Can Users Speak for Themselves? Investigating Users Ability to Identify Their Own Interactive Breakdowns

Bernardo A.M. Mattos, Raquel L.S. Pereira, and Raquel Oliveira Prates

Department of Computer Science, UFMG
Belo Horizonte, Brazil
{bemattos,raqlara,rprates}@dcc.ufmg.br

Abstract. The Communicability Evaluation Method (CEM) is based on Semiotic Engineering HCI theory and involves observing users in a controlled environment and capturing with software the user-system interaction. The analysis involves 3 steps: (1) tagging: watching the user-system interaction video, identifying the communicative breakdowns, associating one of CEM´s utterance (from a predefined set of 13) to the breakdown; (2) interpretation: interpreting the problems that are being indicated by the tagging performed in the first step; (3) semiotic profiling: reconstructing the intended communication being conveyed by the system and the problems identified. Originally CEM requires the evaluator to perform all 3 steps. In this paper we investigate the possibility of users themselves performing the tagging step of the analysis and the costs and benefits of such a procedure. If users are able to identify and tag breakdowns they can directly communicate the problems they have experienced. Our results have shown that user tagging is possible and pointed to various directions in which it could be very useful. We present the case study performed, the results found and discuss costs and benefits of such procedure.

Keywords: Evaluation, user participation, communicability, semiotic engineering, communicability evaluation method (CEM).

1 Introduction

Empirical studies have played a major role in creating the knowledge in HCI available today [11],[1]. Researchers in the field have pointed out the need for more theoretical approaches and a theory of HCI [1] and called attention to the fact that usability is not the only important aspect to be evaluated in user interfaces [5]. They have also criticized the excessive emphasis the HCI community has given to quantitative evaluation and identified the need for more theoretically based methods [5]. This paper investigates the possibilities a theoretical based evaluation method – Communicability Evaluation Method – offers to involve users in part of the analysis of their own interaction with the system.

Semiotic Engineering theory [2] has been proposed as a theory of HCI that aims at explaining (not predicting) the phenomena involved in the designing and using of interactive systems. It is based on Semiotics and takes a communicative approach to

M. Kurosu (Ed.): Human-Computer Interaction, Part I, HCII 2014, LNCS 8510, pp. 521–532, 2014.

HCI. Semiotic Engineering theory perceives an interactive system as a communicative act from designers to users. In the message being sent (i.e. the system) designers convey to users who the system is meant to, what problems it can solve and how to interact with it to solve them. This message is an indirect message that users understand as they interact with the system. In this theoretical framework a good interface is one that conveys to users, efficiently and effectively, their underlying design intent and interactive principles, in other words one that has a high **communicability** [9].

The Communicability Evaluation Method (CEM) [9]. [2], [3] has been proposed within the Semiotic Engineering framework and aims at evaluating the system's communicability. To do so, the method identifies communicability breakdowns – that is, problems that have taken place in the user-system interaction. Each breakdown is associated to an utterance that represents what users potentially may have said to the designers of the system when facing the breakdown. It is as if evaluators "*put words into the users' mouth*". Based on the breakdowns identified and their tagging, the evaluator proceeds to describing the communicability problems of the system, and contrasting them to designers' original intent.

Communicability Evaluation Method has been applied to a number of different contexts [12] [3] [8]. Other works have shown how the interactive breakdowns changed as users interacted with the system over time [10] and how it compares to other evaluation methods [14]. The participation of users in helping evaluators understand their actions has been the focus of all verbal protocol methods [4], [11]. Other researches have investigated in which moment of the evaluation users can best support evaluators' interpretation [13], [7], [6]. Although user participation in supporting evaluation analysis has been explored for other methods, and in spite of CEM authors having raised the hypothesis of the possibility of users performing the tagging step [9], to the best of our knowledge this investigation has not yet been performed. Therefore, in this paper we investigate the possibility of having not evaluators put words in users' mouth, but rather users themselves apply the tagging step of the method. Performing this step requires users to identify the communicative breakdowns they have experienced and to associate it to the tag that describes it. A communicative breakdown occurs when the user does not understand the communication intended by the designer through the system, which can hamper or even preclude the use of the system. If users are able to identify and tag breakdowns they can communicate directly the problems they have experienced.

The first step in investigating the possibility for users to identify breakdowns and associate to them an utterance from the CEM set was to perform regular communicability tests using CEM. As soon as the test was completed the procedure continued and the evaluators guided the participant through the tagging step. The user tagging was conducted in the following steps: (1) brief explanation of what a breakdown is; (2) watching the interaction movie of the test and annotating it with breakdowns identified by the user; (3) brief explanation of CEM's predefined set of utterances used to describe breakdowns; (4) reviewing the breakdowns identified and associating to each one an utterance from CEM's set; (5) a post-tagging interview about their experience.

Eight people participated in the user-tagging investigation. All of them were undergrad or master students and were experienced in the use of technology, but had never taken any courses in or worked with HCI or interface evaluation. Our main findings were that users were able to identify breakdowns and tag them. They felt comfortable with the procedure and with the set of utterances and felt that they could perform it by themselves.

In the next section we briefly present the Communicability Evaluation Method. Then, we describe the user-tagging experiment performed. In the following section we present and discuss the results from the user-tagging experiment. Finally we conclude with our final remarks and possible future directions in this research.

2 The Communicability Evaluation Method

As described, the Communicability Evaluation Method (CEM) is based on Semiotic Engineering and aims at evaluating qualitatively the communicability property of an interactive system. The CEM involves evaluators observing users interacting with a system in a controlled environment. The preparation and application steps of the CEM are very similar to other user tests in controlled environments [12], [2]. A few aspects that are important to highlight are (1) user interaction with system must be recorded by use of an interaction recording software; (2) two evaluators are recommended: one to conduct the test and another one to take notes during the test; (3) an interview with the user after the test is strongly recommended. Once the test has been applied, the evaluator proceeds to the analysis of the data collected.

The analysis of the data requires three steps: tagging, interpretation and semiotic profiling [9], [2], [3]. In the **tagging** step the evaluator watches the movies of user-system interaction and identifies communicative breakdowns that have taken place. To each communicative breakdown the evaluator associates one utterance from a predefined set of 13 utterances (see Table 1) that best describes it. Utterances are stereotypical expressions users could potentially express (to the designer) when having difficulties interacting with the system, such as *Where is…*<specific function>? In this step the evaluator "*puts*" words in the participant's mouth [9].

Once breakdowns have been identified and tagged, the next step in the analysis is interpretation step, in which the evaluator assigns meaning to the tagging done in the previous step and evaluates whether there are or not communicability problems with the system. Finally, in the last step, the semiotic profiling step, the evaluator reconstructs the global message being sent from designer to users, and then contrasts this global intended message with the problems identified.

The interpretation and semiotic profiling steps require expertise in CEM and Semiotic Engineering theory, which is not necessarily the case for the tagging step. Thus, authors of the method have raised the hypothesis that users may be able to do the tagging themselves and that this could provide a more precise tagging, since users would utter the expressions themselves [9]. Although the utterances are based on natural expressions, their use is directed by a specific set of symptoms which limit their natural use. The goal in this paper is to investigate whether users could identify

their own communicative breakdowns and associate the expected utterance from the CEM's predefined set of tags, in other words, if tagging could be used to provide a direct communication between users and designers (or evaluators). In the next section we explain the method adopted and the case study performed.

Table 1. Brief description of CEM set of utterances

Utterance	Description
What's this?	Occurs when the user does not know the meaning of an interface element.
Where is it?	The user knows what he would like to do, but demonstrates difficulty in locating it.
What now?	It applies when a user does not know what to do next and thus searches for the next step.
What happened?	The user does not perceive or is unable to assign meaning to the function's outcome, or the system does not present any feedback.
Oops!	The user performs some action to achieve a specific state of affairs, but the outcome is not the expected one. The user then immediately undoes the action.
I can't do this way	The user goes into a path of interaction composed of many steps and decides to abandon the path.
Why doesn't it?	The user insists on an action path that does not produce the expected outcome.
Where am I?	User does not realize the context he is in and tries to perform actions that only make sense in another context of the system.
Looks fine to me…	The user is convinced he has achieved his goal, but in fact he has not.
I give up.	The user runs out of resources (time, patience or motivation) and interrupts a task performance.
I can do otherwise.	The user is unaware of some preferential intended affordance present in the interface and manages to achieve his goal some other way.
Thanks, but, no, thanks	The user understands some preferential intended affordance present in the application's interface but decides to do it in another way.
Help!	The user explicitly asks for help.

3 User Tagging Experiment

The first step in the experiment was to perform a system evaluation using CEM. The user tagging experiment immediately followed this system evaluation in the same controlled environment. The reason for this was to allow users to perform the tagging while they still had their interaction with the system fresh in their memory, not adding an issue regarding the time spent between the test and their analysis of it. The interaction was recorded and the test videotaped. Two evaluators were required: one to guide the user through the experiment, and another one to observe and take notes (behind a one-way mirror). The experiment was organized in 5 steps:

1. **Brief presentation:** Once participants completed a regular CEM test, the evaluators explained to them what the next steps in the process were. At this moment the evaluator explained briefly to them Semiotic Engineering perspective of an interface as a communicative act, the communicability property and the definition of a communicative breakdown.
2. **Identifying communicative breakdowns:** In this step, participants watched the movie of their own interaction, and identified communicative breakdowns they had experienced. The description of each breakdown given by the participant was annotated in the movie by the evaluator. If the participant became too quiet the evaluator encouraged his/her participation by asking questions about what was going on at a point of the movie.
3. **Presentation of CEM utterances:** After the breakdowns had been identified, the evaluator explained to participants the tagging process and the utterances. To do so, a slide show was used in which each utterance was briefly explained, its symptoms presented and an example made available. Two short movies illustrating an example of a breakdown associated to an utterance were shown to participants, and they were asked whether they would like to see other ones. Participants at this point were given a list of the utterances they could refer to during the following step (similar to the one shown on Table 1).
4. **Associating utterances to breakdowns:** In this step, evaluators guided users through the breakdowns annotated in the movie and then users associated an utterance to each breakdown.
5. **Post-tagging interview:** In this step the evaluator conducted a semi-structured interview with the participant about the tagging.

In the case study conducted the software chosen was a personal organizer (Student Life – http://www.tesorosoft.com/studentlife.htm) for college students. The software handles everything related to the life of a typical student, such as classes, homework and tests, contacts, calendar, reminders and a degree tracker. The software was chosen because it did not require any specific domain knowledge, college students were available as volunteers and the in a previous HCI class students had evaluated it and reported having found many communicative breakdowns in interacting with it.

The requirements to select volunteers were: (1) be a college student (undergrad or master level); (2) understand English well (the system´s interface was in English); (3) not have any experience with HCI and interface evaluation; (4) not have used the system before; (5) be experienced in the use of technology. One pilot test was done to adjust the material and how the test would be conducted. The study was conducted with eight volunteers (5 men and 3 women), ages ranging from 20 to 28 years old. Students were distributed into different courses: 4 in computer science, 2 in information science and 2 in engineering. Each test lasted about 1:40 hours. The test of the personal organizer was comprised by 3 different tasks: editing an instructor's information, entering a new class in an existing semester, entering a new semester. All users were able to complete all three tasks. After they completed the tasks they were interviewed about the system (as they would in a regular application of CEM).

The system test lasted around 15 minutes. Once it was over, the tagging step of the analysis was performed as described in the beginning of this section. The explanation required to apply the method was brief – step 1 (brief explanation) took about 3 minutes and step 3 (presentation of utterances) lasted around 10 minutes. At this step none of the participants asked to see any other examples besides the two that were shown. At the utterance association step (step 4) some users (5 of them) added a few (varying from 1 to 3) new breakdowns to their interaction they had not identified in step 2. Finally the post-tagging interview script included asking the participants what they thought about tagging their own interaction; if it had changed their view of the system; whether they felt the explanation on the utterances had been enough for them to perform the tagging; if they felt they had done a good job tagging their interaction; whether they had felt the need for utterances that were not available; what were the difficulties they had had during the tagging and whether they felt they could perform the tagging by themselves.

At the end of the experiment the participants' tagging were analyzed by the evaluators. The analysis was mainly qualitative and aimed at identifying whether users were able or not to identify breakdowns and associate utterances to them and what were the main challenges involved in these tasks.

4 Results and Discussion

The experiment intended to provide indicators to two main questions: (1) Can users identify their own breakdowns?; (2) Can users associate utterances from CEM's predefined set to breakdowns? Our analysis has shown that most of the time users are able to perform the tagging step, but they faced some challenges in doing so.

4.1 Identifying Breakdowns

To verify whether users would be able to identify their own breakdowns, evaluators analyzed the breakdowns they had identified and contrasted them with their own identification of user-system breakdowns. Five out of eight participants were able to identify most (over 55%) of the breakdowns they experienced. One participant correctly identified 85% of her breakdowns. An analysis was done to understand if there were specific situations or breakdowns that users missed. First of all, we checked whether users who experienced more breakdowns missed more in the identification, but there was no correlation between the number of breakdowns found and the number of breakdowns experienced.

The next step in the analysis was to verify what breakdowns had been missed. Note that two of the breakdowns considered in CEM users were not expected to identify (*Looks fine to me*. and *I can do otherwise*.), since they represent some aspect the user did not perceive at the interface. Thus, in order to identify them, at the tagging step users would have to perceive something about the system they had missed minutes before. Thirteen (13)% of the breakdowns that were not identified corresponded to these two types of breakdown. Nonetheless, some users were able to identify a few of

these breakdowns. It usually was possible when during the interaction users learned more about the system and understood something they had missed before.

There were 4 instances of *Looks fine to me* type breakdown that were correctly identified (i.e. 31% of the total number of breakdowns). One of them was due to filling in a numeric field with the wrong information. The participant realized the mistake in the analysis of his own interaction movie. One other participant did not understand what a `Set` button in the interface was for, so she clicked on it after each action unnecessarily, introducing a spurious action in her sequence of actions. Later on in a following task she realized what it was meant for, and during the analysis was able to identify the spurious action in two different moments. The last one was not sure which field she was meant to fill in, so she chose the wrong one.

Out of the two *I can do otherwise* breakdowns that took place, one of the participants was able to identify one of them. The user did not find a piece of information (available in the system) referred to in the task scenario and created a new one. Later on, during another task he found the information and realized he did not have to create it, identifying the breakdown. Although this situation might have been caused by the test situation (in a real context he would have created the information the task referred to at some previous time), it still indicated that the difficulty the user had in finding the information led him to an alternative action.

It was interesting to notice that a great number of *Where is it? What is this?* and *Oops!* types of breakdowns were not identified (together corresponded to 67% of the total of breakdowns not identified). These are usually considered easy to identify by evaluators who are beginners because their symptoms are easily identified (as shown in Table 1). The reason for this came up during the step in which they identified the breakdowns. When the evaluator noticed the symptom and the participant did not identify it as a breakdown, he asked whether they had had any problems at that point. They usually responded that "*No, I was just looking for <something>*" or "*No, I was just checking what that was.*". This means that even though users had to change their attention focus from the task to the interface they did not consider it a problem. This could be either because this was the first time they were using the system, and thus expected to look for functions and ask what interface elements meant as a way to learn the system, or because once they found what they were looking for, then later on they minimized the problem (effect noticed in Post Think Aloud experiments [13]). The former situation was verified in an experiment over time. A study of how breakdowns change over time has shown that these types of breakdowns decrease as users become more familiar with the system [10]. Thus, a future study could verify if once they learned the system, they would still consider these situations natural or would perceive them as breakdowns.

Another situation that happened that is worth noticing was when a participant experienced a breakdown and in trying to solve it went through other breakdown situations. For instance the participant was looking for a specific action, so he waited for tool tips on interface elements (*What´s this?*) and opened options and closed immediately as they realized it was not what they were looking for (*Oops!*). A couple of times the participant identified the higher level breakdown (typically *Where is it?* or *What now?*) but did not identify the smaller ones as breakdowns, probably because they were efforts to solve the higher level one.

The other types of breakdowns that were not identified represented a smaller percentage of the situations, but two cases are interesting to comment. In the first one a participant did not identify a "*I give up*" breakdown, which may seem odd. The participant tried to set the initial and end date of the semester being created, but was unable to do so. She then abandoned the task to do something else. At that moment, it configures a breakdown since she gives up on the action of setting the date. However, in her analysis she did not define it as a breakdown because she said she had not given up yet, she had decided to do another action and go back later to try again. She actually did so, and then when she was still unable to set the dates she identified the breakdown. In this case, what was taken into account by the user were not her actions (she did give up at that point) but rather her global intention to try it again later.

The other situation was that 3 participants when finished a task were not able to tell whether they had succeeded or not. They then went on to try and verify whether what they had intended to do had actually been done. Although, these actions indicate a breakdown – system did not communicate back to them in a way they understood what had been done (symptoms of a *What happened?* breakdown) – even when they commented what their intention was at that point, they did not consider it a breakdown. We believe that one possible reason for it was that they all realized they had achieved their goal, so nothing had gone wrong in the task, so they did not consider that a breakdown.

None of the users experienced breakdowns related to the utterances *I can´t do it this way.* , *Thanks, but no thanks!* or *Help!*.

4.2 Associating Utterances to Breakdowns

Another goal was to verify whether users, having identified breakdowns, could associate the expected utterance to it. The analysis on this step of the experiment yields indicators on how natural the expressions are to users. One potential difficulty is that even though the utterances are natural the situations they apply to are limited or determined by the symptoms. In tagging the breakdowns five out of eight users chose the utterance correctly over 70% of the times. The other three chose correctly over 49% of the times. Two out of these three also had more difficulties in identifying breakdowns. An analysis of the mistaken tags was done to investigate whether there was a pattern in changing one for the other, and also if mistakes took place in a specific context. However, this analysis showed that there were no patterns in making a wrong choice or in the context in which it was made.

Nonetheless, the analysis indicated that most of the choices of wrong tags can be explained by two different situations. The first one, as we had foreseen, was when in a natural context (but not if symptoms described were considered) both utterances would apply. The second situation that led users to choose the wrong tag was when users identified the breakdown as part of a higher level goal or context and associated the tag to this higher level, as opposed to the symptoms of the breakdown. For instance, one participant opened a dialog realized it is not what she was looking for and immediately closed it – the symptoms characterize the breakdown as an *Oops!*. However, she commented that at that point she could not find what she wanted and was lost in the system, so she chose the tag *Where am I?*

It is important to notice that explanation on utterances given to participants was very brief (around 10 minutes). The reason for this short explanation was to investigate the execution cost with a minimal learning cost on the users' side. However, it seems that with a more detailed explanation or even a short training (users would tag some examples first) users could perform even better at the tagging step.

4.3 Evaluator's Accuracy

This study has also provided some indicators about how accurately evaluators can identify problems experienced by users, and correctly tag it – that is, "put the *correct* words in the users mouth". It was interesting to notice that most of the time evaluators correctly identified breakdowns (99%) and tagged it (95%).

There were only two breakdowns that evaluators had not considered as a breakdown and participants identified as such. In the first case, the participant interacted with a dropdown list to select the year of interest. He chose one year, changed to another and went back to the first. Evaluators considered he had understood the interface element and was just exploring the different years. However, the participant identified it as a breakdown and later tagged it with *Oops!* because he had thought the choice of year would change the options in the semester list. When he made the choice and the expected change did not take place, he undid his action by returning the year to its previous value. The other breakdown was when another participant put her mouse over an interface element expecting a tool tip. However, she did it quickly and no tool tip was available and she went on to do something else. The evaluators did not perceive that the cursor had been intentionally moved to the interface element in search of a tool tip (symptoms of a *What´s this?*, which the user correctly tagged) and did not identify the action as a breakdown. Although these two breakdowns are not very serious ones, it shows that sometimes evaluators may miss symptoms when the users' intention is not clear from their interaction with the system. In these situations the participation of the user in the tagging step can be crucial.

One other aspect the case study indicated was that although symptoms limit the application range of the tags minimizing some ambiguities, some still take place. In some situations evaluators perceive the ambiguity and if that is done during the test they have the chance at the post-test interview to understand what happened. However, in other situations the evaluator does not even perceive it and the user is actually the only who could tag it correctly. In this case study both situations were observed.

In this case study instances of the first situation were not solved during the test because participants were expected to continue with the tagging step so these kinds of issues were not raised at the post-test interview. For instance, one participant had filled out some pieces of information and clicked on the button labeled `Set`. He then stopped for a moment and continued his interaction. The `Set` button was not associated to the pieces of information he had filled in. Since he then continued evaluators took that he believed he had to click on that button, and tagged that as

Looks fine to me. However, since it was not clear what had led him to click on the button, during the interview evaluators could ask him why he had done so to eliminate any misunderstanding on their part. In fact at the tagging step the user identified the breakdown and tagged it as an *Oops!*. He explained that as soon as he clicked on the button, he realized he did not have to do so. However, since the `Set` button had not generated any valid action there was nothing to undo, so he just continued the interaction.

There was one situation that the evaluators had no idea the tagging was incorrect until the point the user explained it and tagged it himself. This happened when one participant wandered with the cursor on the screen without actually doing anything, which led evaluators to tag it with a *What now?*. However, the participant explained that he knew what to do next and was actually trying to identify which element in that dialog represented the function he wanted, so he correctly tagged it as a *Where is?*.

The case study shows that the evaluators tagging was very accurate (95%), and although evaluators had experience with CEM, but were not experts yet. Of the few mistakes that did happen, some could have been solved at the post-test interview. However, the system evaluated was a very simple general purpose one and the tasks were simple as well. Further studies considering more complex or domain specific systems that require more complex tasks should be pursued to investigate how it would affect evaluators' accuracy.

4.4 Participant's Comments about Tagging

In the experiment users were interviewed twice, once about the system after the CEM regular test, and another about the tagging after the user-tagging experiment. None of the users changed their opinion about the system after the tagging, however, they were much more critical about the interface and also volunteered many suggestions or comments for alternative designs.

Some of the users clearly were enthusiastic about the tagging, and even when evaluator was explaining the utterances they would make comments about them. For instance, when the evaluator explained the *Where is?* tag a participant said[1] "*Oh! I had a couple of Where is?*" and then when he explained the *What´s this?* tag the participant said "*Oh when I checked out that button it was a What´s this? then*". When asked about the tagging, four participants said they had difficulties in making a distinction between a few of the utterances, and one of them said that he thought that the time they had do learn them was too short. In spite of the difficulties, seven out of eight participants believed they could perform the tagging by themselves. Only one participant said that she felt that some utterances were very simple such as *Where is it?*, but others (she mentioned *Oops!, I can´t do it this way*., and *Why doesn´t it?*) she believed she would have difficulties in using. One participant said: "*I thought it was interesting because you can identify exactly where the problem is, you can distinguish among the many problems in the system.*". Another one commented that: "*This*

[1] All participants' comments were translated from users' first language to English by the authors.

tagging part I thought was really cool because the expressions are the same as the doubts we have. We really identify ourselves!"

They all felt that their participation in the interpretation process was very useful to the evaluation. Some commented that otherwise how could the evaluator have known what was going on. Most users felt that the utterance set was comprehensive. Only one user suggested the inclusion of an utterance, but when he explained what it meant evaluators identified it as being the same description as the *Oops!* tag. One other user commented that it would be nice to have some positive utterances as well to be able to tell when the system had been clear or a goal easy to achieve.

5 Final Remarks and Next Steps

In this paper we have presented a case study that investigates the possibility of users performing the first step in the CEM analysis of their interaction with the system – the tagging step. Although investigation on user participation in supporting evaluators has been done in other contexts [13], [7], [6], it had not yet been done for the CEM. The relevance of investigating this for the CEM is that this research increases the knowledge regarding CEM and its applicability. This is useful not only for Semiotic Engineering Theory research but also for HCI since this community has pointed out the need for more theoretical based methods for HCI [1], [5]. Furthermore, the breakdown identification and tagging is based on natural users' expressions, and could be an appropriate language for user to designer communication about the system.

This study has shown that the user participation during tagging step could provide for a more precise identification of problems experienced. Although in cases like the one investigated (simple general purpose system with simple tasks) more precision may not be necessary since evaluators tagging was also accurate, in others, in which tagging requires a domain specific knowledge (e.g. educational systems) it could have a relevant contribution in eliminating ambiguities.

One other result noticed after having performed the tagging step was that users discussed more aspects of the system's interface and alternative design options. Thus, an interesting future investigation would be to use CEM (or the user tagging step) as a method to involve users in the design process. In this situation positive utterances, as suggested by one of the participants, could be useful to allow users to express the aspects of the system they thought had high communicability.

Participants' interviews about the tagging have shown that users have identified themselves with the tags and felt comfortable using them to indicate the problems they had experienced. In spite of the brief explanation on tagging they received, most of them were able to tag over 70% of breakdowns correctly. Thus, further studies are needed to evaluate if users would be able to tag breakdowns by themselves.

As for the breakdowns users missed, they could probably be solved with a longer explanation or training. In that case benefits of user investment must be weighed, since the time invested and thus cost of the method would increase. At any rate, some of the breakdown misses per se may be an interesting information for evaluators. The fact that users do not consider a breakdown as such could provide indicators on the

higher level task they were focusing on, as well as the breakdown's severity and priorities for redesign. In the case study presented the focus was on the tagging step, so the impact of users tagging on the other analysis step (interpretation and semiotic profiling) was not investigated. Nonetheless, such a case study could provide more indicators on the potential benefits of user tagging.

Finally, this paper was a first step in the investigation of the cost and benefit of user tagging in CEM. It has shown that user tagging is possible and pointed to various directions in which it could be very useful. Furthermore, the challenges faced by users are potentially challenges for evaluators who are applying the method for the first time. Thus, pointing out the challenges and potential reasons for them is useful for evaluators who want to apply the method, as well as for educators who teach them.

Acknowledgements. We thank all the participants of the tests and also Fapemig and INWeb (MCT/CNPq/ grant 57.3871/2008-6) for partially supporting this work.

References

1. Carroll, J.M.: HCI models, theories, and frameworks: toward a multidisciplinary science. Morgan Kaufmann, Menlo Park (2003)
2. de Souza, C.S.: The semiotic engineering of human-computer interaction. MIT Press, Cambridge (2005)
3. de Souza, C.S., Leitão, C.F.: Semiotic engineering methods for scientific research in HCI. Morgan & Claypool, Princeton (2009)
4. Ericsson, K.A., Simon, H.A.: Protocol analysis: Verbal Reports as Data. MIT Press, Cambridge (1993)
5. Greenberg, S., Buxton, B.: Usability evaluation considered harmful (some of the time). In: Proceedings of ACM CHI 2008, Italy, pp. 111–120 (2008)
6. Frøkjær, E., Hornbæk, K.: Cooperative usability testing: complementing usability tests with user-supported interpretation sessions. In: Proc. of CHI 2005, pp. 1383–1386. ACM (2005)
7. Guan, Z., Lee, S., Cuddihy, E., Ramey, J.: The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: CHI 2006, pp. 1253–1262. ACM (2006)
8. Mattos, B., Prates, R.: An overview of the communicability evaluation method for collaborative systems. In: IADIS International Conference WWW/Internet, pp. 129–136 (2011)
9. Prates, R.O., de Souza, C.S., Barbosa, S.D.J.: A method for evaluating the communicability of user interfaces. ACM Interactions 7(1), 31–38 (2000)
10. Prates, R.O., Barbosa, S., de Souza, C.S.: A case study for evaluating interface design through communicability. In: Proceedings of DIS 2000, pp. 308–317. ACM (2000)
11. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-computer interaction. Addison-Wesley, Reading (1994)
12. Sharp, H., Rogers, Y., Preece, J.: Interaction Design: Beyond Human Computer Interaction. Wiley (2011), Case Study 14.3 Communicability Evaluation, `http://www.id-book.com/casestudy_14-3.php` (last visit Januaury 2014)
13. Teague, R., et al.: Concurrent vs. Post-Task Usability Test Ratings. In: Proc. CHI 2001, pp. 289–290. ACM Press (2001)
14. Salgado, L.C.C., Bim, S.A., de Souza, C.S.: Comparação entre os métodos de avaliação de base cognitiva e semiótica. In: Proceedings of IHC 2006. SBC, pp. 158–167 (2006)