

# User experience: Evaluating UME admissions and educational tests

**Michael Persson**

Department of Informatics, Umeå University  
Historiegränd 8A, 90734 Umeå  
davaeorn@gmail.com  
+46767800097

## ABSTRACT

This report is a documentation of a number of evaluations performed on the Umeå University web based applicant test suite, and the subsequent discussion about its perceived issues from an user experience design perspective. It is found that there are quite a few things to improve.

## Author Keywords

UX, Tests, Think aloud, Heuristic evaluation

## INTRODUCTION

User experience (or UX) is a field that concerns every part of a user's interaction with technology, and what they feel about it. It is a very rich term, and hard to describe, in part because there is nothing that is not experience. *"We are always engaged in experience even when we are trying to stand back from it to describe it."* (Wright & McCarthy, 2004) We know that beautiful things, or things that make us feel good, are also perceived to work better. (Norman, 2004).

Some weeks ago, the Communication and International Relations Office in Umeå University approached the students of the HCI masters programme regarding their need to receive feedback on their interactive test suite for student prospects. After being shown some initial interest in the undertaking of evaluating the test suite, they compiled a formal project plan, which was subsequently shared with the students who had shown interest in the evaluation.

The test suite is comprised out of two tests, the admissions test (where the users can input their high school curriculum and average merit values to find out which programmes they are likely to be accepted to), and the educational test

where the users through grading the veracity of a statement can be given field and programme recommendations.

The project plan underlines a number of things, summarized below. The motivation of the project is mainly based on the current usage statistics of the tests, of which one has a frighteningly low average time spent. The beneficiaries of the project are the users, or the potential students who are going to be the test takers in the future. The purpose of the project is for the involved human-computer interaction students to evaluate the tests and make suggestions to make improvements and solve shortcomings. The planned effects of the evaluation should be an improved user experience, ultimately assisting in creating a long-term increase of applicants to the programmes of Umeå University. The project is to be realized through collaboration between the human-computer interaction students (who chose to participate in the evaluation) and the Communication and International Relations Office.

Through a proper empirical study based on the think aloud method and an inspection method called heuristic evaluation, the intent is to become familiar enough with the test suite to provide an informed contribution to whether or not a major redesign is needed.

## METHODS

To evaluate the test suite, two methods were chosen: think aloud and heuristic evaluation. Given the limited resources and the scale of the situation, this evaluation seemed to naturally attract methods that would not require a lot of external equipment or rely on quantitative, objective results. Since the purpose of the evaluation is to build a general understanding of the current state of the test suite in order to make informed design suggestions, it would not be reasonable to go in barrels blazing, or as Wiberg would have it: *"Nielsen, however, may have a point when the reference is to test design in which it is better to use resources to test more frequently with fewer users than to*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

spend all the resources on one test with a larger number of subjects.” (2003).

### **Empirical method: Think aloud**

Think aloud is a qualitative empirical method in which the evaluator instructs the subjects to perform a task and narrate their reasoning for all the actions they choose to perform. The evaluator should take care to not intervene in the actions or questions of the user, being sort of a ‘fly on the wall’ in the context. As noted by Wiberg, “*When the ‘fly on the wall’ approach is used the evaluator influences the subject less, and it is easier to compare the data from different subjects.*”. (2003) There is, however, a tradeoff to be considered: Since the subject is less likely to talk without queries from the evaluator, the data may require more interpretation. This means that the interpretation of a think aloud session is heavily dependent on having an experienced evaluator who can understand what the user is doing without having to ask them for specifics. Due to the proximity of the note taking to the actual interaction, the evaluator also needs to be able to filter the interesting results from the background noise “on the go”. (Wiberg, 2003)

The think aloud sessions were performed on an availability basis, with the only user requirement was being to be able to understand Swedish, as per the requirements of the test. After explaining the protocols for think aloud testing, the subjects were told to perform both the admissions test and the educational test, in whichever order they preferred. The answers are translated from Swedish. They are presented in the results section as either personal quotes, or as statements perceived by the evaluator. Furthermore, some of them are clarified by means of parenthesis and some by paraphrasing, where the evaluator felt it to be necessary.

### **Inspection method: Heuristic evaluation**

Heuristic evaluation is an inspection method where the evaluator, usually an expert in the field, uses an established list of guidelines, or heuristics, to evaluate an artifact. This method is used in order to determine whether or not the object of evaluation is structurally sound. The revised Nielsen set of heuristics was chosen for this evaluation. The set of heuristics is described as follows (Nielsen, 1994):

1. Visibility of system status
2. Match between system and real world

3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

The heuristic evaluation was carried out in a checklist fashion, where Nielsen’s revised set of interface design heuristics was used in a linear order, from top to bottom.

## **RESULTS**

### **Heuristic evaluation: UMU Admissions test**

1. Shows a progress meter in form of step fractions (four steps). No evident issues.
2. The language is easy to understand and not specific to some advanced field.
3. There is a button to return to the previous step, but there is no reset button in the end of the test.
4. Radio buttons and easily noticeable buttons to go forward and backwards in the test mostly follow the standard, except for the fact that radio buttons can be unselected, which is unconventional. Items in a scroll list normally don’t act as immediate links, but in this test, they are used in that way, which is also deviating from the norm.
5. The test is too small in scope to allow any major design flaws.
6. There could be a bit clearer reflections on the values that were input in earlier dialogs, but there aren’t a lot of dialogs, and they are all fairly concise.
7. As with number five, the test is too small to warrant expert shortcuts.
8. Minimalistic with some exceptions (such as the fact that the results are hidden beneath links to lists with every program).
9. There are no error messages. The minimal (grade) requirements could have been made clearer.
10. There is no real documentation, all the instructions for the test is documented within the test itself.

## Heuristic evaluation: UMU Educational test

1. It shows a progress bar, but it doesn't seem to be uniform (incrementation per step is inconsistent).
2. The language is understandable and at the level you can expect from the target audience.
3. There is only one button to proceed to the next question, so technically it's impossible to do any mistakes. There is, however, also no readily apparent way to go backwards in the test, until you find a strange secondary text menu in the top right corner which will let you both go forwards and backwards. There is also no reset button (or even a way to step backwards) as the test is completed, and a page reload is needed.
4. Except for the problems in point 3, it's standardized, and every checkbox and button follows regular conventions.
5. There could be fewer bugs. For example, it's possible to disjoint the viewport so that every subsequent question will be shifted to the left, breaking the readability of some questions and the general impression of the test itself. There was no discernible way to resolve this issue without reloading the entire test.
6. There is no way to see the way the questions are graded or how they impact your result. This could be to keep the focus on the process until it's complete, and as such be a conscious design choice.
7. As with the admissions test, expert shortcuts doesn't seem to be in the scale of this test either.
8. Except for the strange secondary text menu, its aesthetics are acceptable.
9. Evaluator has not been able to produce any error messages in the test.
10. There is no external documentation.

## Think aloud evaluations

### *Think aloud session 1: Female university student*

#### *Admissions test*

Subject enjoyed the movie clip. She remarked on the radio buttons being able to be unmarked. She also found the animations helpful in order to create the illusion of moving forward.

"Why did it open the pages in different windows? How can I compare them? Can I not do it in the same window?" (Regarding the result page, where every item in the list would open a new browser window) "I would bookmark the ones I find interesting."

"It's nice that you can go back." (Regarding correcting input mistakes in earlier steps by backtracking)

"How can I get to know more about the things I'm not good at?"

#### *Educational test*

"It doesn't have a video..."

"No percentage scale on the slider. How do I know how much every choice means?"

"Can I not go back? If I want to change my answer because of subsequent questions, how can I do it?"

At this point, the subject found the secondary text menu back button. Mixed reactions. Went back to change some of her answers.

"What will happen to a student with bad self confidence?" (Regarding the input slider)

"First they ask me about my interests, then they ask me about my skills – are they related?"

Subject notices that it is not possible to go backwards from the multiple checkbox questions.

"This is 'crafts' again. Why?" (Regarding the same questions seemingly showing up multiple times)

"Oh my god, I can share it on Facebook!" (Subject resorts to sarcasm)

"If I press this, will I lose the other things? It's Flash, so I can't right click..." (Subject clicks the button)

"I would like to click the other links, to see all the things I'm not good at."

"There should be a way to take the test again."

### *Think aloud session 2: Male university student*

#### *Educational test*

"How can I judge my own skills?"

Subject is using extremes, the slider being placed fully to either left or right, or left in the middle. He remarks that some subjects would have been interesting to study, after which he proceeds to mark them as "not interested".

"An education with a focus on recreation? What?"

"Haven't I already answered this question?"

"Why do I have to answer these things again?" (Regarding the checkbox dialog, which asks him to represent his interests with checkboxes)

"Is this the third time this question is asked?"

After clicking through the result list leading to the faculty program pages, he notes that some of the programs don't actually have a web page yet.

"Extremely long test, which probably could have been cut by two thirds. How relevant is it to know about my knowledge?" (As compared to what he is interested in)

#### *Admissions test*

"No chance in hell I'll remember what I've read" (Regarding the high school grades)

"Quite a painless way to get to know what programs you are eligible for, assuming that you know what you want to study."

#### *Think aloud session 3: Male university student*

#### *Admissions test*

"Weird, I can't get into anything. I'll have to change it to 2.0. (Regarding not remembering exactly what average score of the high school grade or his national test score)

#### *Educational test*

"I know how to use this slider out of habit."

"I'm a typical swede, I don't want to commit to a choice this early in the test." (Subject seems to have arrived at a personal conclusion of the test mechanics)

"Trick question?" (Again, seems to want to avoid being categorized by the test)

"There are many different levels on the slider, it's a bit annoying to consider how many degrees you are being judged on."

Subject is briefly distracted by an external banner in the column right of the test window.

"What do they mean with 'technical subjects'?" (A few questions later) "I think that 'technical subjects' are related to IT."

"I don't want to burn any bridges..." (Subject tries to manipulate his test answers in order to get good scores in the fields he is interested in)

"I would have used the tests if I was looking to start studying now, but I felt like as if was a bit steered towards economical studies."

#### *Think aloud session 4: Male, former university student*

#### *Educational test*

"Some questions seem to be related to previous answers."

"I'm not sure whether or not the questions are randomized." (A few questions later) "There doesn't seem to be a red thread..."

"The progress bar should be more opaque and it should be absolute rather than percentage based."

"Is the check box part a way to narrow down the results?" (Subject notices that the progress bar jumps from 35 percent to 70 percent after the check box dialogue) "Maybe they should narrow it down like that in the beginning of the test?"

#### *Admissions test*

"I think I understood the point of the test." (After having seen the introductory movie clip)

"There should be a button to mark that you haven't read the course."

"I don't understand why anyone would share and like this on Facebook."

"I didn't expect that to happen when I clicked on an item in a scroll list." (Regarding the scroll bar items being hyperlinks)

"They should have some link to a resource to help you calculate the merit value of your grades."

#### *Think aloud session 5: Male, no university background*

#### *Educational test*

"The sliders aren't numbered."

"What does 'sustainable growth' mean?"

"What is included under 'recreation'?"

"Some questions seem to be repeating."

"What's the difference between 'society' and 'social studies'?" (Note: some things are lost in translation – the names of the fields are more similar in Swedish)

"What's the difference between 'crafts' and 'building'?"

"Are the questions being limited by my previous answers?"

"It seems strange to ask me about things I have already answered..."

"This is the third time they are asking me about math."

"The result staples doesn't seem to have that much of a difference."

#### *Admissions test*

"I can't choose 'F' as a grade." (Subject initially seemed to believe that you could choose your specific grades in the course selection window)

"I don't know my grades or my average merit values."

"The box isn't big enough to show the entire name of some of the courses."

## DISCUSSION

Throughout the results of the evaluation, one can notice many common themes that arise between the different think aloud sessions. One common complaint seems to be that the educational test seems to ask the same questions over and over. Another one is that it seems to be too time consuming. These two points seem to be quite connected. Since I do not have access to the source code, it is hard to tell if there is some algorithm in play or if the questions just appear in a random order (which is also reflected by the think aloud subjects), but it feels safe to say that the test could really need to be more streamlined in regards of time expenditure. As one user said: *"Extremely long test, which probably could have been cut by two thirds. How relevant is it to know about my knowledge?"*

While some of the users did not even consider going backwards in the educational test to change an answer, the ones who did had problems finding the option to do so. The progress bar has also been reflected upon negatively as it behaves in a strange fashion and didn't really seem to help the users to keep track of the progress. As those points are also reflected by the heuristic evaluation, it seems safe to conclude that the structure of the interface is in dire need of maintenance.

There has also been some confusion regarding the web interface standards, and some of the contributions from the think aloud have actually reinforced and revealed some issues that were overlooked in the initial heuristic evaluation. One example of something that was overlooked was the fact that scroll list items normally shouldn't act as direct links (the heuristic evaluation was later updated to reflect upon this fact). Another example of something that was found in the heuristic evaluation but reinforced by the think aloud results is the matter with radio buttons, something that left some users wondering. The quotes *"There should be a button to mark that you haven't read the course."* and *"I can't choose 'F' as a grade."* both reflect on the fact that it was not at all clear that there was a possibility to 'uncheck' the radio buttons to show that the courses have not been taken or passed.

Another interface issue seemed to be the fact that the slider in the educational test was experienced as too analogue. Some concerns were had regarding the effect of self-confidence on the answers, and it is interesting to think about whether or not this was considered in the implementation of it. While it seems reasonable that for a single user, they are going to act in the same fashion for the

duration of the test and as such still get a relatively reliable test result, the disparity between the average distance from the center the slider is placed between users would make it hard to generate valid statistics in the context.

A major issue with the admissions test, which could partially be explained by the fact that the test subjects were not newly graduated high school students, was the fact that it was hard to remember their grades, which courses they had read, and what average merit values they possessed. Considering the fact that the Swedish school system has reformed grade and merit score systems almost incessantly in recent times, it doesn't come as a surprise that people have a hard time remembering or calculating their actual values. One user suggested leading the user to an external source: *"They should have some link to a resource to help you calculate the merit value of your grades."*, but it seems more streamlined to use the public information status on school merits to automatically track the merit values, possibly by requesting access to some database and allowing the user to enter his social security number at the start of the test for the purpose of identification.

The general attitude of the end result of the test suite was, however, quite positive. There seemed to be no confusion about whether or not the intended functionality was useful, albeit mostly in the favor of the admissions test: *"Quite a painless way to get to know what programs you are eligible for, assuming that you know what you want to study."* The introductory video is the only instance of multimedia in the current testing experience, and it was the only thing that a user expressed a longing for when moving from the admissions test (which housed the video) to the educational test: *"It doesn't have a video..."*

The design expressed a desire to be spread in social media, but the evaluations showed that the users were not ready to utilize the Facebook connection to share it: *"I don't understand why anyone would share and like this on Facebook."*

The initial intention was to use Wibergs revised heuristics (2003) to view the test suite from yet another angle, but after the first two studies it became apparent that the tests had been built with utility rather than experience in mind, which would render that particular endeavor excessive. While the evaluation subjects were useful to (relatively) quickly get an overview of the programmes in the different

fields, it's unlikely that they had a particularly engaging experience in the process.

Given the results of the evaluations, it seems like it is time to start thinking about a redesign of the tests, with the user experience in focus. There is no lack of ideas for such a redesign, some of which can be found interspersed throughout the discussion, but a full redesign is outside the scope of this report. Hopefully, the cooperation with the Communication and International Relations Office will continue when they return to work this fall.

## REFERENCES

- Wright, P., & McCarthy, J. (2004). *Technology as experience*. MIT Press.
- Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. Basic Books (AZ).
- Nielsen, J. (1994). Heuristic evaluation. *Usability inspection methods*, 17, 25-62.
- Wiberg, C. (2003). *A measure of fun: Extending the scope of web usability* (Doctoral dissertation, Umeå University).