Popup Workshop:
Preliminary Analysis of User Testing

Susan L. Hendrix
Department of Computer Science
University of Colorado at Boulder

28 Feb 2007

1   Introduction

This analysis of my user testing so far comes from a discussion with Clayton Lewis, in which I posed the question of whether I had done enough user testing. His suggestion, that I felt was a valuable one, was to take a look at user testing so far, what I had set down as the aims of user testing, and see if the two matched. This document is the result. It is an informal look at what I have done in this area and what I plan to do next. I come to the conclusion (quite spoiling the ending) that I have enough user testing at this point for the dissertation, and requesting correction if anyone feels this is not the case.

The evaluation plan is available at: http://www.cs.colorado.edu/ hendrixs/evalplan.pdf, and the reader may refer to it if necessary. I have also placed a companion document, that contains pictures of the popups for each test subject, at http://www.cs.colorado.edu/ hendrixs/committee/usersummary.pdf. This was too large to email.

The next section describes the test subjects so far. In Section 3, changes made to the evaluation process are described. Most of the process in the evaluation plan was followed and this section describes some additions, deletions and alterations. Section 4 looks at the observations made and whether they are adequate to support the areas discussed in the evaluation plan. The evaluation plan also listed some areas in which more experience was needed than that provided by earlier observations of users–the early work with 5th grade students, for example. These are discussed in Section 5. Finally, Section 6 details what work remains in terms of data analysis, and my conclusions about why I feel no more users are needed.

2   Users

The original plan was to recruit 5-8 students of ages 10-13. At present, 5 students with a larger age range have been tested, as two of the children were 6, two were 11 (twins), and one was 12. The following table summarizes user names, ages, number of popups and length of testing. The number of hours of testing is approximate, as each tape is counted as one hour (some may be slightly less). Emily is still in testing; her totals represent the testing so far, with the estimates of the final totals in parentheses.

A short summary of the work of each child might be in order. The interested reader may find a .pdf document with pictures from each session of testing at http://www.cs.colorado.edu/ hendrixs/committee/usersummary.pdf. This will become Appendix D of the dissertation. A short summary of each session will be added in the final document. I am currently going through the tapes, that have been converted to DVDs, recording times for later

Name	Starting age	No. Sessions	Hours	No. Pop-ups Completed
Ursula	6 (turned 7 during testing)	8	10	13
Richard	6	5	5	7
Daisy	12	7	11	7
Peggy	11 (turned 12 during testing)	8	8	4
Emily	11 (turned 12 during testing)	10 (12)	12 (15)	7 (9)

analysis, and preparing the summaries. Readers may wish to refer to this companion document, that was a bit large for email. (Or request a paper copy.)

Ursula was 6 on beginning testing. I was somewhat hesitant about the use of the program with a 6 year-old, but I knew that Ursula did a lot of paper crafts on her own, and decided she would be an interesting subject. She was very shy and said little, even when prompted in early sessions. In later sessions, she answered questions but was still quiet. I was amazed at her ability to make the pop-ups. Beginning with a simple frog (made by hand) and a simple abstract on the computer, she developed a facility with applied structures. She also had a predilection for bunnies and turtles. But it is interesting that after we returned to testing after a forced break of several months, the bunnies and turtles disappeared and she began to make more abstract shapes.

Richard was another 6 year-old. He is the brother of Daisy, an older subject. Richard also showed some ability with applied elements, and from the start was adding to elements and changing them. He showed a lot of imagination in his constructions, and liked telling elaborate stories about them. He could not write yet–this was a problem for the younger subjects, but he asked me to write words under tabs on some of his pop-ups. He was also much more forthcoming than Ursula. He would have continued longer, but his family was moving overseas, so we had to stop at 5 sessions. (We did manage the final interview before he left.)

Daisy was 12. She is Richard's older sister, and also had to move, although since she had started earlier, we did more. She showed a love of abstract shapes and faces–similar to those I saw with 5th graders earlier in the development of the program. At one point I suggested that she might illustrate a story she liked, as she was a great reader. She brought in a fantasy book about owls, wanting to make an owl to put up in her room. This was the largest pop-up made in user testing. It is more a paper sculpture than a pop-up. Everything pops up and folds except the talons, that were attached with double-sided tape in order to allow it to be folded for traveling.

Peggy and Emily are fraternal twins. They showed very different approaches, and I plan to devote a single chapter in the dissertation to their work, as opposed to choosing one child as a case study. Peggy was not someone who likes arts and crafts. She made only 4 pop-ups, but they were quite complex, and she spent a long time on them. She used applied elements, and added her own elements taken from books. Her pop-ups were the most mechanically complex of any user's, and she showed an interest in looking at the how-to books and understanding the more complex elements.

Emily loves art, and takes oil painting classes. Her testing is not yet complete; I estimate 2 more sessions of 2 hours each will be required at this point, that includes the final interview. She is making a book, with a series of characters all connected to each other–a man, his cat, the mouse eaten by the cat, and so on. She has added material to the computer generated elements, reshaped them, and added her own elements like flaps and tabs. She should complete the last animal (an elephant) in the next session, and plans one more pop-up with all the animals together. She is unusual in that she planned a book from the start, and every pop-up has been made with that in mind.

To summarize the users, the lower number of users were tested in my 5-8 subject estimate. However, the users were quite varied in age and approach, and they made more pop-ups than I would have estimated (I had assumed 5 pop-ups per user, and will have 40 instead of the 25 estimated for 5 users.) I will have 49 or 50 hours

of video to analyze.

3 User testing procedures

In Sections 4.2 and 4.3 of my evaluation plan, I detailed my methods, and what I intended to have done before testing. This section will take a brief look at what I actually did start with and do. What changed, and what remained the same? Overall, the plan was followed, although there were some changes.

In terms of what needed to be accomplished before the students started, the only thing dropped was the list of tasks and evaluation of the software by the methods of task-centered design. This still may be done, but not before testing. Version 2.0 of the software, with an improved viewer and applied structures was complete (although a few bugs were discovered and fixed during user testing.) I had planned to add 4 applied elements, and did 2. This was because I discovered that the M-fold was an easy enough adaptation of the V-fold that there was no reason to add it, and that the attached plane was simply a piece added to the parallelogram, that had its name changed to "tent" for user testing. (Standard, easy to remember names were chosen for each of the elements and used in all helps and when referring to them with the users: step, angled step, beak, tent and v-fold). In addition, there was no recording tool used. The video was trained on the screen, and it should be adequate to show the actions during construction.

Children did no work at home–everything was done during testing. The basic workshop environment was as described, with professional pop-up books, books on how to make pop-ups, and materials and tools available. Besides making pop-ups, we spent some time looking at pop-up books (with the video on), so I am able to make observations on their reaction to the pop-ups of others.

The procedure was the same as stated, except that one user made only 4 pop-ups instead of 5. All of the rest of the students made more than 5. In addition, during the final interview, the students were also presented with two pop-up books of *Alice in Wonderland* in very different styles and asked to compare them, in addition to describing the 3 pop-up books they had seen before. Also, during the sessions the users, particularly Peggy and Emily, were asked about the "rules" that the computer used during change. No video tape was shown to the users in the final interview, but they were shown still photos of each pop-up they had made, and asked about their construction.

In addition, I plan on a follow-up session with each user. Since two of my former users are now living overseas, and I would like to be consistent in the way I do this, this follow-up will be via email with the help of parents. I plan on doing the follow-up soon with the 3 earlier users and will wait at least a month for the twins. Questions will be asked about where the pop-ups are now, if they have done any more, and what they remember best about pop-up making, among others (analysis of the tapes should reveal areas that I might want to clarify for each user.)

4 Observations so far

In Section 4.1, I listed those questions about observations that I needed to make to confirm or falsify my hypothesis that the program could help children learn the craft of pop-up making, and also other observations I wished to make. This section revisits those questions. Are they answered, or at this point can they be answered with more analysis of the data on hand, rather than requiring more subjects? Under each heading from the evaluation plan, I'll list the questions and what I think has been done or needs to be done.

4.1 Craft skills

Craft skills deal with whether the users got enough experience in pop-up making to pick up skills and whether they improved those skills.

> *Do the children make 4 or more pop-ups successfully. (That is, are the pop-ups operational?)*

All of the users made 4 or more pop-ups. All of the pop-ups made were operational except for the owl–perhaps the most impressive construction, but one designed as a sculpture rather than a standard pop-up. This question is answered in the affirmative.

> *Do these pop-ups show an increasing level of difficulty? (Difficulty is in terms of 1) more elements 2) smaller elements 3) more use of applied elements 4) more use of combinations of elements and 5) use of asymmetry.*

A detailed summary will be prepared for the dissertation, but with 40 pop-ups to consider, enough data has been collected. The answer is mixed, with users showing fairly complex pop-up making abilities from the start. Richard's pop-ups probably did not grow in complexity. However, Emily's have done so. (Looking at the difference between Freddy Squarehead, her first pop-up, a simple but effective collection of non-applied elements, and Tap-dancing Cow No. 47, two separate parts of the cow sitting on v-folds with an accompanying flap, the cow is much more difficult.)

> *Do these pop-ups show an increasing technical sophistication in terms of changes made to the basic pop-up as made with the software. (This includes cutting away of material, addition of material, changing cuts, addition of other elements that the software does not support, and decoration options beyond those provided by the software.)*

Once again, it depends on the user, but in general designs grew more sophisticated. There are enough pop-ups that a more detailed analysis can be completed.

> *Do the children use the Viewer window to check the operation of the pop-up, and to locate colliding elements or elements that do not open properly?*

This has been seen. For instance, Emily used the viewer to check the position of Bart the Elephant's trunk and whether it would hit the face. The video tapes can be used to establish how often the users consulted the viewer.

4.2 Craft knowledge

Craft knowledge deals with whether the users learn about terminology and other information about pop-up making.

> *Do children refer to the elements in their drawings by name–either the name used by the program, or one they have devised themselves?*

Standard names were established and used during testing. I have observed that the children picked them up, and they were used during the final interview with the analysis of the professional pop-ups. How often awaits more extensive analysis of the video. (I did notice that *tent* and *step* were often confused.)

Does this change over the time that they are working with pop-ups, or become more frequent?

Once again, this will require more work with the videos.

Do children exhibit some understanding of the reason some pop-ups work and some don't? (This may be either expressed explicitly, or in their building actions.)

This was discussed at least occasionally during sessions, and in detail with Peggy and Emily. More analysis of the tapes is required to see just how much they learned about this.

4.3 Craft appreciation

Craft appreciation deals with developing an aesthetic for dealing with pop-ups–for instance, being able to look and critique a pop-up made by someone else.

Are children able to describe the construction and actions of a pop-up made by someone else?

The short answer is "Yes." Some of the children did a pretty good job, in fact, in the opening session.

Are they better able to do this after making pop-ups using the tool?

Do their feelings and judgments about pop-ups change? In particular, are they backed up with pop-up knowledge?

These questions are related. In general, the children appeared to have a better description after making the pop-ups. In particular, they have a better vocabulary for it. But a more detailed analysis needs to be done.

4.4 Other observations

There are some general observations that I have made at this point. 6 year-olds are, as one would suppose, quite different in their approach. First, they are less interested in the computer. They want to get their hands on the real paper. This may be a function of having less computer experience. However, both of the younger users do use computers for games, and were not afraid of using them. They also want to be shown the program first, rather than diving in and playing with it. Since the younger students used scissors and not craft knives, they had trouble with any small designs. Also, 6 year-olds are hampered somewhat by not being able to read and write, at least in making pop-ups with stories. This can be partly overcome by having an adult write for them, but it does mean they are less likely to make a story book. All this said, however, the younger subjects were able to design and build pop-ups using the tool. This was somewhat surprising, as I had assumed a starting age of 8 or so.

Children prefer symmetric designs when they design on the computer, but do not always use them. In addition, symmetric designs may become asymmetric through added pieces after leaving the printer. (In particular, the reader is directed to some of Ursula's work, Figures D.4-D.6 in the pictures document, or to Emily's giraffe in Figure D.30.)

As the videos are reviewed, more general observations will undoubtedly arise. Some of the other questions that I was interested in and will consider are the following:

Software support and interface:

1. What interface features support or hinder the user and in what ways?

2. Does the viewer window add to the user's ability to spot colliding or non-moving elements?

3. Does the use of the viewer change during the time they use the program. (For instance, is it something they play with at first, then ignore–or do they increase their use of it?)

Children's design:

1. What are the processes that children use to design pop-ups? (Similarities of approach among different children? An example might be the picky "I'll get the mouth just right before thinking about eyes" vs. "slap some stuff on and play with it.")

2. Are there some approaches that are more successful?

3. What is the role of pop-up constraints in the design?

Cognitive changes and learning:

1. Does the use of the software foster cognitive changes or learning? How? (I'm using paper tests of visualization and spatial reasoning, and talking to them about pop-ups before and after.)

2. Do the students learn the constraints involved in the design of pop-ups? (This has been discussed with the subjects during testing.)

Entering experience and capabilities:

1. What defines a successful pop-up maker? This includes factors such as age, visualization and spatial skills.

2. Many children may have varying levels of previous experience with pop-up books and pop-up creation. They may have differing experiences and skills with tools and other paper crafts. How do these affect their performance?

All of these questions will be considered in analysis of the video tapes. I believe that I will be able to make some observations about most of them, given the wide range of abilities, backgrounds, and age of the subjects I have recorded.

5   Previously identified untested areas

I had listed (in Section 2.3 of the Evaluation Plan) some areas in which I needed more experience. These are valuable to revisit, as I think they are all covered at this point.

> *One of the greatest outages in my experience is that I have only one pop-up from each student. Therefore, I have no knowledge about how the use of the tool, and their pop-up making skills, might change over time.*

I now have 4-13 pop-ups from each of 5 users. I think this is a large enough sample.

> *Another outage is that I have no data on the cognitive processes of children when using the tool, either in terms of what skills they started with, or how they learned. Part of this is the small time I have been able to spend with each student.*

My feeling is that this is an area that I will never cover in as much depth as I would like. However, I do have information about beginning experience, and a lot of comments made over the course of working from 5-15 hours with each of the users. There is a great deal of information waiting in the videos.

> *I have no experience with children attempting to build pop-ups with applied elements. These are more difficult, since they required fitting and gluing in addition to cutting and folding. Also, they require more decoration and alteration (cutting into shapes for instance) in order to make a picture. Of course, this means that they allow more variety and creativity. It is possible that use of these elements may make the process a completely different experience for the children.*

Children use applied elements readily. I do believe at this point that they enhance the experience, and more detail on this can be provided in the dissertation. In particular, however, I think those elements make it much easier for children to make pop-ups that are more than simple abstract designs.

> *Possibly the most important interface element for feedback to the designer is the Viewer with the animation of the pop-up. Since the Viewer is not a proper 3D display yet, I do not have good information on how children will use it. They played around with it, but did not use it in any systematic way. Whether or not this will be the case over time is a big question.*

More work needs to be done with the tapes to answer questions about how the display was used. However, the user testing was done with the improved display, that was much better than the original. The pop-up appeared as a true 3D object, that could be rotated to any orientation.

6   Conclusion

Emily should be finished in a few weeks. All the videos have been put on DVD, and I am starting to go through the tapes, preparing a rough summary, and a set of time stamps to make it easier to go back to relevant portions for more detail. In addition, I am beginning to write the dissertation, and I will be analyzing the pop-ups produced, the interviews and test results in more detail.

I do not believe that more users are needed at this point. The data collected (40 pop-ups and 50 hours of video estimated when Emily is done) is adequate to answer the questions posed. I have also filled in the gaps in my experience with users. I was concerned about the large age gap between the younger and older users, but I believe that this is actually a better mix than my original plan of only using subjects from 10-13 years of age. It has certainly extended my estimate of the usefulness of the program in younger children. The age range of the users is greater than was originally planned, even if the number of children is at the lower end of the range I'd anticipated having. Also, I have a larger number of pop-ups than I had anticipated. The most important job at this point is to analyze the video, that should take a fair amount of time. The law of diminishing returns tells me that more data will at this point lead to a less detailed analysis.

I would appreciate any feedback that you might have.