

Final Report

Evaluating the “diet feedback”
of Dieetinzicht.nl and a prototype



Course

IN4179 Intelligent User Experience Engineering (IUXE)

Version

Final (may 2008)

Group 1

Steven Bos	1319671
Hannes Smit	1228897
Galatea Iatraki	1398741
Thomas Kararides	1366475
Sacha Panic	1367803
Iulia Chiriacescu	1300695

Introduction

This report is the second of a series of two on designing and evaluating an interface to improve the *diet feedback* interface of the website Dieetinzicht. This newly designed interface is simply called "Prototype". Whereas the focus of the first report (the Design Report) was on designing the Prototype, this report is on evaluating it and seeing if our claims in the Design report hold.

The objective of this evaluation for us student engineers, as stated in the Design Report, is to check through: confirm that the design phase was correct and complete and to examine if the Prototype is effective, easy to learn and suitable for its purpose in the Linda scenario. The claims related to this objective (when using the Prototype) are summarized as the following:

1. The user can find what he/she consumed over a certain period
2. The user can identify unhealthy ingredients
3. The user can find alternatives of certain ingredients
4. The user is provided with some sort of health-scale
5. The user can tell more about several ingredient details (fat amount, calories...)

Though normally hard numbers on user performance, experience/satisfaction, trust, number of errors, etc, are attached to claims, we discussed in the Design Report that we are too inexperienced to claim any numbers. This evaluation will be the baseline for future related evaluations and therefore the results are presented elaborately in this report.

In the first chapter the final version of the Prototype is discussed and compared with the Prototype from the Design Report. The second chapter focuses on our usability testing method, with subjects like the test procedure, measurements and more. The third chapter is the central *raison d'être* and is about the results of the evaluation. In chapter four we allow insight in our group discussion on the results. The final chapter "Conclusions", returns to the claims and objective stated above. Also some recommendations for future work are given there.

We tried to make this report accessible for selective (speed) readers by including chapter introductions. In general we tried to keep the amount of text compact and created visualizations where possible. In some cases we refer to the Design Report, but the general outline can be comprehended without it.

Enjoy reading,

Group 1

Table of Contents

1. THE FINAL PROTOTYPE	1
§1.1 FUNCTIONALITY	1
§1.2 USER TASKS/ACTIONS.....	1
§1.3 INTERFACE.....	2
2. USABILITY TESTING METHOD	3
§2.1 PARTICIPANTS.....	3
§2.2 SCENARIO & TASKS.....	4
§2.2.1 Scenario.....	4
§2.2.2 Tasks.....	4
§2.3 PROCEDURE.....	5
§2.3.1 Preparation.....	5
§2.3.2 Introduction.....	5
§2.3.3 Test.....	6
3. RESULTS.....	7
§3.1 EFFECTIVENESS.....	7
§3.2 EFFICIENCY.....	8
§3.3 USER SATISFACTION	9
§3.4 FINAL RESULTS WITH COUNTERBALANCING.....	11
4. DISCUSSION OF RESULTS.....	13
5. CONCLUSIONS AND FUTURE WORK	14
§5.1 CONCLUSIONS.....	14
§5.2 FUTURE WORK.....	14
REFERENCES	15
APPENDIX A: QUESTIONNAIRE PROTOTYPE.....	A
APPENDIX B: QUESTIONNAIRE DIEETINZICHT.....	B
APPENDIX C: PARTICIPANT DATA.....	C
APPENDIX D: FEEDBACK / COMMENTS / SUGGESTIONS.....	D

1. The final Prototype

This chapter briefly describes the final Prototype, designed and discussed in our Design Report. Because the Prototype was almost finished during the design phase no major changes on the interface have been propagated. Behind the GUI we created an XML system to load the Dieetinzicht data in our Prototype, so that Dieetinzicht and our Prototype used the same data for inferring "diet feedback". The first paragraph mentions the goal of our final Prototype. The second paragraph on User tasks/actions, is about the general tasks that could be completed with both the Dieetinzicht and Prototype interfaces. The last paragraph discusses our interface in little detail.

§1.1 Functionality

As the HTA in section 3.2.1 of our Design Report shows, our Prototype can be used to perform the task 'Show diet feedback'. This task allows users to look at their diet report, and get a general impression about the current status of their diet program.

§1.2 User tasks/actions

The HTA shows that there are 3 paths which can be followed when executing the task 'Show diet feedback':

1. A user selects a period of interest, and uses our new interactive diet overview page to get a short summary to check the progress of the diet, and obtain advice on how to proceed. (**Fig. 1**)
2. A user selects a period of interest, and uses our new interactive diet overview page to get to the detailed week and/or day nutritive value summaries. (**Fig. 2**)
3. For the completeness, the third step to print a diet report is included in the final design (yellow top right square) but it is not implemented and tested because it has no claims attached to it in this research.

To support the user when performing these actions, we developed a Prototype GUI. We set out to try to improve the current GUI both on a functional and an on a emotional level.



Fig 1. Diet overview in 3 steps

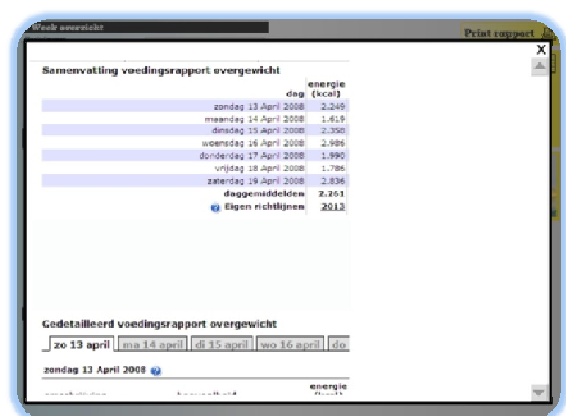


Fig 2. Detailed Diet view

§1.3 Interface

The State Transition Network in section 3.3 of our design report illustrated the three core actions which must be executed to perform the task 'Show diet feedback'. These actions are all implemented in the final Prototype. First, the user can pick a week, then the user can browse nutritional values (optionally in detail), and finally the user can look for advice on his or her diet. The two illustrations, fig. 1 and 2, reflect these steps in the GUI, where they are represented by a clearly identifiable, interactive and animated window. A lot of work is also done on creating interactive elements like sound feedback or highlights. All these specifics are mentioned in the Design Report.

2. Usability testing method

This chapter discusses our usability testing method. The “design” of our evaluation is based on theory by Benyon et al[1] and Neerincx [3]. We carefully made up a personae with traits mentioned in paragraph 2.1. We also considered a scenario where a test subject had to perform tasks in, which is presented in paragraph 2.2. In the last paragraph, 2.3, we present our test procedure with extra attention to measurements. It should be mentioned that in practice, some parts of our test procedure did not come to fruition, which is mentioned in chapter 4 “Discussion of the results”.

§2.1 Participants

This section provides a more detailed description of the participants to our experiment. In our previous report we mentioned Linda, our personae. By introducing Linda in our concept we have made an abstraction of our end users. We will regard the end users of the website as being characterized by a series of traits. By using these characteristics we can design a system that would be the most suitable for our target user.

The next table presents the traits of Linda that are the most relevant for our design.

Characteristic	Linda
General	Middle aged woman Slightly overweight Dutch speaking.
Physical	Good general health condition Clear vision, good perception of colors Able to use a computer
Psychological	Able to pay attention to solving tasks Able to memorize and recognize steps Able to associate meaning to colors
Computer experience	Average experience Comfortable with e-mail and web browsers Understands the part-whole relationship Perceives proximity
Medical knowledge	Basic understanding of nutritive values and BMI values

In the end we evaluated eight test subjects in a controlled environment, including one man (see appendix C: Participant Data). The results that we present in chapter three are based on 6 subjects, because:

- Participant three, who was the only male, not only did not match our personae, but also was an expert in usability testing. When testing, he paid little attention to the tasks given to him but instead tried to explore the application in search for bugs. The gathered data was therefore not usable.
- Participant six, who matched the Linda personae completely, apparently had too little time participating and filled in answers in the questionnaire without consolidating the application. Furthermore, when evaluating the second

interface, she completed it in under 1 minute, just copying the answers of the previous questionnaire (which is truly wrong to do because the asked questions are similar, but not the same as will be discussed in paragraph 2.3).

Being persistent, in the end we could evaluate six female participants that gave acceptable results and more or less fit our personae.

§2.2 Scenario & Tasks

This paragraph states the usage scenario and its related tasks. Compared to the Design Report there are minor changes in the task as we will discuss.

§2.2.1 Scenario

For testing and evaluation purposes we used one global scenario. In this scenario Linda, the personae that represents our target user has been using the Dieetinzicht website for more than two weeks. She has kept track of her daily diet by filling in the diary. She has come to the point where she is interested in reviewing the progress of her diet so far, by using the 'my report' feature of the Dieetinzicht website.

§2.2.2 Tasks

We designed a set of tasks which can be used to verify our claims. We then asked our test subjects to execute the tasks in a controlled environment. The users' feedback gathered afterwards enabled us to check if our improved design actually realized the claims. The tasks and related claims they cover, are listed in the table below:

Number	Task	Claims covered
1	Identify the most unhealthy ingredient in the past 2 weeks	2
2	How much calories does the most unhealthy ingredient in the past 2 weeks have	3
3	Find an alternative ingredient for the most unhealthy ingredient in the past 2 weeks	5
4	Find out what you ate during 3 specific days	1
5	Find the week with the lowest BMI	4
6	Find the week with the healthiest eating habits	4

While thinking of questions for the questionnaire we realized that task number five cannot be done in the case of Dieetinzicht. On first sight, Dieetinzicht shows the BMI value in a certain week, but this value is static and changes as soon as you modify it and thus there is no possibility to keep a history record. Our Prototype does provide this feature, but we decided to scrape this task to prevent biasing in any way. Scraping the task could be done because claim four is also covered with task six.

Furthermore we changed task three slightly by modifying the amount of days from three to one. This was done because doing the same trick three times is not really interesting to measure.

§2.3 Procedure

This section will describe the procedure used during the actual evaluation. The procedure consists of three parts: preparation, introduction and testing. In the testing part special attention is paid to measurement (what to measure, how to measure, in which order, etc).

§2.3.1 Preparation

We first started with brainstorming questions [4] that would involve one task so that we know that only a particular task was evaluated. The difficulty was that the question should apply on both interfaces. We reckon that the questions in the two questionnaires should be as similar as possible, with only a minor change to prevent literally copying the answer when a participant proceeds to the next questionnaire and sees the same question. For example the questions:

- “find out what you had for breakfast on Sunday 13th and
- “find out what you had for lunch on Sunday 13th

involve executing the same task but a different answer comes out when asked. The final questionnaire involved only 5 tasks with 13 questions (see appendix A and B).

To prepare for the experiment we had to fill the database of Dieetinzicht with real data. And to make the data believable we needed to base the diet on proper and unhealthy meals. We used recipes from AH to fill one week with healthy meals (Week A) and our own experience for a week with unhealthy meals and a lot of snacks in between (Week B). After that we had to load the same data in our Prototype to allow fair comparison (so Week A was again the healthy week). After that we completed the questionnaires ourselves to see what the correct answer should be. We then prepared for learning Morae in advance [5,6].

The actual experiment took place at the second floor of the EEMCS building, where we were provided a full blown Morae evaluation set-up. Normally we would use a webcam to capture the facial features when evaluating but unfortunately this was not possible in the first session. It was available the second session, but we ignored its data because we didn't have the same data set and then we would infer conclusions based on incomplete data.

As mentioned in paragraph 2.1 we evaluated eight participants and downsized that to six. To prevent biasing [2] in any way we let 3 participants start with the Dieetinzicht interface and let the other 3 start with the Prototype.

§2.3.2 Introduction

The participants were given a brief introduction on the context and goal of the test and evaluation session which they were about to participate in, and that the goal of the session was to compare the users' experience of the Dieetinzicht interface with that of our Prototype.

The instructions were:

For the Dieetinzicht evaluation the participants were instructed the following:

- Only to click on the *My report* buttons, not other functions of the website. This was demonstrated by one of the team members.
- To follow the task and related questions in the questionnaire and come out of the evaluation room when done.
- That when a task is too difficult to be executed, it can be noted down in the questionnaire.

For the Prototype evaluation the participants were instructed the following:

- To follow the task and related questions in the questionnaire and come out of the evaluation room when done.
- That when a task is too difficult to execute it can be noted down in the questionnaire

Since the scenario used for testing and evaluation states that our persona has been using the site for a few weeks, our initial thought was to give the participants about 10-15 minutes to briefly familiarize themselves with the (relevant part of the) interface which they were about to perform the tasks on. Due to time constraints though, we were not able to give every participant that chance, so that some users had to proceed directly to task execution. This gave some variation in our results as is discussed in chapter four.

§2.3.3 Test

The users were given all the time they needed to complete all the tasks without any help or any kind of influence from the team members. In this way time varied from user to user which gives us the opportunity, with a little help of Morae remote console, to examine all sorts of usability measurements/objectives.

Benyon et al. [1] presents a table with common usability metrics. As an usability objective we choose to do an overall usability evaluation, since we have nor the equipment nor the experience to handle measurements of trust, emotion and situation awareness. The chosen usability object implies that we measure:

1. *effectiveness* by measuring the percentage of tasks successfully completed. When a question is correct, it means that the task was successfully completed (see paragraph 3.1)
2. *efficiency* by measuring the time and clicks to complete a task. Afterwards we inspect the Morae data and add anchors where a user finishes a task, so that we know how long a task takes and how many clicks were necessary for that user (see paragraph 3.2)
3. *user satisfaction* by measuring three rating scales for satisfaction. One scale for the overall review, one scale per task and one scale of perceived time (to see whether the user had the feeling that using a interface took too much time) (see paragraph 3.3)

The order in which the measurements are done is as follows:

- First fill in two questionnaires (and thus do step 1 and 2 twice)
- Afterwards inspect the Morae data for both interfaces (and thus do step 3 twice)

During the test the participants were offered a drink and afterwards the possibility to leave their email to have this report send to them.

3. Results

This chapter presents the core of this document: the results. It is divided in four paragraphs. The first three are: measurements of effectiveness, efficiency and satisfaction. The last paragraph is on the final results when taking counterbalancing into account. Not much text is included; the tables and graphs usually speak for themselves. In the next chapter these results are discussed.

§3.1 Effectiveness

Dieetinzicht	Task 1	Task 2	Task 3	Task 4	Task 5
Correct answer	A	113 kcal	Frites & kroket	Spareribs	Various
Joyce	✓	✓	✓	✗	Couldnt find
Lisa	✓	✓	✓	✓	Pizza
Elly	✓	✓	✓	✓	Biefstuk
Gytha	✗	✓	✓	✗	Rundergehakt
Jenna	✓	✓	✓	✓	Biefstuk
Jessica	✗	✓	✓	✓	Couldnt find
Total Correct	4	6	6	4	Not counted
Total	20/24				

Prototype	Task 1	Task 2	Task 3	Task 4	Task 5
Correct answer	A	1 kcal	Brood, ei, etc	Spareribs	Various
Joyce	✗	✓	✓	✓	Magere spareribs
Lisa	✓	✓	✓	✓	Pizza
Elly	✓	✓	✓	✓	Magere spareribs
Gytha	✓	✓	✓	✓	Magere spareribs
Jenna	✓	✓	✓	✓	Magere spareribs
Jessica	✓	✓	✓	✓	Magere spareribs
Total Correct	5	6	6	6	Not counted
Total	23/24				

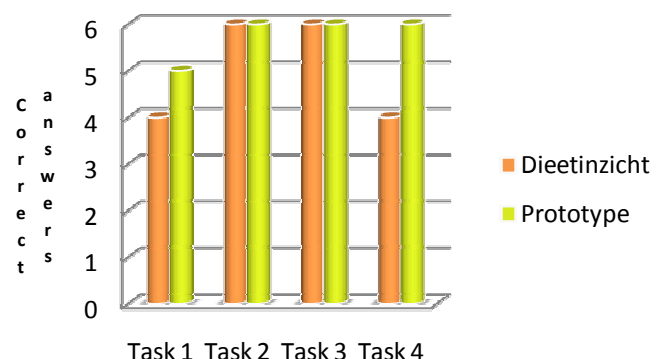
Interpretation of results:

In the Prototype one error was made when executing the tasks. Participants made a total of four errors when doing the same using the Dieetinzicht interface.

The effectiveness of the Prototype is therefor evident: in tasks 1 it performs 17% better and in task 4 it performs 33% better.

Note: Task 5 is not counted in these two tables because the measurement appeared to be uncomparable in practice.

Correctly answered questions in questionnaire



§3.2 Efficiency

Duration of completing tasks

Dieetinzicht	Task 1	Task 2	Task 3	Task 4	Task 5	Total
Joyce	144	23	1	168	44	380
Jessica	170	24	20	75	55	344
Jenna	105	8	25	280	540	958
Elly	77	57	93	11	0	238
Gyta	176	112	74	142	51	555
Lisa	75	16	48	52	17	208
Mean	125	40	44	121	118	447
Average deviation	38,8	29,7	28,2	75,3	140,7	62,54

Prototype	Task 1	Task 2	Task 3	Task 4	Task 5	Total
Joyce	215	106	13	74	0	408
Jessica	98	27	33	14	0	172
Jenna	27	122	33	170	0	352
Elly	300	392	214	1	96	1003
Gyta	60	126	172	46	5	409
Lisa	16	85	32	9	1	143
Mean	119	143	83	52	17	415
Average deviation	92,1	83,0	73,4	46,4	26,3	64,24

Clicks for completing tasks

Dieetinzicht	Task 1	Task 2	Task 3	Task 4	Task 5	Total
Joyce	4	2	1	8	Didn't	15
Jessica	24	5	5	13	7	54
Jenna	11	1	4	14	64	94
Elly	7	6	5	39	0	57
Gyta	15	11	5	8	5	44
Lisa	9	2	2	9	2	24
Mean	12	5	4	15	16	48
Average deviation	5,2	2,8	1,4	7,9	19,4	7,34

Prototype	Task 1	Task 2	Task 3	Task 4	Task 5	Total
Joyce	6	9	2	5	0	22
Jessica	15	12	5	2	0	34
Jenna	2	6	7	44	0	59
Elly	21	47	10	0	5	83
Gyta	3	13	13	3	0	32
Lisa	5	17	3	1	0	26
Mean	9	17	7	9	1	43
Average deviation	6,2	9,9	3,3	11,6	1,4	6,4

Interpretation of results:

To measure the efficiency we measured two variables the time and clicks needed for task completion.

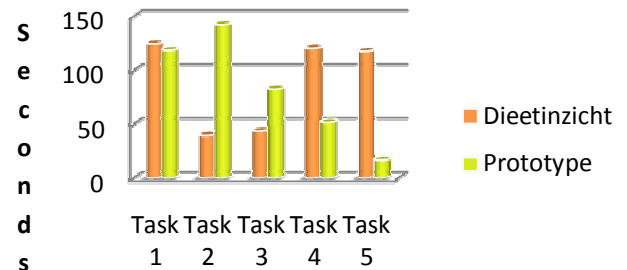
Interestingly, the Prototype is performing worse in task 2 and 3, while the Dieetinzicht interface is performing badly on task 4 and 5. In total the Prototype does win in terms of efficiency but surprisingly: only barely.

In total:

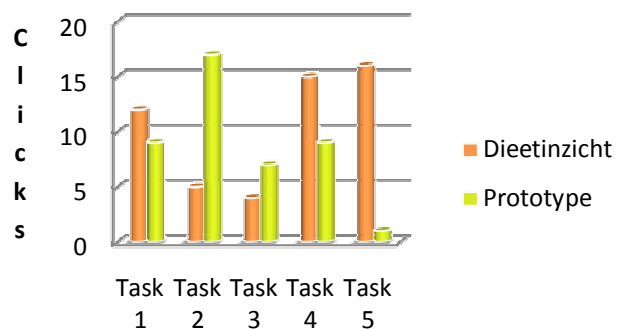
- Less time for Prototype: 34 seconds and thus uses 92 % of the total time used for Dieetinzicht.
- Less clicks for Prototype: 9 clicks and thus uses 83 % of the total clicks used for Dieetinzicht.

Note: Joyce used no clicks, but couldn't find an answer in Dieetinzicht, task 5. If we would insert a value by ourselves, for example the mean of the time used by the other participants, the end-result would be the same. So we skipped task 5.

Time needed for task completion



Clicks needed for task completion



§3.3 User satisfaction

Feedback per task

Dieetinzicht	Task 1	Task 2	Task 3	Task 4	Task 5	
Joyce	3	1	1	1	-	
Jessica	-	1	1	4	-	
Jenna	1	1	1	2	4	
Elly	4	1	1	4	3	
Gyta	4	2	2	2	3	
Lisa	3	1	1	2	2	
Total Mean	3	1,16	1,16	2,5	3	2,16

Prototype	Task 1	Task 2	Task 3	Task 4	Task 5	
Joyce	3	5	1	3	1	
Jessica	1	5	1	1	1	
Jenna	1	1	1	4	1	
Elly	1	5	1	-	3	
Gyta	1	3	2	2	1	
Lisa	2	4	4	4	4	
Total Mean	1,5	3,83	1,67	3,5	1,83	2,47

Feedback per task scale: Very easy (value 1) --- Very difficult (value 5)

Perceived time / Overall review

Diëtinzicht	Perceived time	Overall review
Joyce	1	2
Jessica	3	4
Jenna	2	2
Elly	3	3
Gyta	2	3
Lisa	1	2
Total Mean	2	2,67

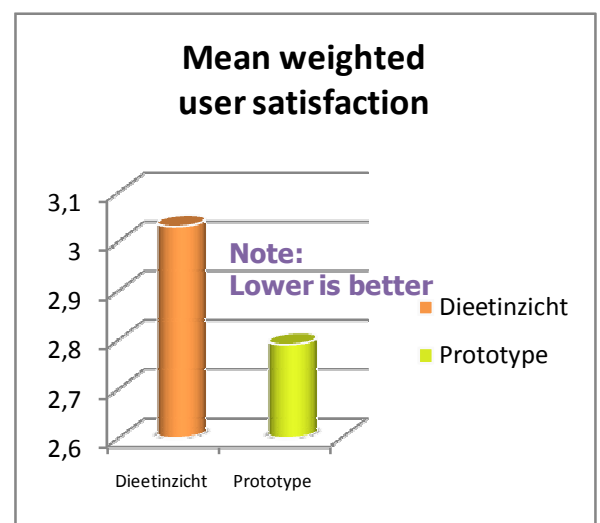
Diëtinzicht	Perceived time	Overall review
Joyce	2	3
Jessica	1	1
Jenna	1	2
Elly	3	1
Gyta	1	2
Lisa	2	4
Total Mean	1,67	2,16

Perceived time scale: task completion took faster then expected (value 1)
 task completion took not faster not slower then expected (value 2)
 task completion took slower then expected (value 3)

Overall review scale: Very usable (value 1) ---- Not very usable (value 5)

Interpretation of results:

- The average feedback per task is higher for the Prototype than for Diëtinzicht. This is remarkable because it means that tasks were considered more difficult using our prototype than using the Diëtinzicht interface.
- The amount of time the participants needed was equal to what they expected for the Diëtinzicht interface, but less than they expected to spend on our Prototype.
- The usefulness of the interface itself was neutral-to-relevant (between value 2 and 3) for both interfaces, but our prototype was considered slightly more useful.



§3.4 Final results with counterbalancing

Group 1	First Prototype	->	Second Dieetinzicht	Gytha, Joyce, Jessica
Group 2	First Dieetinzicht	->	Second Prototype	Elly, Jenna, Lisa

While analyzing the data below we decided not to compare Group 1 with Group 2, because the participants can differ so strong in a small testgroup that even basic statistics would be unreliable. We therefor have chosen to only compare the interface themselves using the same data presented in the tables in paragraph 3.1, 3.2. and 3.3.

Effectiveness	First evaluate	Then	Scale
Group 1	Prototype	Dieetinzicht	
Mean# of correct answers	3,67	2,67	In correct answers
Group 2	Dieetinzicht	Prototype	
Mean# of correct answers	4	4	In correct answers

With regards to effectiveness, all we can conclude is that the Prototypes returns more correct answers when it's tested at first. When it's tested at last, no differences appear.

Efficiency	First evaluate	Then	Scale
Group 1	Prototype	Dieetinzicht	
Mean duration per task	65,93	85,27	In seconds
Mean clicks per task	5,87	7,78	In clicks
Group 2	Dieetinzicht	Prototype	
Mean duration per task	99,87	93,60	In seconds
Mean clicks per task	11,67	11,20	In clicks

With regards to efficiency:

- The time it took to perform one task per user on average was high in Dieetinzicht, compared to our Prototype. Our Prototype works more efficient although it seems that group 2 takes more time for both applications.
- The amount of clicks used per task per user on average was a lot higher for the second group. In both groups the Prototype needs fewer clicks than Dieetinzicht, no matter if it was tested first or at last.

Satisfac tion	First evaluate	Then	Scale
Group 1	Prototype	Dieetinzicht	
Mean feedback per task	2,07	2,03	Very easy (1) ---- Very difficult (5)
Mean perceived time	1,33	2,00	Less time than expected (1) --- More time than expected (3)
Mean Overall review	2,00	3,00	Very useful (1) -- Not very useful (5)
Group 2	Dieetinzicht	Prototype	
Mean feedback per task	2,07	2,57	Very easy (1) ---- Very difficult (5)
Mean perceived time	2,00	2,00	Less time than expected (1) --- More time than expected (3)
Mean Overall review	2,33	2,33	Very useful (1) -- Not very useful (5)

With regards to user satisfaction:

- The feedback per task (the perceived difficulty per task) using the Prototype seems to be less easy than using Dieetinzicht in both cases. No matter the order. It is still in the category *easy* (= value 2,0) though.
- The perceived time using our Prototype gives the user the feeling that it takes *less* time than they would expect, when the participants have no knowledge about Dieetinzicht. On the other hand it takes exactly as much time they expect when Dieetinzicht was tested first.
- The overall review using our Prototype was considered more useful when tested first and thus Dieetinzicht is considered less useful when tested second, although the usefulness stays equal when tested at vice versa.

Simple comparison

Group 1	(Gytha, Joyce, Jessica)	Prototype	Dieetinzicht
Effectiveness			
	Correct answers	+	-
Efficiency			
	Duration	+	-
	Clicks	+	-
Satisfaction			
	Overall review	+	-
	Feedback per task	-	+
	Mean perceived time	+	-
Total		5	1

Group 2	(Elly, Jenna, Lisa)	Prototype	Dieetinzicht
Effectiveness			
	Correct answers	=	=
Efficiency			
	Duration	+	-
	Clicks	+	-
Satisfaction			
	Overall review	=	=
	Feedback per task	-	+
	Mean perceived time	=	=
Total		2	1

4. Discussion of results

In this chapter we discuss the results of our experiment in terms of functionality and usability. We will use data acquired from the questionnaires and data from Morae.

In the previous chapter we can see that the functionality was increased in the Prototype. The tasks that the users had to complete were in accordance with the claims. We can see in terms of efficiency, that the tasks were completed with only one mistake in the Prototype, whereas in Dieetinzicht there were four mistakes and furthermore task five was not completed by two users.

Using Morae we could measure the time and the amount of clicks necessary for each task to ultimately measure the effectiveness. The Morae data shows that on average the users spent less time using the Prototype than Dieetinzicht. Also, we see that the users spent *less time than expected* using the Prototype while in the case of Dieetinzicht the *expected* and *actual needed time* were the same.

The *mean* amount of clicks was *slightly less* (and thus better) for the Prototype even though the users clicked more than necessary just to get acquainted with the interface. Some users report that they had difficulties in finding where to click in the Prototype, while this was clearer in Dieetinzicht. On one hand this shows that the interaction elements of our interface are not emphasized enough. On the other hand, the Prototype still outperforms Dieetinzicht which means that after spending more time using the system, the user can get *faster* (in a lot of cases), and can get *smarter* feedback. We have to keep in mind that the speakers were not functioning in our Morae set-up, which prevented the users from being notified by sound feedback in clickable areas. Another factor that influenced the amount of clicks in both interfaces is that due to time constraints the testing protocol was not always executed. As a result some users were not given the time to familiarize with the interface before starting the test.

With respect to counterbalancing, the results show that the Prototype works better when it is used first. Both the amount of clicks and the time needed to perform the tasks are lower in this case. This comes in contradiction with our expectances. One reason might be that the two interfaces are totally different. Getting used to Dieetinzicht takes less time than getting used to the Prototype. Still in both groups the Prototype outperforms Dieetinzicht.

With respect to the user satisfaction. We were surprised to see that according to the comments (see Appendix D), some test participants prefer the Dieetinzicht interface. Even though the majority prefers the Prototype, the average feedback on the tasks, which is question 12 in the questionnaire, is higher for the Prototype than for Dieetinzicht. This means that tasks were considered more difficult in our Prototype than in the Dieetinzicht interface. The above state that we did not completely manage to design a user interface to accommodate our personae.

Another remark that we have to make has to do with the dataset used. As the database was not completely filled in, some users got confused. This thing had a negative effect on time and the amount of mouse clicks. Finally the users reported that our Prototype was more useful for supporting "their" diet than the Dieetinzicht interface.

5. Conclusions and future work

§5.1 Conclusions

Insert Conclusions with reference to claims

Based on the above and the discussion in chapter 4 we can conclude to the following list of things that went right and wrong:

WRONG:

test setup not adequate/provided (webcam, software etc.)
 role playing /persona could be better (protocol during test)
 too much design/ GUI? Issues→ made test more difficult
 Mouseover/ magnifying: interaction elements could be better designed.
 difficult to design a good questionnaire
 using testers from EWI/ MKE building → biased! → test results unusable
 difficult to measure learning curve? for users
 -More test users

RIGHT:

- in general it seems that Prototype is better on functional /emotional level
 - the general procedure (design-test-evaluation) was good- users found the test/questionnaire relevant

§5.2 future work

From the conclusion we can state that it is very likely that the Prototype is better on both functional and emotional level compared to the diet feedback interface of Dieetinzicht. Of course this should be proven more elaborately by evaluating more participants.

One of the strengths of Dieetinzicht is that their site lay-out is consistent and engineered from a certain point of view. Implementing our Prototype without also changing other functionality of the website (to make it consistent again) will only scare users. We do see future work in seriously redesigning the entire website so that the Prototype could fit in. As mentioned in our Design Report, the website is not user friendly and not stimulating to use. Most importantly it takes too much time using the website (filling in the diary, getting feedback, etc). In the scenario of a complete redesign, our Prototype (with some little tweaks) could become a feature that attract users, by giving smart feedback and at the same time stimulates them, to use it more frequently.

References

Books

- 1) Benyon, D., Turner, P., and Turner, S., (2005). Designing Interactive Systems: People, Activities, Contexts, Technologies. Harlow, England: Addison-Wesley (Pearson Education), Chapters 12 (mainly paragraph 12.3 and 12.5), 15, 17.

Papers

- 2) Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty – heuristics and biases. *Science*, 185(4157), 1124-1131.
- 3) Neerincx, M.A et al. Expert-Based Usability Review Method of TNO (2008)

Webreferences

- 4) Taylor-Powell, E., Renner M. Collecting Evaluation Data: End-of-Session Questionnaires <http://learningstore.uwex.edu/pdf/G3658-11.PDF>
- 5) Learning Morae : http://yukon.twi.tudelft.nl/iuxe/weblectures/morae_vincent_04/morae_vincent02.html
- 6) Learning Morae: <http://yukon.twi.tudelft.nl/iuxe/tutorial/manager03/manager01.html>



Appendix A: Questionnaire Prototype

Taak 1: In welke week heeft u het gezondst gegeten vergeleken met de dagelijks aanbevolen hoeveelheden (= een dieetrichtlijn)
Week A: van 6 – 12 april of Week B: van 13-19 april?

Vraag 1: Mijn gezondste week was:	<input type="radio"/> Week A: van 6 - 12 april <input type="radio"/> Week B: van 13-19 april <input type="radio"/> Ik weet het niet precies
Vraag 2: Indien u een week gekozen heeft, hoe makkelijk vond u het om uw keuze te maken?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 2: Hoeveel calorieën bevat een yakult light, gedronken tijdens uw ontbijt op zondag 6 april

Vraag 3: Heeft u kunnen vinden hoeveel calorieën een yakult light bevat?	<input type="radio"/> Ja, ... kcal <input type="radio"/> Nee
Vraag 4: Indien u dit gevonden heeft, hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 3: Zoek uit wat u voor ontbijt had op zondag 13 april

Vraag 5: Heeft u kunnen vinden wat u voor ontbijt had?	<input type="radio"/> Ja, oa. <input type="radio"/> Nee
Vraag 6: Indien u uw ontbijt heeft gevonden hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 4: Vind het ingrediënt met de hoogste vetwaarde in week B (van 13 - 19 april)

Vraag 7: Heeft u dit ingrediënt kunnen vinden?	<input type="radio"/> Ja, het was <input type="radio"/> Nee
Vraag 8: Indien gevonden, hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 5: Vind een vervangend ingrediënt met minder vet voor het ingrediënt gevonden in vraag 7. Indien u vraag 7 met "Nee" beantwoord heeft, kunt u vraag 9 en 10 overslaan.

Vraag 9: Heeft u een vervangend ingrediënt voor het ingrediënt uit vraag 7 kunnen vinden?	<input type="radio"/> Ja, bijv. <input type="radio"/> Nee
Vraag 10: Indien gevonden, hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Dit was de test van het Prototype, hieronder volgen nog 3 algemene vragen over het Prototype en test zelf.

Vraag 11: Wat vindt u van de tijd die u in het algemeen nodig had om de taken te voltooien?	<input type="radio"/> Het voltooien van de taken nam (te) veel in beslag <input type="radio"/> Het voltooien van de taken ging sneller dan verwacht <input type="radio"/> Het voltooien van de taken ging niet snel en niet langzaam
Vraag 12: Hoe relevant vond u de taken om inzicht te krijgen in "uw" dieet?	(kleur één rondje in) Heel relevant – o o o o o – Niet relevant
Vraag 13: Hoe bruikbaar vond u het systeem om er bijv. uw dieet mee te ondersteunen?	(kleur één rondje in) Heel bruikbaar – o o o o o – Niet bruikbaar

Hieronder is ruimte voor uw eventuele commentaar/suggesties:

-----Einde Enquete-----

Appendix B: Questionnaire Dieetinzicht

Samenvatting voedingsrapport energie

	energie	voet
dag	(kcal)	(kWh)
maandag 11 februari 2008	1.234	0,003
dinsdag 12 februari 2008	1.234	0,003
woensdag 13 februari 2008	1.234	0,003
donderdag 14 februari 2008	1.234	0,003
vrijdag 15 februari 2008	1.234	0,003
zaterdag 16 februari 2008	1.234	0,003
zondag 17 februari 2008	1.234	0,003
gemiddelde	1.234	0,003
daggemiddelde	1.234	0,003

Let op: de berekende daggemiddelde energie is een schatting. Deze is gebaseerd op de "dag"-waarde van de periode van dit rapport heeft u verspreiden getuigt op de eigen producties. Het is daarom mogelijk dat de berekende daggemiddelde energie afwijkt van de werkelijke daggemiddelde energie.

Taak 1: In welke week heeft u het gezondst gegeten vergeleken met de "eigen richtlijnen" (= de dagelijks aanbevolen hoeveelheid) **Week A: van 6 – 12 april of Week B: van 13-19 april?**

Vraag 1: Mijn gezondste week was:	<input type="radio"/> Week A: van 6 - 12 april <input type="radio"/> Week B: van 13-19 april <input type="radio"/> Ik weet het niet precies
Vraag 2: Indien u een week gekozen heeft, hoe makkelijk vond u het om uw keuze te maken?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 2: Hoeveel calorieën bevat brinta, gegeten tijdens uw ontbijt op zondag 6 april

Vraag 3: Heeft u kunnen vinden hoeveel calorieën een brinta(ontbijt) bevat?	<input type="radio"/> Ja, ... kcal <input type="radio"/> Nee
Vraag 4: Indien u dit gevonden heeft, hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 3: Zoek uit wat u voor lunch had op zondag 13 april

Vraag 5: Heeft u kunnen vinden wat u voor lunch had?	<input type="radio"/> Ja, oa. <input type="radio"/> Nee
Vraag 6: Indien u uw ontbijt heeft gevonden hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 4: Vind het ingredient met de hoogste vetwaarde in week B (van 13 - 19 april)

Vraag 7: Heeft u dit ingredient kunnen vinden?	<input type="radio"/> Ja, het was <input type="radio"/> Nee
Vraag 8: Indien gevonden, hoe makkelijk vond u het om het te vinden?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Taak 5: Bedenk een vervangend ingredient met minder vet voor het ingredient gevonden in vraag 7. Indien vraag 7 met "Nee" beantwoord, sla dan vraag 9 en 10 over.

Vraag 9: Heeft u een vervangend ingredient voor het ingredient uit vraag 7 kunnen bedenken?	<input type="radio"/> Ja, bijv. <input type="radio"/> Nee
Vraag 10: Indien gevonden, hoe makkelijk vond u het om het te bedenken?	(kleur één rondje in) Heel makkelijk – o o o o o – Heel moeilijk

Dit was de test van het Prototype, hieronder volgen nog 3 algemene vragen over Dieetinzicht en test zelf.

Vraag 11: Wat vind u van de tijd die u in het algemeen nodig had om de taken te voltooien?	<input type="radio"/> Het voltooien van de taken nam (te) veel in beslag <input type="radio"/> Het voltooien van de taken ging sneller dan verwacht <input type="radio"/> Het voltooien van de taken ging niet snel en niet langzaam
Vraag 12: Hoe relevant vond u de taken om inzicht te krijgen in "uw" dieet?	(kleur één rondje in) Heel relevant – o o o o o – Niet relevant
Vraag 13: Hoe bruikbaar vond u het systeem om er bijv. uw dieet mee te ondersteunen?	(kleur één rondje in) Heel bruikbaar – o o o o o – Niet bruikbaar

Hieronder is ruimte voor uw eventuele commentaar/suggesties:

-----Einde Enquete-----

Appendix C: Participant Data

ID	Test date	Gender	First interface to test
1	1-05-08	Female	Dieetinzicht
2	1-05-08	Female	Prototype
3	1-05-08	Male	Dieetinzicht
4	8-05-08	Female	Dieetinzicht
5	8-05-08	Female	Prototype
6	8-05-08	Female	Dieetinzicht
7	8-05-08	Female	Dieetinzicht
8	8-05-08	Female	Prototype
<ul style="list-style-type: none">Marked in red are users whose data is discarded from the statistical evaluation.Visible in this table is the randomness of which interface is seen first and that we did two different evaluation sessions.			

Appendix D: Feedback / Comments / Suggestions

Prototype	Feedback/Comments/Suggestions
Elly	<ul style="list-style-type: none"> • "I was confused because some percentages were empty at the beginning"
Gytha	<ul style="list-style-type: none"> • "It was unclear for me in which screen would appear after clicking on the bars and on the document icons" • "The colour-indication works well" • "When answering the question about the largest amount of fat, I wasn't sure if I answered correctly, because it's possible that I ate a lot of ribs..." • "Organised"
Jessica	<ul style="list-style-type: none"> • "Comments: I was very busy for a long time to look up what I had eaten per day. Later I learned that I had to press on the magnifying glass button." • "At question 6 I filled in that I could find it very easily, but that's because I had to look for this during the question before." • "Your site is very colorful and has a positive, clear impression. It is very easy to see what may be used as a substitute."

Dieetinzicht	Task 1
Joyce	<ul style="list-style-type: none"> • "In Dieetinzicht it's more clearer where to click"
Elly	<ul style="list-style-type: none"> • "It was a little late when I found out the possibility too select a period of 2 weeks. This prevents from searching between two weeks." • Suggestion: "I'm thinking about the possibility to sort total values for energy and fat etc."
Gytha	<ul style="list-style-type: none"> • "You have to scroll a lot." • "It's unclear at which day you're looking."
Jenna	<ul style="list-style-type: none"> • "The schematic representation of the nutritional value gives a clear overview and is therefore good for supporting a diet." • I just do not know how user-friendly it is to insert the data, because it was already there. • If you follow a diet, it is pleasant to determine in advance what you will eat. If you're looking at what you've eaten afterwards, it is already too late. • Maybe it's helpful to have a list of some foods, including the nutritional values. So that you can determine in advance. Then question 9 will be a little easier to answer. • Maybe it's a nice idea to include in the schema what you've done well in a given week, what's missing in your diet and what you've had too much of. With colors for example. Or perhaps you'll get a little neurotic then • I thought it was fun to participate! • (already done for 15 minutes)"
Jessica	<ul style="list-style-type: none"> • "Comments on Dieetinzicht: A very complex, not user-friendly website. It lasted very long before I understood the data, moreover, there were no colours used. It's not really motivating if you have to fill in what you've eaten from day to day."