8
User evaluation study

A test with 12 participants was made to see how UISKEI compares to other two prototyping techniques: paper prototyping and prototyping using Balsamiq. The goal was to evaluate the difficulty not only in drawing a user interface, but mainly in defining its interactive behavior. The evaluator asked the participants to create and simulate a prototype of a login screen using the three different tools, evolving the prototype through three cycles of iteration:

- 1st cycle → Create the login screen prototype with a single checkbox, which may lead to two different outcomes during simulation.
- 2nd cycle → Add another checkbox to the previous prototype, increasing the number of possible outcomes to four.
- 3rd cycle → Discuss about how much effort is needed to add yet another checkbox, raising the number of possible outcomes to eight.

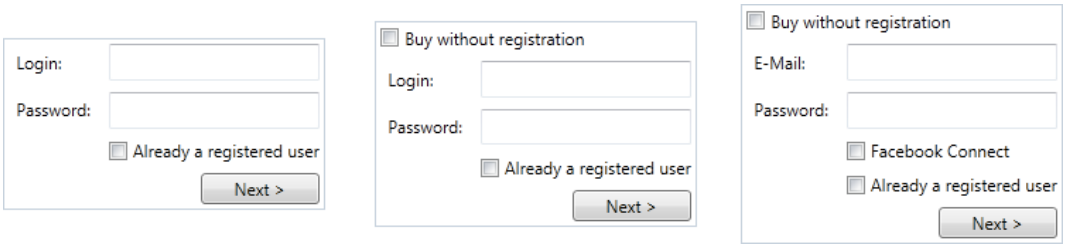

Figure 29: The evolution of login screens through the cycles.

The hypothesis of the test is that UISKEI should have a poor performance in the first cycle, since its language to add elements and ECAs is unknown to most of the participants, but then it would improve in later cycles, as participants learn the language and benefit from having only one mock-up with coded behavior.

We expect that Balsamiq would perform well in the beginning, due to its extensive collection of widgets and the well-known drag-and-drop paradigm, but the need to duplicate screens to show the behavior using only navigation would make it harder to use as complexity increases.

According to (Hammond T. A., 2009), “Pen and paper provide a freedom of interaction that is still preferred to a computer automated design tool, even though users want the sophistication of analysis and simulation capabilities of a computer-understood diagram”. So we expect the same pattern for paper, since the addition of an element is extremely easy by drawing, but as the prototype evolves and becomes more complex, some changes may require the participant to redraw the prototype, eventually making it very difficult to perform the simulation on-the-fly.

In the following section we present the evaluation method used. Section 8.2 presents and analyzes some results, while Section 8.3 shows some participants' opinions expressed during the evaluation.

8.1 Evaluation method

The experiment followed a within-group design, comparing the performances of the same participants on all three tools (paper, Balsamiq and UISKEI), thus requiring a smaller sample than if each participant was only exposed to a single tool. This had the negative effect, however, of them learning from the experience of previous tools and getting better in completing the tasks (Lazar, Feng, & Hochheiser, 2010, p. 48). To avoid the learning effect, we randomized the order in which each participant used the tools, so the learning effect of a user is offset by another one. Consequently, the entire data set is not significantly biased by the learning effect (Lazar, Feng, & Hochheiser, 2010, p. 52).

Before using each tool in the first cycle, videos were shown to introduce the tools and to explain how to add elements and define the behavior. A “cheat sheet” with the main language used in UISKEI (containing Figure 18 and Table 3) was also provided. After using each tool in cycles 1 and 2, as well as after the discussion of the 3rd cycle, the participant was asked to answer a short questionnaire, containing 10 grading questions, as follows:

1. How easy was it to understand what needed to be done to add the interface elements?
   (1: very hard, 5: very easy)

2. Once you knew what to do, the effort needed to create elements was:
   (1: very high, 5: very easy)
3. How different were the resulting interface and what you had in mind?
   (1: very different, 5: very easy)
4. In general, how did you like the way to create elements?
   (1: hated it, 5: loved it)
5. How easy was it to plan what needed to be done to create the required behavior?
   (1: very hard, 5: very easy)
6. How efficient was the definition of the planned behavior?
   (1: very inefficient, 5: very efficient)
7. How easy was it to create the new behavior?
   (1: very hard, 5: very easy)
8. The definition of the new behavior required an effort:
   (1: very high, 5: very easy)
9. In general, how did you like the way to define behaviors?
   (1: hated it, 5: loved it)
10. Once a behavior is defined, what do you think about its representation?
    (1: hard to understand, 5: easy to understand)

All questions were formulated in a way that higher scores meant better results. The first four questions are related to the creation of the prototype interface, whilst the remaining questions are related to its behavior.

After doing the tasks, participants also went through a quick interview, questioning them about which tool they would use in the situations described below, and why:

1. In an early development stage, while exploring the idea space, where different solutions are considered and constantly changed, focusing only in the interface.
2. When the idea is clearer, a solution was chosen and a single prototype needs to be built, focusing still only in the interface.
3. Considering that they needed to define the behavior of the chosen solution.

The complete test script can be found in “Appendix B: Evaluation study script“.

8.2
Evaluation results

Overall, the hypothesis of an increase in the scores given to UISKEI as the cycles progress was confirmed by the test. After the end of the second cycle, UISKEI only received scores lower than the other tools in questions 5 and 7, showing that the logic behind defining ECAs is not easily grasped. By the end of the third cycle, UISKEI’s average scores in all questions were greater than the other tools and all greater than 4.0. The average score of each tool in each question can be seen in Figure 30. The complete test results can be seen in “Appendix B: Complete test results“.

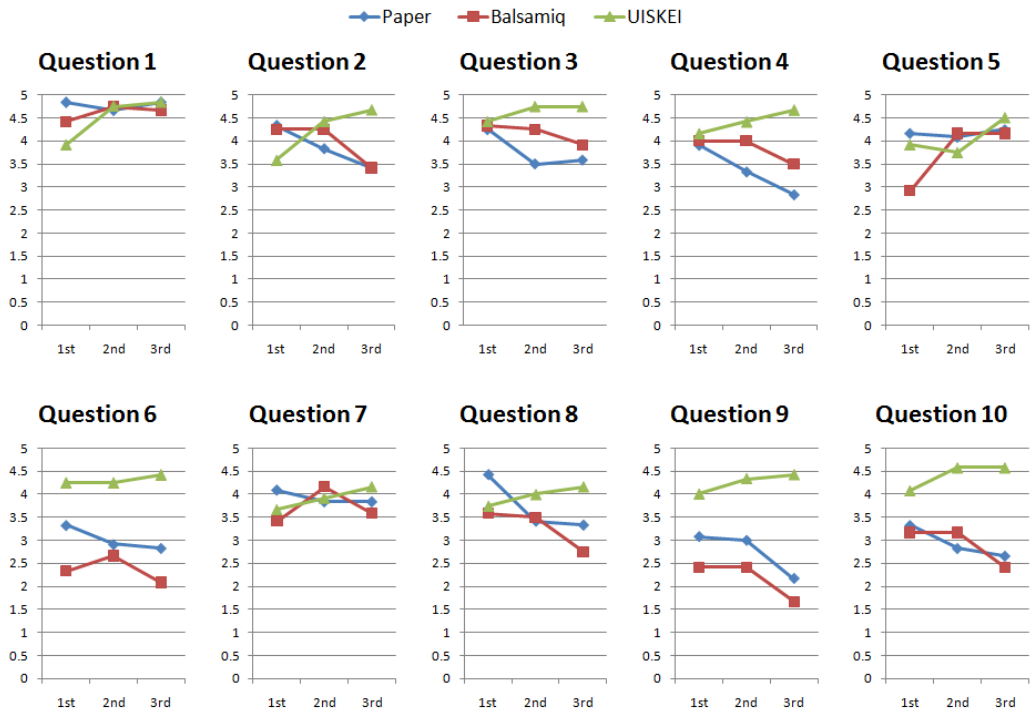

Figure 30: Average scores per question.

The lowest scores in the third cycle were given in questions 7 and 8 (both with an average score of 4.17). Compared to the other tools, the average of question 7 (paper had a 3.83 average while Balsamiq had a 3.58) suggests that participants faced difficulties in handling new behaviors with all the tools, but the answers to question 8 (in which paper got a 3.33 score and Balsamiq, 2.75) show

that the participants considered UISKEI as a more effortless way to solve these problems.

In the same cycle, the questions with the biggest differences from other tools were questions 6 (+1.58 from paper and +2.33 from Balsamiq), 9 (+2.25 from paper and +2.75 from Balsamiq) and 10 (+1.92 from paper and +2.17 from Balsamiq), all related to the interaction definition. In general, participants liked having a single interface (contrary to the multiple ones created in Balsamiq) and a previously defined behavior (opposed to the "on-the-fly" simulation of paper). Question 10 results also show that the mind-map representation of ECAs was well accepted.

Another good indicator of UISKEI's success was the answers to the "hated it / loved it" questions. Question 4 is related to the user interface and shows that the added complexity is quickly perceived in the paper technique, which faces an almost steady decrease in its scores in all cycles, while Balsamiq decreases only in the last cycle. UISKEI, on the other hand, has a steady increase of its scores, showing that the language to add elements, once learned, is well appreciated by users. The "hated it / loved it" interaction question (question 9) showed yet another pattern, with a steep decrease in the last cycle for both paper and Balsamiq, while UISKEI received an almost constant score, showing that users liked the way that the increased simulation complexity was handled.

Analyzing the question in groups (the interface building question - 1 to 4 - , the interaction building ones - 5 to 10 - and all the questions grouped together), it is possible to see that UISKEI achieved good results. Moreover, Figure 31 shows that the standard deviation of UISKEI's answers (the error bars in the bar graph) was smaller than for the other tools, suggesting that the participants seemed more in agreement when evaluating UISKEI.

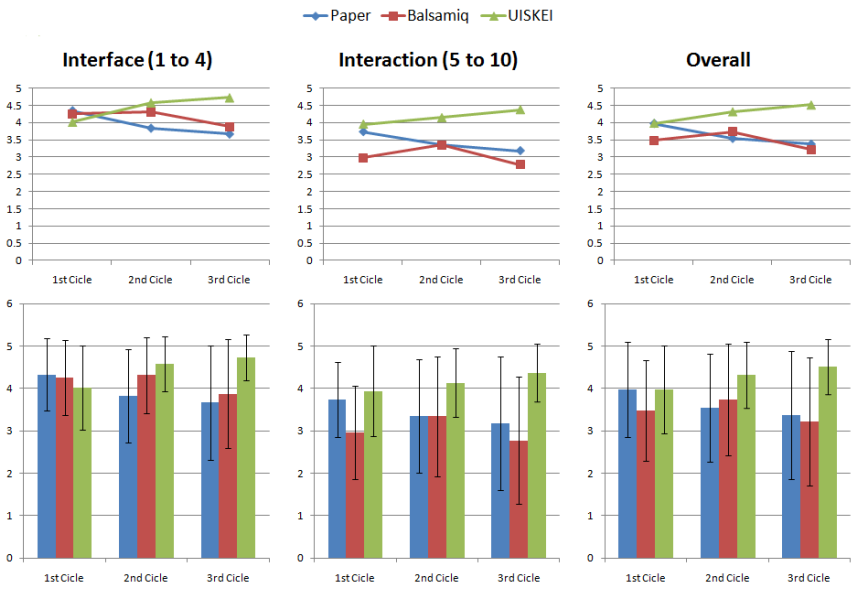

Figure 31: Average score per group of questions.

The interview results were also in favor of UISKEI, as can be seen in Figure 32. In the first question, while 25% of participants chose paper, 33% of them chose UISKEI. This near tie indicates that UISKEI's sketching method is comparable to paper, giving the desired "paperless prototyping" feeling to the participants. Balsamiq's results in the first two questions may be a result of its vast library of elements and features, such as alignment options and gridlines. However, the power of ECAs is shown in the answers to the third question, in which the vast majority chose UISKEI over the other tools.

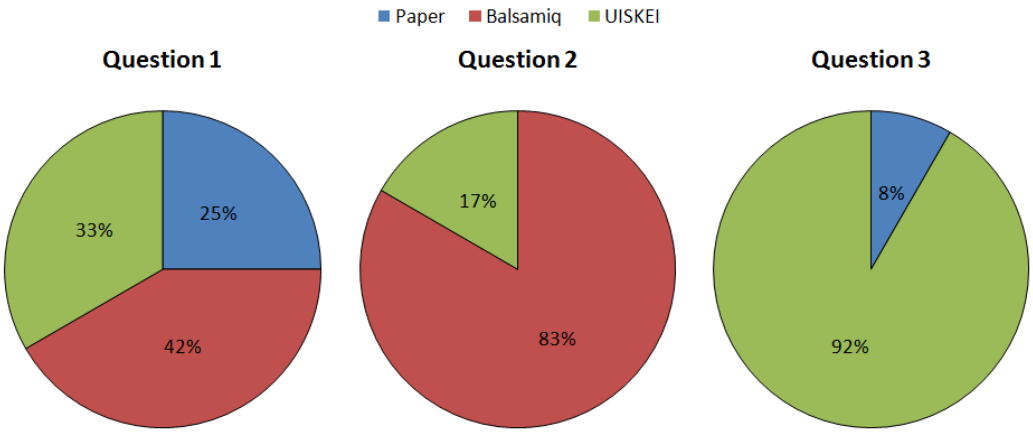

Figure 32: Summary of interview results.

8.3
Participants' opinions

Most participants did not know that it is possible to create an interactive prototype with paper. After seeing the introductory video, some participants already complained about the work involved. When they assumed the "computer" role during the prototype simulation, the number of complaints increased. One of the participants (p6) stated that "using paper may cause confusion in the 'computer'". This opinion was later reinforced by another participant (p8), who called him/herself as a "486" (referencing Intel's older line of microprocessors) during paper simulation, summing up that "Paper is fun, but not much practical".

The overall opinion about drawing on paper was that it is good for a rough sketch, but difficult to make changes. It can be summarized in p1's declaration: "it really complicates in the sense that you don't have too much flexibility once you have already drawn something. I think that you become too restricted to what you have or you start from the scratch, which is certainly not efficient". This shows a great disadvantage of paper, since the exploration of different solutions may be limited due to the effort of making changes in the prototype.

Regarding the paper simulation, some terms used were "boring", "disgusting" and "hell." The fourth participant pointed out something very interesting, saying that "despite being easy to do and easy to simulate, you don't have the register of what was happening, unless you film it, of course, but even so, you don't have the register of the used logic or even the errors that happened."

Balsamiq, on the other hand, divided opinions, ranging from people who loved it and others who hated it. Amongst the "lovers", the most praised features were the smart gridlines and the overall look-and-feel of the elements, showing that once you have these "aesthetics" facilities, they turn into a major concern. Actually, comparing the Balsamiq prototypes with the other ones, the disposition of the elements was much more similar to the ones pictured in the script. This focus on the detailed look-and feel was not observed in UISKEI, since, as a sketching tool, the focus should be more in the overall interaction and structure rather than on the "aesthetics" (Landay & Myers, 2001).

The "haters" focused their dislikes on the limitation of navigational actions, having to duplicate the prototype due to the lack of conditionals. The participant

p11 said that “it is a hell having to replicate (…) if there isn’t an ‘if’, nothing works”. In comparison with UISKEI, it was considered “less dynamic”, as said by p5, who simply said “Balsamiq is static, UISKEI is dynamic”.

Regarding UISKEI, the most common comment was about the “learning curve” and the terminology used. This visual programming problem was already stated in the work (Schmucker, 1996), which said “the tools for producing these applications often require months or even years of study to use effectively, and more often than not require the use of programming languages that are difficult to use even for professional programmers (e.g., C++)”.

However, the participants envisioned that, once learned, it would make the design process easier. P11 said that “in UISKEI, the learning curve is a little high, a little higher than normal, but once you get it, it is piece of cake”. Another participant, p4, added “it is something that you can take a while to learn, but once you learn it, it will be way faster to use”.

The behavior definition process fits the “learning curve” observation, since most users are not familiar with either pen-based interaction or pie menus. Besides the difficulties, it was well accepted, as described by participant p8: “I found UISKEI quite cool. I liked this way of connecting events, of creating conditions and actions, way practical and well integrated to the pen, easy to use even without mouse and even with the pen not being so precise”. This participant was not the only who complained about pen interaction issues: p1 lacked the use of keyboard, since writing recognition was not featured in Brazilian Portuguese (language of the test). One of the participants chose to experiment the writing recognition, translating the text to English, but found that entering text is slower than the keyboard, result also obtained in (Frye & Franke, 2008) regarding writing code.

The ECA representation was also praised. Comparing to Balsamiq and its replicated interface, p4 said that “UISKEI’s way is more interesting, because you can see all the conditions that are happening at the same time”. Another participant, p2, praised the way the ECAs were presented, saying that “what I liked is that it is very compact, you can work in a organized way (…) The interface is compact, things are shown in the right place and the actions are simple”.