the children took part in this study. Their ages ranged from 24 to 37 years old. For this phase of the study, an evaluation was made of the "Child's World" software through the Pedactice method [17] questionnaire, where we asked them to monitor the context of use of the system together with the schoolchildren and to respond to the questionnaire within eight days.

After this experience, at a second meeting, the teachers could return the completed questionnaire to the researcher. Later, they responded to an interview on the software and, most importantly, the method, and presented their analyses on the appropriateness of their profile.

As to the analysis of the results of applying the Pedactice method [17], we shall follow the same structure presented in the questionnaire - which was implemented in full with the educators of these schools. However, what will be presented and discussed are only the issues referring to the third block of ques-

tions, which refers to the criteria of usability. It is appropriate to emphasize that the questionnaire does not provide a definition of these criteria. The following are therefore results by topics:

- Ease of learning and use – With regard to this criterion, the teachers' evaluation was that the system is easy to learn and easy to memorize. It was curious to note that one of them replied that the ease of the software is related to the maturity of the user, but she did not describe whether she was referring to the child user or the adult user. We conclude that this teacher could be putting into focus the child's age or the child's level of literacy, but given that the software offered spoken resources to guide use, then these difficulties could be softened, thus making it easier to use.

- Quality of scientific content - As the quality of the scientific content, the responses collected in the questionnaires indicated that the proposed content is of a good quality, but cannot be seen as the "main" form of "teaching" a given theme. The teachers responded that "Child's World" presents content which is appropriate for the age group that it is intended for, but its use must be tied to forms of education experienced in the classroom, otherwise it would not be of value on its own.

- The quality of the graphical interface - The graphic quality of the "Child's World" was approved unanimously. We identified that the answers point to complete satisfaction regarding the graphical language of the system, to the extent that, the respondents said that the screens are attractive, colorful and fun, such as to arouse the students' interests and satisfy their childish curiosity through metaphors related to the school playground, the animal kingdom and so on.

- Motivation and overall satisfaction (from the students' point of view) – As to motivation and overall satisfaction, all the teachers responded "yes", the system does have several elements of motivation and is satisfactory, taking into account their observations that the children used the software without a "hitch" throughout the study.

About of results obtained from the TICESE method [18], in the third stage of the field study, an experiment was conducted in which six designers, aged between 20 and 35 years old, from a company specializing in the development of educational software were invited to use the "Child's World" educational software and thereafter to evaluate it by applying TICESE [18]- the Technique of Inspection of Ergonomic Compliance for Educational Software, by [14]. Just as was done with the children and teachers, the

designers were asked to use and evaluate TICESE [18] and its appropriateness in relation to the software and the profile of the stakeholders, during a period of 21 days. Then, the designers were interviewed about the software and the method of evaluation - TICESE [18] – the evaluation checklist.

One of the main difficulties pointed out by the designers in applying the checklist was precisely the "detailed approach" of the manual to explaining the method and its criteria, and the fine detail of the form, when it presents several subdivisions of criteria, where, in an absolutely quantitative way, the checklist investigates the ergonomic compliance of the system. This discontent of the designers regarding the extent and redundancy of the checklist was also identified, later, from the interviews. Based on this point, instead of evaluating criterion by criterion, which would also make our analysis overlong and redundant, we have opted to present a discussion of the more meaningful results and the discrepancies that were found by using TICESE [18].

- One criterion that deserves to be discussed was that of "consistency", where it was seen that three designers indicated 100% ergonomic compliance of the software with this criterion and the other three designers indicated there was 0% compliance. As the checklist does not provide space for the descriptive recording of the reasons for these values, we infer that the disparities were due to the fact that they found it difficult to repeat the "routes" (pathways of clicks for reaching a certain activity).

- As to the criterion of "protection against errors", the designers gave discrepant percentages to the system: two of them said that there was 100% ergonomic compliance, one said there is 50% compliance and the other two gave 0% compliance. We believe that these results show that the TICESE [18] manual may well not have been clear about the definition of "error", or else, that some designers may have considered some errors as irrelevant, while others consider the errors should be prevented.

- The criterion of "evaluation of learning" also offered widely divergent results. Our reflection on this is that designers have no way of evaluating the "evaluation of learning" except if present in a context of learning, unless they can evaluate their own learning through the system.

- We believe that this was the reason for the disparity between the percentages of this criterion. We conclude that some have attributed 0% compliance as they did not have access to the end users of the software in the learning situation, just as others may have given other values, because they judged that they

themselves learned something when they used the system.

- The same situation came about with criterion of "compatibility", as two designers indicated 100% user-system compatibility, while three of the other evaluators indicated 0% and the sixth one gave 71%. Our question: Now, if the system is compatible with a user with a given profile, how could it be regarded as completely incompatible with another user who has the same profile? The analysis of the checklists completed by the designers, once again, points up inconsistent results, for it does not seem to make sense for one evaluator to give 100% to a criterion, while another gives it 0%.

Given these findings, we can state that TICESE [18] does not permit us to know the reason for the values assigned to the evaluation. This aspect bothered one of the designers who said that, simply, to give "marks", without discussing them, can lead to very "dangerous" results, as stated in his final interview about the method in itself. We therefore conclude that, by applying the checklist, we identified designers' dissatisfaction from the fact of their foreseeing that their results present divergences, although they had all stated that, as general rule, the software has good usability. Some commented that if they had to respond again to the 33 questions, their answers would probably be different from those of their first contact, and the first evaluation.

Finally, what attracted the most attention is that some designers felt it so necessary to evaluate the software qualitatively that they decided to describe, on the back of their checklist, extracts from the opinions given by their work colleagues, and one of them even presented annotations made throughout the process, given that he was not happy to have in his possession only the percentages that, he says, were not significant.

7. Conclusions and matters arising

It was found that the conduct a field study, involving three methods, is a complex task, because one cannot compare them due to their specificities, nor can one fail to collect data arising from the application of each of them. The results have appeared through the various reflections made after conducting the experiment and this leads us to believe they could have been completely different, if it had been possible to bring about the physical integration of all the participants in the same context.

As a suggestion for future studies, we emphasize that the researcher needs to understand that in dealing with the usability of educational software, the interactions occurring between users and between them and the system, can throw up many of the problems of usability which would not be seen unless an analysis of the context has been undertaken.

Such a context should be the focus of new research studies on the subject, so that designers, ergonomists (and other professionals) no longer define their strategies for research, without first understanding what alternatives they have for investigating the relationships established in the context of the system.

Finally, it is valid to remember that the original proposal of this research study was to "test tests", to implement and to evaluate the methods for evaluating usability. What was noticeable was that, much more than "testing them", there was a need to know how to conduct them. This is the challenge that remains for other further developments of this research, for, on discovering that the methods need adaptations and that they can be integrated, it is noted that it now remains to open a line of research to find out how to do this.

8. Acknowledgments

These are due to the supervisors of and participants in the research; to UFPE, Dept. of Design/ UFPE, Educandus Ltda., to the Instituto Capibaribe and to Colégio Apoio. To the Ed. Delta and to CAPES for funding this research. We share the merit of this study with these collaborators.

References

- [1] Litto, F. M. & Formiga, M. (2009). Educação a distância: o estado da arte. São Paulo, Pearson Education.
- [2] Zaharias, P. & Polymenakou, A. (2009). Developing a Usability Evaluation Method for E-Learning Applications: Beyond Functional Usability. *Proceedings of International Journal of Human-Computer Interaction*, V. 25, pp. 75-79.
- [3] Hertzum, M. (2010). Images of Usability. In: *Proceedings of International Journal of Human-Computer Interaction*, V. 26: 6, pp. 567-600
- [4] Federici S, & Borsci S. (2011). Usability Evaluation: Models, Methods, and Applications. In: JH Stone, M Blouin, editors. *International Encyclopedia of Rehabilitation*. Available at: www.cirrie.buffalo.edu/encyclopedia/en/article/ Consultado em 14/07/11
- [5] Freire, L.; Soares, M. M. ; Padovani, S. (2005). Quando o usuário é uma criança navegando um software educativo: aspectos culturais influenciando a comunicação pictórica em sistemas computacionais. In: 2o Congresso Internacional de Design da Informação, 2005, São Paulo. *Anais do 2o Congresso Internacional de Design da Informação*. São Paulo: SBDI.
- [6] Tedesco, J. C. (2004) Educação e novas tecnologias: esperança ou incerteza. São Paulo: Cortez; Brasília: UNESCO.
- [7] Cox, K. K. (2003). *Informática na educação escolar: Polêmicas do nosso tempo*. Autores associados. São Paulo.
- [8] Oliveira, C. C.; Costa, J. W.; Moreira, M. (2001). *Ambientes informatizados de aprendizagem: produção e avaliação de software educativo*. Campinas: Papirus.
- [9] Sette, S. S.; Aguiar, M. A.; Sette, J. S. A. (2004). *Formação de Professores em Informática na Educação: Um caminho para mudanças*. Coleção Informática para mudança na Educação. MEC
- [10] Preece, J.; Rogers, Y.; Sharp, H. (2005). *Design de Interação: Além da interação homem-computador*, Bookman, 2005. Trad: Viviane Possamai. Porto Alegre, Brasil.
- [11] Paloff, R. M.; Pratt, K. (2002). *Construindo comunidades de aprendizagem no ciberespaço: estratégias eficientes para salas de aulas on-line*. Porto Alegre: Artmed.
- [12] Cantarelli, E. M. P. (2005). *Software educacional*. Slides/lecture notes. Available at: www.fw.uri.br/~elisa/ Accessed on: 17/07/2005.
- [13] Hanna, L., Risdien, K.; Alexander, K. (2004). Guidelines for usability testing with children. Available at: www.acm.org or www.microsoft.com/usability/ Article, 1997, accessed on 05/12/2004.
- [14] Gamez, L. (1998). *Ergonomia Escolar e as Novas Tecnologias no Ensino: Enfoque na Avaliação de Software Educacional*. (Dissertation – Master's in Human Engineering - Universidade do Minho). Braga-PT.
- [15] Atayde, P. (2003). "Maqsei - uma Metodologia de Avaliação de Qualidade de Software Educacional Infantil" - XIV Simpósio Brasileiro de Informática na Educação - SBIE - NCE/UFRJ.
- [16] Barbosa, S. D. J. & Silva, B. S. (2010). *Interação humano computador*. Rio de Janeiro: Campus-Elsevier.
- [17] Pedactice (2000). *Ficha de síntese para avaliação de potencial pedagógico*. Captured on 20/03/2005, via the site: www.fpce.ul.pt/projectos/ Portugal.
- [18] TICESE (1998) *Manual do TICESE (Técnica de Inspeção de Conformidade Ergonômica em Software Educativo)*. Available at : www.cin.ufpe.br/~case. Accessed: 12/11/2004.
- [19] Soares, M. M. (1998). *Design Participativo: uma Experiência em Design Centrado no Usuário*. In: Soares, M. M. *Translating user needs into product design for disabled people: a study of wheelchairs*. England, 1998. (Thesis, Doctorate in Ergonomics - Loughborough University).
- [20] PCN - Parâmetros Curriculares Nacionais (2005). Available on the site of the Ministério da Educação: www.mec.gov.br/sef/ensfund/paramnac.shtm. Accessed: 12/05/2005
- [21] ISO 9241-1 (1992). *Ergonomic requirements for office work with visual display terminals (VDTs): General introduction*.