Designing Towards a Unique Value Proposition

Iterating Using the Directed Product Reaction Method: A Case Study

Max Liedström Kvelland
maxli412@student.liu.se

Advisor: Johan Åberg
Examiner: Anna Levén

Linköping University
LIU-IDA/KOGVET-G-13/011-SE
Acknowledgements

I would like express special thanks to Johan Åberg. Not only for coming up with the original concept behind combining PRC and Lean, but also for providing assistance and feedback during the writing of this thesis, and during prototype design.

I also wish to acknowledge the help provided by Katarina Liedström in recruiting potential users for the testing phases of iteration 1 and iteration 2.
Abstract

Lean Startup lacks a method specifically designed to evaluate whether a HCI-design is fulfilling the promises delivered by the unique value proposition (UVP). At the time of writing, concurrent think aloud protocols during usability testing and interviews are used to that end. There are, however, a couple of issues bundled with the use of such methods. Using a case study conducted on two iterations of high-fidelity prototypes of a meal planning website currently in development, a method called Directed Product Reaction (DPR) is proposed. DPR is a result of the adaptation of Product Reaction Cards, and the addressing of issues uncovered during case study iterations. DPR aspires to be that specifically designed method needed for assessment of user experience in relation to UVP in Lean Startup, and while the resulting method proposed has far from achieved that goal, it serves as a foundation moving forward.
# Table of Contents

1. Introduction .................................................................................................................. 1  
   1.1 Structure .................................................................................................................. 1  
   1.2 Lean Startup ........................................................................................................... 1  
   1.3 Unique Value Proposition (UVP) ........................................................................... 2  
   1.4 Product Reaction Cards (PRC) .............................................................................. 2  
   1.5 The Idea Behind Combining PRC and Lean Startup ............................................ 2  
   1.6 Research Questions ............................................................................................... 3  
   1.7 Delimitations ......................................................................................................... 3  
   1.8 Source Discussion .................................................................................................. 4  
2. Theory ............................................................................................................................ 5  
   2.1 Lean Startup in Practice ......................................................................................... 5  
      2.1.1 Identifying the Business Plan ......................................................................... 5  
      2.1.2 Creating the UVP ......................................................................................... 5  
      2.1.3 Identifying Problems and Customers ............................................................ 6  
      2.1.4 Creating Your Solution ................................................................................. 6  
   2.2 Explaining Product Reaction Cards ...................................................................... 6  
   2.3 Johan Åbergs Adaptation of PRC ........................................................................ 7  
   2.4 Usability and UX ................................................................................................... 7  
3. Case Study ...................................................................................................................... 9  
   3.1 Case Description .................................................................................................... 9  
   3.2 Procedure ............................................................................................................. 11  
      3.2.1 Iteration 1 ....................................................................................................... 11  
      3.3.1 Iteration 2 ...................................................................................................... 17  
4. Proposed Method ......................................................................................................... 24  
   4.1 Translating the UVP to Key DPR Concept Words ................................................. 25  
   4.2 Expanding the Alternatives .................................................................................. 25
4.3 Directed Product Reaction (DPR) ................................................................. 26
4.4 Applying DPR in Practice (example) ............................................................. 28
5. Discussion ........................................................................................................ 30
  5.1 Limitations of DPR in Iteration 1 and 2 ......................................................... 30
  5.2 Discrepancies in UVP Interpretation Between User and Designer .............. 32
  5.3 Validity and Reliability in DPR ..................................................................... 33
  5.4 Further Research ......................................................................................... 34
    5.4.1 UX Framework ...................................................................................... 34
    5.4.2 "Shadowing" and Emphasis on Interviews ............................................. 35
    5.4.3 Future iterations of DPR ......................................................................... 36
6. Conclusions ....................................................................................................... 36
References ........................................................................................................... 38
Appendixes ........................................................................................................... 40
  Appendix A, interview script ........................................................................... 40
  Appendix B, abbreviation glossary ................................................................. 41
1. Introduction

1.1 Structure

This chapter is dedicated to a short overview of the structure of this thesis. In a nutshell, the thesis consists of six major parts: introduction, theory, case study, proposed method, discussion, and conclusions.

In the introduction, there are short explanations of some key concepts for this thesis: Lean Startup, Unique Value Proposition, and Product Reaction Cards. Also, an argument regarding why Product Reaction Cards may be a good fit for Lean Startup is provided, setting the stage and illustrating the potential need for this thesis. Also presented here are the research questions. There is also a list of delimitations and constraints, as well as a discussion of the three most valuable sources to this thesis.

Following the introduction is the theory section, containing a summary of the practical application of Lean Startup. Product Reaction Cards is then explained in more detail, followed by a presentation of Johan Åberg's adapted version of PRC. Last, a discussion of how Usability and UX relate to Directed Product Reaction is provided.

The next section contains an application of Directed Product Reaction (DPR) in a real interaction design setting by means of a case study. The case is described, as well as a detailed account of the procedure, along with a presentation of the case study results and analyses.

After the theory section comes the section called proposed method. This is where DPR is presented as an answer to this thesis' first research question (how can PRC be adapted to assess whether a design is approaching the values of a UVP or not?). The key elements of DPR are explored, and an example of application is provided.

Towards the end, we find the discussion section, discuss some of the problems and ideas uncovered during the writing of this thesis and during the case study. There are also some recommendations for future research in relation to DPR.

At last, the chapter called conclusions contain an overview of the conclusions arrived at during this project.

1.2 Lean Startup

Lean Startup is a methodology aimed at entrepreneurs attempting to create a new product within a market that entails a great deal of uncertainty, as is the case for most new products and brands (Ries, 2011). Lean Startup, in contrast to traditional earlier startup strategies, is not focused on initial
market researching and extensive planning. Rather, its focus lies in maintaining a high time efficiency, and minimizing waste, while learning as much as possible about the problem before creating a solution (Ries, 2011). According to Eric Ries, who coined the term Lean Startup, it "represents a new approach to creating continuous innovation" (Ries, 2011, p.4). Ash Maurya (Maurya, 2012, p.XXIII) wrote that Lean Startup "[…]represents a synthesis of Customer Development, Agile Software Development methodologies, and Lean (as in the Toyota Production System) practices." and that the goal of Lean Startup is to optimize the use of time, while maximizing the learning about the customers and users.

1.3 Unique Value Proposition (UVP)

Within a Lean Startup, Unique Value Proposition (henceforth abbreviated as UVP) is a key concept. A UVP is a phrase used internally (within the production team) to guide and direct design towards a common goal. An example being "Spend More Time Building Versus Planning Your Business - The faster, more effective way to communicate your business model" (Maurya, 2012, p.31-32). The UVP is also used externally (meaning it is available to customers), and because of that capturing the essence of your product is important, while also making it clear to customers why your product is worthy of getting attention.

1.4 Product Reaction Cards (PRC)

Product Reaction Cards (henceforth abbreviated as PRC) is a method for testing design developed at Microsoft (Benedek, J. & Miner, T., 2002). It is a tool for measuring intangibles (e.g. abstract concepts such as appeal and fresh) in a design. See chapter 2.2 for more details.

1.5 The Idea Behind Combining PRC and Lean Startup

While Lean Startup is far from lacking in terms of variety of possible methods used to test usability, there are to my knowledge no clear-cut methods used to assess whether a design is closing in on a unique value proposition in the early stages of iteration before achieving product/market fit. A method frequently used to assess UVP fit is a qualitative one, the MVP interview (Maurya, 2012), where one of the goals is to learn whether the minimum viable product delivers on the UVP. This is done by conducting a usability test of the product, where the user is asked to complete one or several tasks (for example completing the registration and activation process). A concurrent think-aloud protocol is typically used. Also, semi-structured interview questions are asked continuously in order to obtain the desired information. Preferably, according to Maurya (2012, p.129), the whole design team is present during the interview.
There are a number of issues involved in using a concurrent think-aloud protocol (cTAP) to assess the user experience. First, the design of cTAP lies more in line of focusing on the strategies behind decisions made in interaction with the product (usability) (Kuusela & Pallab, 2000), rather than the experience had when using the product (UX). As mentioned in the later stages of this thesis, Hassenzahl (2008) argue that UX and usability are two separate (although symbiotic) notions. Second, using a cTAP very seem to affect the way in which the product is being used, more specifically causing tasks to feel more cumbersome than they normally would (Ericsson & Simon, 1980). This is problematic as bad UX is facilitated by low usability (Hassenzahl, 2008). As such, I believe a new method is needed to properly measure the UX aspect.

PRC is not a traditional usability evaluation method, since it does not produce a list of usability problems, but rather provides an indication of what direction the user thinks the design is taking in terms of concepts. In other words, it is more inclined towards UX than usability. With UX on the rise, both as a concept and as a standalone goal of design, a method specifically indented to measure UX is surely needed. With this in mind, PRC would seem like a good starting point for the method proposed in this thesis. The general idea is that an adapted version of PRC can serve as a way of assessing whether a design is closing in on the values of a UVP or not. This thesis contains a suggestion of such an adaptation, which has been named Directed Product Reaction (abbreviated as DPR), as well as an example of practical application of such a method by means of a case study.

The hopes of this thesis is that by proposing DPR as method, it can take on the role as a complementary go-to method within Lean Startup to measure UX alongside the usability evaluation method of choice, in place of the current cTAP and qualitative MVP interview segment combination.

1.6 Research Questions
As such, the questions of interest become the following:

1. How can PRC be adapted to assess whether a design is approaching the values of a UVP or not?

2. In what way can this adaptation be applied in practice?

1.7 Delimitations
This thesis focuses on adapting a method to work within a Lean Startup setting, and the method proposed should thus not be assumed to work in other settings, even though that may be possible.

As the case study is conducted in Sweden, there may be possible cultural boundaries affecting the use of this method in other cultures. Note that the proposed method will, however, not take any cultural differences into consideration.
1.8 Source Discussion

The most important sources for this thesis are the following:


*Desirability: New methods for evaluating desirability in a usability lab setting* was written by Joey Benedek and Trish Miner in 2002, and introduces Product Reaction Cards to the world. To my knowledge, it is the latest document on PRC produced by the creators of the method. While perhaps not being the most recent document on the subject, it is the primary source for the PRC method, and that is my ultimate reason for using this source.

*Running Lean: Iterate from Plan A to a Plan That Works* is a book written by Ash Maurya in 2012. In this book, Maurya explains how to apply Lean Startup in a project, providing both theoretical and actual examples in a walkthrough-fashion. Ash Maurya himself is a web-based software developer with extensive experience within the startup scene, being involved with USERcycle (creator) and Mozilla (mentor), to name a few. The book received a ringing endorsement from the man who coined the term Lean Startup, Eric Ries. Ries also voiced a need for the book and validated Ash Maurya as an author of Lean Startup, by claiming "[...] I have heard an overwhelming demand for practical guidance for how to put Lean Startup principles into practice. There is no better person to begin that mission than Ash Maurya." (Maurya, 2012, p.XIII). It is a recent creation, and can therefore be assumed to be up-to-date, but it is also an excellent guide to the application of Lean Startup. These are my reasons for using this source. However, as Maurya is very much involved in the creation and use of Lean Startup, it is doubtful whether he can be considered an objective source of information or not.

*User Experience (UX): Towards an Experiential Perspective on Product Quality* is a paper written by Marc Hassenzahl in 2008, and in it he presents his view on UX in its current state, as well as explaining some of the subjective origins of user experience and emphasizing important aspects of it.
He also provide some definitions from ISO CD 9241-210, and introduces the notions of pragmatic and hedonic qualities of product experience. Hassenzahl argues that there is a separation of usability and user experience, but that they also contribute to each other. Hassenzahl is one of the more recognizable names within the study of UX, and has produced many papers on the subject. Hassenzahl's approach to UX has been criticized as being too reductionist, and while the argument he provides as a response is not convincing enough to be indisputably held for truth, this paper remains an important source for this thesis.

2. Theory

Presented in this section is the theoretical background and foundation of reasoning behind the suggested method, Directed Product Reaction. First and foremost comes a chapter explaining the practical steps leading up to the UVP in a Lean Startup, followed by a more detailed explanation of PRC as a method. Furthermore, we explore the terms Usability and UX in relation to PRC, and discuss what PRC (and as such Directed Product Reaction) actually measure.

2.1 Lean Startup in Practice

There are a couple of steps involved in the creation and usage of the UVP within Lean Startup, but since this thesis is focused mostly on the UVP, a short summary of the steps will be provided.

2.1.1 Identifying the Business Plan

The business plan in Lean Startup comes in the form of a Lean Canvas (Maurya, 2012), which is an adaptation of Business Model Canvas (Ostenwalder & Pigneur, 2010). Ostenwalder and Pigneur define a business model as "a business model describes the rationale of how an organization creates, delivers, and captures value." (Ostenwalder Pigneur, 2010, p.1). The idea behind Business Model Canvas (and consequently Lean Canvas) is to capture this on a single page. The Lean Canvas itself is broken down into 9 segments: Problem, Customer Segments, Unique Value Proposition, Solution, Unfair Advantage, Revenue Streams, Cost Structure, Key Metrics, and Channels.

2.1.2 Creating the UVP

The UVP is a key concept for this thesis. It is a phrase or sentence that is used both internally (to influence design) and externally (to get customers interested in your product). The UVP is what makes the product different and worthy of attention. In other words, "Unique Value Proposition: Why you are different and worth buying getting attention" (Maurya, 2012, p.29). For internal use, a UVP is used to influence design by providing a conceptual common goal for the entire team. For external use, the idea is not to create a UVP that targets as many mainstream customers as possible,
but rather to focus on your early adopter customer segment at first. One of the things that Maurya mentions is that a good UVP includes some of the end results of using the product. According to Maurya, some of the high profile luxury car brands have used a single word to distinguish themselves from others: BMW using Performance, Mercedes using Prestige, and Audi using Design. While these words are not typical models of a UVP, they illustrate one side of the external use of a UVP, namely to distinguish your product from the rest of the market.

Further development of the Lean Canvas will not be discussed here, since it does not cohere with the purpose of this thesis, but is outlined in detail in *Running Lean: Iterate from Plan A to a Plan That Works* by Ash Maurya. Instead, the following chapter will present how to move forward once a Lean Canvas and a UVP have been created.

### 2.1.3 Identifying Problems and Customers

What may be experienced as a huge problem for developers of software may or may not be of great significance to users at all. It is imperative that the problem is correctly understood, and that it is an actual problem users and consumers are faced with, before resources are spent on creating a solution to that problem.

### 2.1.4 Creating Your Solution

Once you have a set of problems that are of immediate interest to users and customers, a minimum viable product is created using the necessary features gathered from your qualitative research results and the lean canvas.

The MVP developed is used to determine how well the solution solves the problem. This is determined by using qualitative and quantitative measures to test if the solution appeals to people, if it is something the users and customers want. Maurya recommends interviews for qualitative data, and cohort reports (based on simple funnel reports) for quantitative data (Maurya, 2012). Further explanation on how to apply Lean Startup in a project can be found in *Running Lean: Iterate from Plan A to a Plan That Works* by Ash Maurya.

### 2.2 Explaining Product Reaction Cards

As written in chapter 1.5, a common way of doing this is through the MVP interview. An adapted version of PRC, however, may serve as an alternative to the MVP interview.

As previously mentioned, Product Reaction Cards is a tool for measuring intangibles (e.g. abstract concepts such as appeal and fresh) developed at Microsoft by Joey Benedek and Trish Miner (Benedek & Miner, 2002). Benedek and Miner noted that usability testing procedures at the time
measured whether users could complete tasks (among other things), but were not as effective at measuring the experience in terms of intangibles. Evaluation of concepts such as fun and enjoyment simply was not conducted in these procedures, and therefore resulted in a lack of information regarding the actual appeal of the product. It was noted, however, that qualitative interviews did allow for evaluation of such concepts, but that some aspects of this approach prevented it from being an optimal method. These aspects include the time consumption of conducting and analyzing the data, as well as inherit social barriers making it harder for users to express their actual feelings (especially if those feelings are negative) toward the product.

In response to this, Benedek and Miner developed Product Reaction Cards, which essentially is a set of 118 cards containing concepts that can be used to describe the product in question. Some example words are Accessible, Busy, Predictable, Relevant, and Stimulating, to name a few (Benedek & Miner, 2002). The set of 118 cards contain 60% positive words, and 40% neutral or negative words. At the end of usability evaluations, the user was asked to choose 5 cards with 5 keywords that best described the product. The selection of cards was done in private to combat the aforementioned social barriers of expressing negative thought. After 5 cards had been selected, the user was asked to elaborate on the reasons behind selecting said cards, which often resulted in specific details about product interaction as well as thoughts about design and concept (Benedek & Miner, 2002).

Benedek and Miner also suggested that the design team may also undergo the PRC procedure in order to describe their own product, which enables a comparison to be made with the results from the user studies, allowing for inferral of information on how much the design vision differ from the measured intangibles of user experience.

2.3 Johan Åbergs Adaptation of PRC

In the PRC adaptation used by Johan Åberg and colleagues, 13 households were presented with 10 concept words (free translation from Åberg, 2013): Health-promoting, Cost-effective, Time-saving, Personal, Appetizing, Eco-friendly, Reliable, Guiding, Inspiring, and Innovative. Households were first asked what words they would associate with the service in question, and were subsequently asked to select the three words that best described the service. Subjects were then asked what prompted them to select that cards that were selected, giving the team some additional qualitative data.

2.4 Usability and UX

Usability as a concept and PRC have a somewhat strained relationship, as PRC does not actually measure usability in the traditional sense. The ISO 9241-11 definition of usability reads:
"Usability: the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."

Combine this with the definitions for effectiveness, efficiency, and satisfaction from usabilitynet.org (2013):

"Effectiveness - can users complete tasks, achieve goals with the product, i.e. do what they want to do?

Efficiency - how much effort do users require to do this? (Often measured in time)

Satisfaction – what do users think about the products ease of use?"

These definitions seem to infer that for a method to be truly considered a usability testing method or usability evaluation method, it must measure effectiveness, efficiency, and satisfaction. PRC does not focus on any of these aspects. One could argue that PRC could be used to measure satisfaction in some form or fashion (depending on the definition of the word), but it is not specifically designed to measure "ease of use". Furthermore, according to usability.gov (2013), a number of quantitative and qualitative metrics are entailed within usability testing: time on task, error rates, successful completion rates, observations of pathways, problems experienced, and comments (among others). PRC contains no quantitative data measures, such as time on task, error rates, and successful completion rates. Regarding qualitative data, as PRC in previous projects have been used complementary rather than as a standalone method, PRC testing has typically been conducted at the end of the standard usability tests, in private. This means that PRC traditionally has not contained any observations of pathways, problems experienced, nor spontaneous comments. While there is a qualitative element within PRC, namely the follow-up conducted after card selection, that alone does not justify the classification of PRC as a usability testing method. This means that we cannot expect PRC to behave or produce as a usability testing method, and PRC should not be held to the traditional standards of such methods. As a consequence, it would be unwise to use PRC (or any adaptation in the same spirit of PRC, i.e. Directed Product Reaction) in the same way a usability testing method is used.

If usability aspects is not what is measured in PRC and Directed Product Reaction, what is it that is actually measured? Consider the term User Experience (henceforth abbreviated UX), often (and many would say to a fault) used interchangeably with the term usability. Hassenzahl, in his article User Experience (UX): Towards an experiential perspective on product quality, provides a two-part definition of UX, the first part being:
"I define UX as a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service."

The second part reads:

"Good UX is the consequence of fulfilling the human needs for autonomy, competency, stimulation (self-oriented), relatedness, and popularity (others-oriented) through interacting with the product or service (i.e., hedonic quality). Pragmatic quality facilitates the potential fulfilment of be-goals."

He integrates the traditional concept of usability into what he calls pragmatic quality. The pragmatic qualities of a system or product allow for the completion of "do-goals" (in other words, tasks), and pragmatic qualities are the result of "good" interaction design in terms of usability and utility. These pragmatic qualities differ from what is termed hedonic qualities, which facilitate the fulfillment or achievement of "be-goals". Be-goals, according to Hassenzahl, is "the driver of experience" (Hassenzahl, 2008, p. 2). While a low degree of usability in a system may bar the completion of be-goals, and while a high degree of usability may increase the chances for fulfillment of be-goals, it is not desired by the user in itself. Instead, desired be-goals may be feelings of competence or stimulation. As such, hedonic quality is a direct factor in achieving a positive experience, while pragmatic quality indirectly support positive experience.

One could argue that PRC, and Directed Product Reaction, measure the hedonic qualities of a product through the subjective experience of users. It does, however, not measure all and any hedonic qualities, but the emotions that users are allowed to express are controlled and filtered by the words provided as options in testing.

3. Case Study

3.1 Case Description

DPR was tested on a meal planning service currently in the formative stages of development, namely Planeatsmile.com.

Planeatsmile.com is a swedish website developed by ÅAB Meal Planning Concepts AB with the ambition to "support people in maintaining a healthy diet, while having a positive impact on health, environment, time consumption, and housekeeping" (translation from ÅAB Meal Planning Concepts AB). The key concepts that have influenced design have been Health, Time, Economy, and Environment (Åberg, 2013). In short, the website is a way for users to discover new recipes and plan
their meals using these, in accordance with a set of preferences. Below (figure 1) is a screenshot showing the starting screen of PlanEatSmile:

![PlanEatSmile starting screen](image)

For example, users are able to adjust how large proportion of suggested recipes should correspond to specific criteria, some of which in the currently released version are:

- Recipe must contain fish
- Recipe must abide by vegetarian standards
- Recipes suggested must be novel (that is, never before suggested to the user in question)

Using the suggested recipes, the user is then able to plan the meals of the next 10 days by drag-and-dropping recipes onto the planning timeline. In the current version, new users have their preferences set to default values for the abovementioned criteria, but there has been an expressed interest in implementing a lifestyle test (for example [http://www.alexit.se/](http://www.alexit.se/) who provides [http://www.livsstilsanalys.se/](http://www.livsstilsanalys.se/)), focusing on dietary habits, to dictate what preferences may fit the user. Prototypes of this lifestyle-test feature, as it would work if implemented, served as the environment for the DPR testing.
3.2 Procedure

In short, the aforementioned suggested method was applied to the case in two iterations. The iterations contained the creation of prototypes and testing using Directed Product Reaction. Both the prototypes and method was redesigned prior to iteration 2 based on the data collected from iteration 1.

3.2.1 Iteration 1

3.2.1.1 Prototypes

Prototypes were made in high fidelity, partly using Photoshop (used for creating graphics), and partly using JavaScript, CSS, and HTML (used for creating interactive behavior). The choice of hi-fi prototypes over lo-fi prototypes was made in order to be as close to the MVP as possible, which as far as fidelity goes is essentially a finished (although basic) product. Two prototypes with deviating characteristics were created, one will be referred to as wizard, and one referred to as profile. Below (labeled figure 2) is a screenshot of the wizard:

Figure 2. Screenshot of the wizard prototype from iteration 1.

Users would answer questions relating to dietary habits, and in the end were recommended a level of improvement fitting of their current dietary level.

Below (in figure 3) is a screenshot of the other prototype, named profile.
Figure 3. Screenshot of the profile prototype from iteration 1.

In this prototype, users were able to select a profile with preset dietary goals.

3.2.1.2 Initial Method

Translating the UVP to concept words
There were many points of concern regarding how to map the values of a UVP to concept words. If the aim was to assess whether a design was moving in the direction of the UVP, the link between the concepts represented by the UVP and the concepts provided for evaluation in Directed Product
Reaction had to be clear. Before explaining how that was to be achieved, let us take a look at some example UVPs provided by Maurya (2012, p.31-32):

- "Spend More Time Building Versus Planning Your Business - The faster, more effective way to communicate your business model"
- "Turn your users into passionate customers - Customer Lifecycle Management Software"
- "The Fastest Way to Share Your Photos and Videos"

Most likely, a number of conceptual words usable as concepts spring to mind when reading these UVPs. "Spend More Time Building Versus Planning Your Business - The faster, more effective way to communicate your business model", for example, may evoke the words timesaving, fast, effective, communicative, efficient, expedient, helpful, and so on. Should the user describe the design using any of these words, one could consider the design to be closer to the UVP than if another irrelevant word, such as compact for example, had been used.

What this means is that what words are included in the Directed Product Reaction is very important, because the words included filter what feelings users are allowed to convey. Because of this, including only positive words related to the UVP would result in all data collected being more or less in favor of the design being tested, as there is no way to express anything but positive feelings. To counter this, we provided words with opposite meanings to the positive words provided, as well as some words that are considered irrelevant in terms of the UVP.

Another thing to keep in mind is that the user interpretation of the UVP may differ from the designer interpretation of the UVP, meaning we expect the design to convey different experiences based on who we are. This thesis and the related testing has been conducted under the assumption that the difference in interpretation between designers and users is small enough to ignore. See chapter 5.2 in the discussion for thoughts about how to act in situations where the difference in interpretation is presumed to be large enough to consider.

**Directed Product Reaction**

How was the notion of translating from UVP to concept words applied in practice to DPR? At the time of testing, PlanEatSmile’s UVP was "A flexible and adaptable way to eat healthy" (freely translated from Swedish). Using this UVP, three primary concepts were derived: flexibility, adaptability, and health. These three concepts were then expanded using synonyms, where flexibility was represented by both flexible and formable, adaptability by adaptable, personal, and accommodating. Health was only represented by one word, healthy. A negative version of each word were also supplied, as well as three words that could apply to the prototype, but that did not have any direct relation to the
UVP. As such, the concepts of the initial Directed Product Reaction consisted of these 15 words (freely translated from Swedish to English):

<table>
<thead>
<tr>
<th>Positive words</th>
<th>Negative words</th>
<th>Irrelevant words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptable</td>
<td>Rigid</td>
<td>Professional</td>
</tr>
<tr>
<td>Flexible</td>
<td>Clumsy</td>
<td>Tasteful</td>
</tr>
<tr>
<td>Formable</td>
<td>Awkward</td>
<td>Compact</td>
</tr>
<tr>
<td>Personal</td>
<td>Unengaged</td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>Unhealthy</td>
<td></td>
</tr>
<tr>
<td>Accomodating</td>
<td>Troublesome</td>
<td></td>
</tr>
</tbody>
</table>

Design success was measured in how well the respective UVP keywords ranked among users (for each customer segment), ranging from no users selecting none of our UVP keywords whatsoever (total failure) to all users including the respective UVP keywords in their top 3 (high grade of success). For example, if zero users picked the keywords adaptable, healthy, or flexible, in their top three, the design would be assumed to be fairly weak. On the other hand, if all users picked all three of the main concepts in their top three, the design would be deemed rather successful.

**Instructions and user recruitment**

After being recruited, users received an e-mail containing an explanation of the testing procedure, as well as login information and information about what dates and times the users were allowed to complete the test.

A total of 11 users were contacted, 6 were asked to test prototype profile, while 5 were asked to test prototype wizard. These users were recruited based on the following criteria:

- Membership of targeted customer segment
- Coherence with previously constructed personas
- Ease of access

The prototypes were tested separately using Directed Product Reaction, where prototype wizard was tested by 2 users (3 chose not to participate), and prototype profile was tested by 5 users (1 chose not to participate).

As PRC at the time had only been used in conjunction with, and as a complement to, other evaluation methods (Benedek & Miner, 2002) users being tested with PRC were already accustomed to the
product in some way due to having completed other usability tests prior to PRC. This meant that PRC in itself did not require users to complete tasks involving the product being tested. In iteration 1, a set of tasks involving the entirety of the product are introduced in every test to make users accustomed to the product in the same way usability testing did prior to PRC tests in its original format.

A total of 7 tasks were designed to give the users a tour of the website as a whole and make them acquainted with it before they proceeded to test the prototype and complete the Directed Product Reaction test. These tasks involved the users testing the different functions and features of the website, such as planning a week of meals, as well as adjusting the food preferences, and replacing ingredients in recipes with healthier alternatives. To avoid users focusing too much on the website, the focus on the prototype was emphasized in the e-mail instructions, both as the main target of evaluation, and development.

**Format of web-based DPR**

After completing the assignments and completing either prototype wizard or prototype profile, depending on group, users were asked to move on to the web-based test.

Users were asked to select exactly three alternatives. The users were allowed to complete the test anytime between 18.00 and 22.00 within a time span of 4 days (Monday to Thursday). After submitting the results, a text message with their name was sent to me, and the user was immediately contacted for a qualitative follow-up (in style of a semi-structured interview). The goal was to gain
more detailed information regarding the reasons why users selected the cards that they selected, while also trying to determine if the method used was appropriate. As mentioned earlier, Hassenzahl's first-part definition of UX states that it is a *momentary* evaluative feeling. As such, it was important that the DPR test was conducted as soon as possible after the user had been in contact with the product.

### 3.2.1.3 Results

The results collected from the PRC testing in iteration 1 were as follows:

<table>
<thead>
<tr>
<th>Themes from iteration 1 interviews.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Testing</strong></td>
</tr>
<tr>
<td>The right words available</td>
</tr>
<tr>
<td>Good amount of words</td>
</tr>
<tr>
<td>The test duration was satisfying</td>
</tr>
<tr>
<td><strong>PlanEatSmile</strong></td>
</tr>
<tr>
<td>Changing to healthy alternatives was simple</td>
</tr>
<tr>
<td>You got to choose what recipes to use</td>
</tr>
<tr>
<td>Not the kind of recipes kids enjoy (too many components mixed together, it felt complicated)</td>
</tr>
<tr>
<td><strong>Prototypes</strong></td>
</tr>
<tr>
<td>The prototype did not affect the result</td>
</tr>
<tr>
<td>Nothing was missing from profiles</td>
</tr>
<tr>
<td>The wizard was simple to complete</td>
</tr>
</tbody>
</table>
Themes have been freely translated from Swedish. The themes are sorted based on what area they refer to. Testing themes are themes that refer either to the PRC method, or the test duration as a whole. PlanEatSmile contain themes referring to the website as a whole. Themes listed under Prototypes refer either to the wizard, or profile.

All users were interviewed within 7 minutes after completing the PRC test, except in one case where contact occurred after 25 minutes.

3.2.1.4 Analysis
As you can tell from looking at table 1, the wizard suffered from a high percent of users ultimately choosing not to participate in the testing, which in turn meant very sparse results.

Furthermore, during interviews many users spoke in referral to the website as a whole, rather than the prototyped part (wizard or profile), despite the emphasis on the prototype in the instruction e-mail. When asked during interview, 3 users revealed that the prototype had no effect on their answers in the Directed Product Reaction test. This is also consistent with the finding that a majority of the themes referred to the website as a whole, and not to the individual prototypes. This is most likely a result of the majority of tasks (7 out of 8) being carried out outside of the prototype, in order to acquaint the users with the website as a whole. Users seemed to believe that the words included were sufficient in order to express the feelings they had for the product, with only one user reporting that he/she would have answered something else had he/she been able to type in text freely.

Regarding the amount of words, most users responded that the amount was enough. As for the length of the entire test, all users seemed to think that the time taken to complete the process was adequate.

Another issue with the method used in iteration 1 was that data collected from testing indicated that the design was either positive, negative, or irrelevant. There was no spectrum of results. This could prove problematic, since a user that may have been slightly inclined towards the positive end of a word pair, such as healthy and unhealthy, would be likely to have answered healthy in the same way as a user that found the design to convey health more strongly. This also ties in to some data that emerged during the interviews, where one user reported that the negative words felt too negative.

3.3.1 Iteration 2

3.3.1.1 Prototype
For iteration 2, the data gathered in iteration 1 was used to iterate and redesign the prototypes in an attempt to improve the overall user experience and to further reinforce user emphasis of the chosen
UVP keywords. This ultimately led to a convergence of two prototypes into one, which was subsequently used in the DPR testing. The prototype, named profile selector, is shown in figure 5 below:

Figure 5. Screenshot of the profile selector prototype from iteration 2.

This prototype worked similar to profile from iteration 1, with the difference that users were able to adjust values of dietary goals directly prior to selecting a profile (such as how large proportion of meat in recipes would be exchanged for fish).
3.3.1.2 Refined Method

Instructions and user recruitment
Users recruited for iteration 2 received an e-mail containing an explanation of the testing procedure, similar to the one sent in iteration 1. The times and days allowed for testing were similar to the ones in iteration 1, with the exception of the time window, which was reduced to 19.00 to 22.00 (from 18.00 to 22.00).

A total of 7 users were contacted, with 6 users successfully completing the DPR testing. Users were selected based on the same criteria as in iteration 1, with the additional requirement that they had not been participating in iteration 1.

The data collected in iteration 1 was used to refine the Directed Product Reaction in a number of ways. Many users in iteration 1 focused on the website as a whole rather than the prototyped part. As such, iteration 2 did not contain any tasks set outside the prototype, but users were provided with a scenario instead. The scenario was written with the personas developed in an earlier stage of website development in mind, and described a situation thought typical for members of the target customer segment. Some of the functions offered by PlanEatSmile as a whole were outlined, and the role of the prototype in relation to the website as a whole was explained. Users were also asked to play around with the settings for the different profiles, and were informed how to restart the prototype in order to try out different settings and selections.

Changes to DPR
As in iteration 1, iteration 2 involved a web-based test, with some alterations. In iteration 2, users were able to rate each chosen concept on a scale from one to four, to produce data on a spectrum. This was done to allow for greater variation in results, and differentiate users that felt strongly about a concept from those only slightly inclined towards that concept. For example, for the concept health users were able to rate how well the product conveyed feelings of unhealthy and healthy, on a scale from one to four, where one was unhealthy and four was healthy. Just as in iteration 1, the primary concepts were flexibility, adaptability, and health. In iteration 2, flexibility was represented by Rigid - Adaptable, Clumsy - Flexible, and Awkward - Formable. Adaptability was represented by Unengaged - Personal, and Troublesome - Accommodating. Health, in turn, was represented by Unhealthy - Healthy.

As such, the concepts available were:

1. Rigid - Adaptable
2. Clumsy - Flexible

19
3. Awkward - Formable
4. Unengaged - Personal
5. Unhealthy - Healthy
6. Troublesome - Accommodating

The scale was introduced to try and assess how well a certain concept (such as health) is conveyed in the design, and give a greater variety of opportunities for data analysis. Users were still (as in iteration 1) required to choose and rate exactly 3 themes that according to them were the most descriptive of the prototype.

*Figure 6. Screenshot of web-based DPR (Swedish).*

The above figure shows how the web-based DPR looked in iteration 2. Only 2 out of 6 concepts are displayed, but the pattern is recurring for all concepts.

### 3.3.1.3 Results

All users were interviewed within 6 minutes after completing the PRC test, except in one case where contact occurred first after 73 minutes. There was also one case where the user had misunderstood the scoring system, and had rated what was important to him/her, rather than rating the product as it was experienced by him/her. As such, provided below is a table (table 3) adjusted excluding data collected from the latter user, who misunderstood the scoring system:
Table 3. Adjusted concept scores from iteration 2 DPR testing.

<table>
<thead>
<tr>
<th>Adjusted scores</th>
<th>Stel-Anpassbar</th>
<th>Klumpig-Flexibel</th>
<th>Ansträngd-Formbar</th>
<th>Ongagerad-Personlig</th>
<th>Ohälsosam-Hälsosam</th>
<th>Besvärlig-Tillmötesgående</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Concepts were scored on a scale from 1 to 4. An example of how scores were interpreted is shown in the table below (table 4):

Table 4. Score interpretation using the Rigid-Adaptable concept as an example.

<table>
<thead>
<tr>
<th>Rigid - Adaptable</th>
<th>Score</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Very rigid</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Somewhat rigid</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Somewhat adaptable</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Very adaptable</td>
</tr>
</tbody>
</table>

That means that, using row 1 of table 3 as an example, the product was thought to be very adaptable, very personal, and very healthy. Other concept scores are interpreted in the same way, with 1 and 4 being extremes at each end of the negative to positive spectrum.

Some of the most recurring themes found during interviews are listed below in table 5, note that data collected in the one interview that took place 73 minutes after the user completed the testing has been excluded:
### Table 5. Recurring themes from iteration 2 interviews.

<table>
<thead>
<tr>
<th>Profiles</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy to adapt according to taste</td>
<td>2</td>
</tr>
<tr>
<td>Nothing missing from profiles</td>
<td>2</td>
</tr>
<tr>
<td>Wanted more alternatives in profiles</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DPR testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey words were adequate</td>
<td>2</td>
</tr>
<tr>
<td>Survey words were not fully adequate</td>
<td>2</td>
</tr>
<tr>
<td><strong>Survey was easy to understand</strong></td>
<td>3</td>
</tr>
<tr>
<td>Survey was hard to understand</td>
<td>2</td>
</tr>
<tr>
<td><strong>Survey word quantity was adequate</strong></td>
<td>3</td>
</tr>
<tr>
<td>Survey words were alike</td>
<td>2</td>
</tr>
<tr>
<td>Easy to select concepts</td>
<td>2</td>
</tr>
</tbody>
</table>

Themes have been freely translated from Swedish. The themes are sorted based on what area they were in referral to. Themes regarding the prototype are listed under *profiles*, while themes referring to DPR testing is listed under *DPR testing*. The most recurring themes are highlighted in bold font.

3.3.1.4 Analysis

As opposed to iteration 1, iteration 2 did not suffer from many dropouts, as 6 out of 7 users successfully completed the test. The test was not conducted without complications, however, as one user was not interviewed until roughly 73 minutes after answering the DPR. Another user misinterpreted the survey and answered it according to what he/she thought were the most important aspects of a prototype of this sort. As mentioned in results, this led to the exclusion of interview data in the first case, where the interview was conducted 73 minutes after testing. DPR data was kept, however, since the DPR test occurred in conjunction with the prototype testing. In the second case, where the user had misinterpreted the survey, interview data (themes) was kept, but DPR data was discarded.

There are two aspects that are important to consider when analyzing the results using DPR. Concept selection, and concept scores. Ideally, users would select the most important concepts for the prototype, and give them all the maximum score (four). In iteration 2, the available concepts were:

1. Rigid - Adaptable
2. Clumsy - Flexible
3. Awkward - Formable
4. Unengaged - Personal
5. Unhealthy - Healthy
6. Troublesome - Accommodating

Each one of these are related to the three primary concepts (flexibility, adaptability, and health) derived from the UVP, which was "A flexible and adaptable way to eat healthy" (freely translated from Swedish). The three most important concepts for this iteration would be the directly translated ones, meaning Clumsy - Flexible, Rigid - Adaptable, and Unhealthy - Healthy. In other words, using this method, a 'perfect' design in relation to the UVP would be a design where all users selected Clumsy - Flexible, Rigid - Adaptable, and Unhealthy - Healthy, and scored 4 for all of them. There is, however, a problem with such a hard drawn line between a measured optimal design and everything else. For example, if a user would select one of the alternatives for flexibility or adaptability (such as Troublesome - Accommodating instead of Rigid - Adaptable), the design would be less than optimal even though the user may have referred to the same notion (that it caters to the needs of the individual). However, if we loosen the requirements on the idea of an optimal design to include the alternatives, the selection process would essentially be rendered meaningless, and we would have to rely solely on the score. This leads us to the first problem with the method in iteration 2 - that some concepts are too alike (see table 5), and that indirectly renders the selection process meaningless.

This, in turn, results in the surfacing of another (greater) problem - the UVP to DRP concept mapping technique used was inadequate.

It is believed the reasons for these errors were:

- Partly due to an error in interpretation and subsequently translation.
  Had the interpretation been correct, the UVP "A flexible and adaptable way to eat healthy" in the case study would not have contained the word flexible. Instead, something along the lines of smooth or perhaps easy would have been more appropriate, making "An easy and adaptable way to eat healthy" the translated result. This would in turn have resulted in fewer synonyms to flexible and adaptable, as flexible would not have been present at all.

- Using synonyms in DPR concept selection.
  For example, one user solved the problem by avoiding the two concepts that was considered indifferent from each other. The original idea behind using different phrasings for the same concept was to 'cover all bases' from an interpretation standpoint, but evidently that may not have been the most effective way of achieving that. At the very least, it had the unintended side-effect of causing problems with concept selection.

The aforementioned two problems must be solved in order to move forward with the DPR method.
Starting with the more important mapping problem, moving away from synonyms of the same concept seems crucial given the data from iteration 2. Furthermore, the ratio between the concept words for each primary concept was uneven (recall for example that the notion of health was only represented by unhealthy - healthy). It is uncertain exactly how this imbalanced ratio affected the results. My belief is that the following steps can be taken in future iterations in order to solve the problems uncovered in iteration 2.

Introduce another step in UVP to DPR concept translation, in the deriving of primary concepts prior to forming DPR concepts. Primary concepts would be the themes of the UVP. It would be important in this stage to eliminate synonyms, and only keep one concept per theme. This would facilitate a fairer ratio between the subsequent DPR concepts. These primary concepts are usually easily distinguishable, or even directly translated, from the UVP itself. Using "A flexible and adaptable way to eat healthy", we can distinguish adaptable (direct translation) and healthy (direct translation). It is important that, as flexible and adaptable can be thought of as referring to the same notion, only one of them is selected as a primary concept. The reason being that the primary concepts guide the ratio between DPR concepts, and having two similar primary concepts would likely result in DPR concepts that are too much alike each other. Another UVP example where direct translation is not possible in the same way would be "Spend More Time Building Versus Planning Your Business - The faster, more effective way to communicate your business model", where we can distinguish clever (from spending resources in the right place) and time-saving (from being fast and effective). One important aspect of deriving primary concepts before moving on to DPR concepts is that it would be easier to avoid synonyms in DPR concepts, as well as maintaining a fair ratio between them.

Avoid using synonyms for DPR concepts.

Providing short descriptions of each DPR concept. This may prove useful in facilitating the same or similar interpretation of words across subjects.

As such, the following method is proposed.

4. Proposed Method

This section outlines the method proposed based on the results and analyses within the scope of this thesis. The method suggested in this thesis contain the following elements:

- How to translate the UVP into DPR concept words
- Expanding the alternatives
4.1 Translating the UVP to Key DPR Concept Words

Translating a UVP to DPR concepts involves three steps. Figure 7 shows an outline of the process.

The first step in translation is to derive primary concepts. These primary concepts are the essence, or themes, of the UVP. These concepts, in turn, direct the forming and phrasing of the key DPR concepts. Selection of these DPR concepts indicate design success, as they can be thought of as representative of the UVP. In other words, using this example, if all users picked *Dumb - Clever*, and *Inefficient - Efficient*, and scored four for them both, it would indicate an optimal design. However, only allowing selection of concepts that are beneficial towards the assessment of the design is not enough.

4.2 Expanding the Alternatives

More alternatives are needed, as users must be allowed to completely avoid the key concepts, should they want to. As such, it is recommended to have at least as many alternative concepts available for selection as you have key concepts. There are three things to keep in mind when creating these alternative concepts. First, they must be applicable to the design in some way. Using something like *Hard - Soft* for a software interface would not make much sense. Second, they cannot be synonyms to key concepts, or refer to the same notions. For example, if there is a key concept *Dumb - Clever*, using *Stupid - Smart* is not a good idea. The alternative concepts are not to be confused with the key concepts in the eyes of the users.

It is possible to use the concept pool established in PRC to select alternative concepts, as long as the abovementioned criteria is met.
4.3 Directed Product Reaction (DPR)

The DPR test itself consists of the following elements:

1. Acquaint the users with the product.
2. Allow users to select DPR concepts.
3. Conduct qualitative interview.
4. Analyze the data.
5. Iterate.

The following sections explain the above steps more in-depth.

Acquaint the Users With the Product
Once the key concepts and alternative concepts have been established, they can be incorporated into the DPR test. DPR can incorporate either a scenario, tasks, or both, in order to acquaint the users with the product being tested. It is important that the users have tested the product in order to gain a basic understanding of how it works and to form an opinion. If users have conducted usability testing prior to DPR, it is not necessary to perform this step.

DPR Concept Selection
As user experience is believed to be very much momentary, it is important that users move on to the DPR concept selection as soon as possible after testing the product. It is recommended to collect information on the time expended between prototype testing and DPR concept selection, to make sure that concept selection occur as soon as possible after prototype testing. The selection process involves selecting exactly X number of the provided concepts, and deciding a score of one to four for each selected concept. The number of concepts required for selection is determined by the number of key concepts provided (see 4.4 for an example). Each concept should be provided with a short description, in order to avoid any deviating interpretations in the minds of the users. The test can be carried out either in person or via other mediums (such as the internet). This case study involved a web-based approach. If conducting the test in person, it is important to leave the room while the user performs the selection of concepts, as to allow the user freedom to consider, select, and score concepts in private.

Qualitative Interview
Likely, semi-structured interviews are the best way to approach the interview step. Prepare a script prior to the test, focusing on questions regarding the instructions (make sure the user did not misinterpret the instructions), their selection (what, in the design, prompted their selection of the particular concepts), and if there is anything they would like to add or say that could not be conveyed via the concept selection. After completion of the DPR concept selection, the qualitative interview is
carried out. It is recommended that the researcher spends a moment to review the DPR selection result prior to conducting the interview, and have the results influence the questions in the interview. The qualitative interview is very important, not only to give the users the opportunity to express themselves freely about the product, but also to tie up any loose ends or get answers to any questions that has arisen as a result of the DPR selection. There are a lot of things to consider when conducting qualitative interviews, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* by John Creswell (2012) contain a section of valuable material. Also, see chapter 5.3.2 "Shadowing" and Emphasis on Interviews for one more reason why interviewing is important in DPR, specifically.

**Analyzing and Interpreting Results**

Analyzing the data is done in three steps. First, analyze and interpret the DPR selection results. If users have selected the key concepts, that indicates that they stand out more than the alternative concepts. This may be for positive and negative reasons, whichever it is become apparent when looking at the score. Table 5 below shows how to interpret concept scores using *Dumb - Clever* as an example.

<table>
<thead>
<tr>
<th>Dumb - Clever</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Very dumb</td>
</tr>
<tr>
<td>2</td>
<td>Somewhat dumb</td>
</tr>
<tr>
<td>3</td>
<td>Somewhat clever</td>
</tr>
<tr>
<td>4</td>
<td>Very clever</td>
</tr>
</tbody>
</table>

Analyzing the concept scores is the second step of data analysis. If all key concepts have been selected and scored four, in the eyes of the users, the design is optimal in regards to the values of the UVP. The resulting score from each DPR concept should be contrasted against the maximum score possible (four, in this case).

Finally, analyze the interview data. This is done by means of a light thematic analysis, where recurring themes and topics are distinguished from individual opinions. These should (given that the interview was successful) provide information as to how specific design features affected the scores, and therefore also provide clues on what to work on in order to improve the scores. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* by John Creswell (2012) contain more detailed information on how to analyze data from qualitative interviews.
Iterate
Once the data analysis is complete, the data collected should support another iteration, whereby the process is repeated. If the data does not support another iteration, there may be a need to extend the sample size in order to collect more data.

4.4 Applying DPR in Practice (example)
In this section, an example of how the abovementioned proposed method can be applied is presented using a fictive example product. The example described is barebones, but should provide a better understanding of how to apply DPR.

Product and Users
The example product is a mobile application, used to plan bus trips within a city. It is mostly designed and marketed towards younger smart phone users between 15 and 30 in age. The UVP used is "Plan Your Trip Reliably, Efficiently, and Effortlessly".

Deriving the Primary Concepts
The first step is to derive the primary concepts from the UVP. As you may remember, they are meant to represent the essence of the UVP. In this example, they are Reliability, Efficiency, and Effort. It is important that primary concepts are not similar enough to be mistaken for one another. For example, if the UVP would have been "Plan Your Trip Reliably, Simply, and Effortlessly", one could have derived the primary concepts Reliability, Simplicity, and Effort. In this case, both Simplicity and Effort could refer to the same notion, namely the level of hardship involved in planning a trip using the application. As such, only one of them is used as a primary concept, and either one may be chosen.

Forming Concepts for Selection
In this case, we derived the primary concepts Reliability, Efficiency, and Effort. These can be both negative and positive. As such, they are already possible to use as concepts in the selection process. All that is needed is to provide both the negative to positive extremes in adjective form. For this example, that would mean Unreliable - Reliable, Inefficient - Efficient, and Difficult - Easy. These are now our key concepts, as they are derived from the primary concepts, which in turn represent the essence of the UVP. As such, they are the concepts with which we determine success. As the number of key concepts (three, in our case) determine the number of required concepts for selection, to make the selection process meaningful, we must provide alternatives to these key concepts.

Expanding the Alternatives
As you may remember, the number of alternative concepts is determined by the number of key concepts. Three, in our case. The reason is that users should be able to avoid selecting any of the key concepts.
concepts, should they want to. The alternatives should be applicable to the design, and should not be easily confused with the key concepts. In our case, we could use Rigid - Adaptable, Busy - Clear, and Boring - Fun. These are all applicable to the product in some way, and are unlikely to be confused neither with each other nor with the key concepts.

Testing
The first user participating, Jennifer, is asked to perform a couple of tasks involving the prototyped product. She is asked to plan a trip, change a planned trip, and check the ticket prices. These tasks are designed to give the users the opportunity to form an opinion of the product. The design of tasks may vary, as they are simply there to make the users acquainted with the prototyped product. She completes the tasks after a session lasting for 10 minutes.

Selection Process
Immediately after completing the tasks, Jennifer is asked to start the concept selection. The concepts available are the key concepts and alternative concepts derived earlier in the process, namely Unreliable - Reliable, Inefficient - Efficient, Difficult - Easy, Rigid - Adaptable, Busy - Clear, and Boring - Fun. All of them are written on separate cards, along with a scale of 1 to 4. There is also a short description on each card, explaining how the concept manifests in the design. For Unreliable - Reliable, it reads "the grade at which you believe the application provides information that is useful and correct". For Inefficient - Efficient, it reads "the expected amount of time saved by using this application". Difficult - Easy reads "how easy or difficult it was to use the application". These descriptions are provided for all concepts available for selection, and explains how to interpret the concept in relation to the product. She must select exactly three concepts, and score them on a scale from 1 (most extreme negative) to 4 (most extreme positive). As soon as it is confirmed that Jennifer has understood the instructions, the researcher leaves the room, closing the door behind him.

Interview
When Jennifer is done, she opens the door, and the researcher reenters the room. The researcher collects the concept cards, while quickly eyeing and noting the concepts that have been scored. Subsequently, the researcher initializes the interview. First and foremost, the goal is to make sure that Jennifer has completed the selection correctly. The researcher asks her if she found the test or selection difficult, and if so, why. She seems to have completed the test and selection without any issues at all, and as such, the researcher shifts focus towards her selected concepts and concept scores. Jennifer has selected the concepts Inefficient - Efficient with a score of 3, Difficult - Easy with a score of 2, and Busy - Clear with a score of 2. She explains that planning your trip using this
application felt a lot more efficient than planning it manually, but that the interface seemed somewhat busy and cluttered, and that it was a bit hard to use as a result.

**Analyze and Interpret**

The selection and scoring was

*Inefficient - Efficient*, with a score of 3

*Difficult - Easy*, with a score of 2

*Busy - Clear*, with a score of 2

which means that the prototype was perceived as somewhat efficient, but also somewhat difficult and somewhat busy. 2 out of 3 key concepts were selected, which means that they stand out compared to the others. The scores indicate that it was for a negative reason, however. This means that more effort must be put in to improve the feeling regarding the *Effort* concept. Clues to how this can be done is often found in the interviews, such as Jennifer saying that the interface was somewhat cluttered, which may have resulted in a product more difficult to use.

## 5. Discussion

### 5.1 Limitations of DPR in Iteration 1 and 2

There were a couple of issues outlined in the respective analyses of data from iteration 1 and 2, and while they were discussed then and there, it may prove useful to have a summary of the issues uncovered.

**If the tested prototype is part of (or a feature of) a larger website, using tasks involving the larger website may result in unreliable data.** This became apparent in iteration 1, where 7 out of 8 tasks were situated outside of the prototyped feature. This resulted in users losing focus on the prototype, and speaking about the website rather than the prototype. As such, the data regarding the design became questionable, as separating opinions of the prototype from opinions of the larger website was complicated. In iteration 2, this issue was countered by removing all tasks situated outside the prototype, and an increased emphasis on the prototype in the instructions. This resulted in data that almost exclusively was in referral to the prototype rather than the website as a whole, but some users still expressed some opinions regarding the website as a whole during the interview. If taken to the extreme, one could simply not mention the larger website at all, which would further facilitate prototype isolation. However, it is uncertain what effect this would have on DPR scores, as the
Users opting out of testing at the last minute. This proved devastating for the wizard prototype in iteration 1, as 3 out of 5 users chose not to participate in the end. This remains an issue for DPR as it does for any method with a limited number of users.

An extended amount of time passing between prototype test and DPR concept selection. As mentioned multiple times, UX is thought to be momentary. Therefore, conducting the selection and interviews as soon as possible after users have been exposed to the prototype is important. In iteration 2, however, there were one user who completed the DPR selection 73 minutes after testing the prototype. For that case, it was decided to not include the data collected from that particular user in the results and subsequent analysis, as the quality of data regarding the experience may deteriorate over time.

Misunderstanding instructions and/or instructions are inadequate. Users may misunderstand the instructions, resulting in data of varying quality (depending on the misunderstanding). In iteration 2, one instance involved a user completely misinterpreting the point of selecting DPR concepts, which resulted in DPR scores that were not in referral to neither prototype nor website. This data was not included in the results nor subsequent analysis.

Interpretation of concepts. How users interpret DPR concepts is important, as it very much influences their selection and scoring. In iteration 1 and 2, concepts were similar to each other. This was done to avoid users disregarding concepts due to poor choice of words. Unfortunately, this resulted in a spread of data, as some concepts could be thought of as referring to the same notions. In extreme cases, this could render the selection process meaningless. To prevent this in further iterations of DPR, the proposed method contain short descriptions coupled with each DPR concept to facilitate the same interpretation between designers and users alike.

Culture-specific data. As mentioned in 1.7 Delimitations, the data was collected in a specific sub-culture (i.e. the target customer segment) in a specific culture (Sweden). It is uncertain if the results could be reproduced in another part of the world, even if we disregard the subjective and contextual-sensitive nature of UX itself. As UX is reflected in the data collected, this becomes a strength of the method, as context and culture is very important when designing HCI-products for a specific target customer segment. However, it can also be a weakness if the target customer segment is too broad, as testing such a large population would become problematic.
High-fidelity prototypes only. DPR has been tested and designed to work using high-fidelity prototypes only, as would be the case in a Lean Startup at the stage in development where one would employ the DPR method. DPR may or may not prove useful when working with low-fidelity prototypes, but that remains uncertain at this point.

End goals not presented (unless part of UVP). The end goals of using a product is currently not represented in the DPR concepts, unless they are an explicit part of the UVP and therefore warrant their own individual concepts. A side-effect of this is that products of various types can produce similar or identical DPR concepts. At present, this is believed not to be a problem, as the same concepts can be used to describe a broad selection of products. Both a software HCI-product and a cell phone may be described via the concept good looking, for example. The same concept would be used, but would refer to different notions, as a good looking cell phone most likely is not good looking in the same sense a software product is. It ties in with how concepts are to be interpreted, which in the proposed method is handled by supplying descriptions of each concept as to facilitate similar interpretations across the sample users and designers. It is unclear at present whether this is an issue or not, and may be something worth investigating in future iterations.

The number of selected concepts is arbitrary. Currently, users are expected to select three alternatives, which has worked well so far. However, the required number of selected concepts should reflect the number of key concepts available. There is no point in requiring users to select three concepts if there are only two key concepts available. As such, the proposed method contains an untested solution to this (not yet encountered) problem, where the number of concepts required for selection is the same as the number of key concepts available.

5.2 Discrepancies in UVP Interpretation Between User and Designer
Associations between words are culturally as well as socially constructed (Foley, 1997) (Jiang, 2000), meaning that whatever conceptual words designers associate with the UVP may or may not be shared with a potential customer base. It may even differ greatly among individuals within the same cultural groups and sub-groups. In other words, words have different meaning to different people, and people may therefore interpret the UVP differently, and as such expect the design to behave in differing ways. It is important to realize this potential discrepancy to attempt to avoid discarding or promoting design on faulty grounds.

As an example, the UVP "Turn your users into passionate customers - Customer Lifecycle Management Software" might translate into the concept words engaging and/or immersive.
Consider this situation, while developers may believe immersive is the word that best describe the
concepts of the UVP (and therefore should be the word that measure and determine success), the users instead believe *engaging* is the word most associated with the UVP. In short, users could interpret the UVP differently from designers. Measuring success only using the word *immersive* in this situation would most likely be a mistake. Users would potentially have selected the concept word *engaging* in their top three, but it is not included in the test since developers were not aware that *engaging* is a concept word that reflects the UVP, and only include *immersive* in the test. As a worst case scenario, this may result in a good design being discarded due to poor concept word mapping.

The case study involved the mapping words from the UVP as direct as possible to Directed Product Reaction concept words. Some synonyms of these direct translations were also provided. That strategy proved to be less than optimal, and some issues related to it was uncovered. Furthermore, that sort of mapping is largely dependent on the format of the UVP, as some UVPs lend themselves well to this strategy while some do not. This strategy also relies on the designers idea of what words and synonyms are used by the customer segment involved in the testing, but it can be used in conjunction with frequency data on what words are most commonly used by that customer segment in an attempt to circumvent this issue.

### 5.3 Validity and Reliability in DPR

In order to assess if a method such as this actually measures how far away the current design is in relation to the UVP, we must first know the truth about that exact relation. We must know for certain that Design X fulfills the UVP Y to 50%, and in order to obtain that information we must have a method that exactly measures that. However, if we had such a method, neither of us would be sitting here with this thesis in front of us. In other words, we cannot know the truth if it cannot be measured, and we cannot tell if we measure the truth unless we know the truth first. Attempting to solve this issue here is most definitely outside the scope of this thesis, as that would be an immense feat in itself because of its paradox nature. However, the following paragraphs contain a discussion of what that question entails in relation to the method suggested in this article.

It may not make sense to consider validity and reliability in the traditional sense, but rather, to talk about trustworthiness (Seale, 1999). Even though DPR is not a pure qualitative method (being more of a mixed method) it seems fitting to discuss validity from a qualitative perspective rather than a quantitative one. According to Seale (citing Lincoln and Guba, 1985, p. 294), there are four primary qualities of a trustworthy qualitative study: credibility, transferability, dependability, and confirmability.
Credibility refers to how believable the results of the qualitative method is to the studied subjects. In the case study, and DPR as a method, credibility is facilitated via the means of interviews. The data collected are direct subjective accounts of experience provided by the subjects themselves, and therefore supports a level of credibility that can be assumed to be high. Higher levels of credibility could possibly be achieved by allowing subjects to view and confirm the collected interview data prior to analyzing.

Transferability refers to how well the results from the sample can be applied to the population. In the case study, for example, participating users corresponded to criteria based on earlier developed personas that outlined a spectrum of typical users. This aspect is one of the most important ones in terms of DPR as a method since the design is most likely targeted to a broader audience than the sample subjects. As such, increasing the levels of transferability in future iterations of DPR must remain an important task.

Dependability, in turn, refers to how well the method describes the context within which research is conducted. It is also important to describe what implications this may have on the procedure and results. Within DPR, dependability is mostly influenced by how well the researcher describes the prototype(s) and/or the setting in which the test took place. In the case study, this was done by providing screenshots and short descriptions of the prototypes tested by the subjects. No effort was made to describe the subjects immediate surroundings and context, as they had the freedom to conduct the test from any location they desired.

Finally, confirmability. The notion of confirmability within DPR is complicated, the reason being that many scientists (Law, et al., 2009) (Desmet & Hekkert, 2007) seem to believe that UX is a momentary, subjective, and highly context-sensitive. As such, ensuring confirmability in a method such as DPR is seemingly impossible. As the subjective user experience depends on many variables that cannot be controlled (such as mood and other contextual parameters), it is highly uncertain if the results could be confirmed or corroborated by other researchers, even if using the same subjects and an identical setup.

5.4 Further Research

5.4.1 UX Framework

As mentioned in 2.4 Usability and UX, what DPR can be said to measure is hedonic quality (Hassenzahl, 2008). Hedonic quality is the prime facilitator of be-goals, and as such is the most important aspect of user experience (UX). While the concept of hedonic quality provides a useful way of speaking and reasoning about the aspects facilitating good UX, it does not provide any structure or
categorization of different types of aspects of UX. My belief is that such a structure can prove useful as a framework for translating a UVP to DPR concepts. A UVP can essentially be derived or deconstructed into a number of notions and ideas about design. An example would be the case study conducted in this thesis, where the UVP "A flexible and adaptable way to eat healthy" can be thought of as containing the primary concepts flexibility, adaptability, and health. These concepts would in turn map against DPR concept words, which in turn could be categorized as different aspects of UX. This categorization would serve two purposes, First, one could include a specific number of concepts for each category, to make sure that the ratio between concepts remain fair. Second, it would make for a method that is easier to generalize than the one used in iteration 2 of the case study.

There exists a framework that could prove to be of assistance. A framework presented by Desmet and Hekkert (2007) as a "general framework for product experience that applies to all affective responses that can be experienced in human-product interaction". This framework could serve as the basis for categorizing DPR concepts, as well as regulating the ratio between concepts. The three levels presented in the framework are aesthetic experience, experience of meaning, and emotional experience.

Aesthetic experience refers to a products ability to please our sensory modalities. Examples include beautiful, pretty, and neat.

Experience of meaning include the assignment of personality traits (or other expressive characteristics) to a product, as well as assess the personal and symbolic value of a product. Examples include luxurious, feminine, understandable, masculine, confident, relaxed, and busy.

Lastly, emotional experience refers to what you would typically expect when talking about emotions. Examples include desirable, fun, frustrating, inspiring, or disappointing.

A framework such as this may prove useful when determining what, and how many, concepts to include in DPR. Also, using this framework as a mould for translating UVP to concepts, users would be allowed to express elicited emotions from all possible aspects of the experience. Focusing on the structure of UX in potential future development of DPR seems appropriate, and is recommended to consider in future research.

5.4.2 "Shadowing" and Emphasis on Interviews

The creation of alternative concepts to fill out the DPR concept pool may be more complicated than what it is made out to be in the proposed method. There may be a third aspect to it not mentioned in 4.2 Expanding the Alternatives. Consider this scenario. The key concepts in DPR are Dumb - Clever, and Inefficient - Efficient. The alternative concepts are Busy - Clear, Ugly - Beautiful, and Taught -
Intuitive. All of these concepts can exist in a design simultaneously. This means that users can experience the design as mostly clear, beautiful, and intuitive, and therefore select those three concepts in DPR concept selection. However, that does not mean that the design is not clever or efficient, but rather that it is not more clever or efficient than clear, beautiful, or intuitive, and the latter three shadow the first two. This is intentionally inherent within the method, and the design focus should be to convey the values of the UVP (the key DPR concepts). As such, these values should be what stands out in the eyes of the users. The problem arises in that it is hard to tell how far off clever or efficient is from standing out, as no data is collected on anything but the three most prominent concepts. A possible solution may be to remove the restriction of selecting and scoring exactly three concepts. That would however result in a method starting to resemble the UX cousin of SUS (Brooke, 1986), which may or may not be a good thing. Another solution, possibly easier and closer to hand, would be to tackle this issue during interviews. The researcher has to try to collect detailed data on what prompted the user to select the way he/she did. With this, the need for skilled and attentive interviewing in further research is emphasized, as a good interview is imperative to collect the data not collected by the DPR concept selection.

The semi-structured interview script used in this case study is provided as an appendix (appendix A).

5.4.3 Future iterations of DPR
Future iterations of DPR would do well attempt the method suggested in section 4, Proposed Method, as it contains some elements not yet tested, such as deriving primary concepts prior to DPR concepts. The method itself needs more work, and is far from finished. Some issues and ideas have been discussed in this section, but undoubtedly it is only the tip of the iceberg.

6. Conclusions
Looking back to the research questions for this thesis, they were:

1. How can PRC be adapted to assess whether a design is approaching the values of a UVP or not?
2. In what way can this adaptation be applied in practice?

The method proposed in section 4, Proposed Method, is the suggested answer to question one. The method, Directed Product Reaction, is based on Product Reaction Cards, and consists of a number of steps. These steps involve the translation (or mapping) of a UVP to primary concepts, which in turn direct the creation of concept words. These concept words are then subject to a selection process, in where users select the three concepts that they find the most representative of the product. The
concepts are then scored on a scale, from a negative to positive impression, giving an indication of the product user experience. This method is without a doubt not the only answer to the research questions in this thesis, not the final answer, and possibly not the best answer. What is meant by this is that there are most likely numerous ways to adapt PRC towards that end, and the method suggested in this thesis may not be the best way. Furthermore, more work is needed to realize the potential of a method such as DPR. A potential held not only within the context of Lean Startup, but possibly within HCI-design as a whole.

Question 2 was answered through the completion of the case study for this thesis, where DPR was used in conjunction with high-fidelity prototypes to assess how well they conveyed the values of a UVP. See chapter 3, Case Study, for a detailed explanation.
References


Åberg, J. (2013). Questions regarding the course 729G30, Bachelor Thesis. [e-mail] (Personal communication, 22 Feb 2013).
Appendixes

Appendix A, interview script

Appendix A is the script used in the qualitative interviews of the case study, freely translated from Swedish to English.

Did you find the test difficult?

Did you miss anything in the profile selector?

What was it about the prototype that prompted you to select and score the way you did in the survey? Were there any details that especially contributed to your answer? Why?

Did you feel like the alternatives available were sufficient? Were you able to express your opinions?

How did you view the amount of words available? Too many? Too few?

Was it difficult to select concepts?

Did you find the test taking too long to complete?

Miscellaneous
Appendix B, abbreviation glossary

UVP - Unique Value Proposition. A phrase used internally (within the production team) to guide and direct design towards a common goal, and externally (with consumers) to gain their attention.

PRC - Product Reaction Cards. A method for testing design developed at Microsoft in 2002 by Joey Benedek and Trish Miner.

DPR - Directed Product Reaction. A heavily adapted version of PRC attempting to measure intangibles in a lean setting in order to guide design. Presented as a response to one of the research questions for this thesis.

UX - User Experience. A momentary, primarily evaluative feeling (good-bad) while interacting with a product or service.

MVP - Minimum Viable Product. A barebones version of a software product launched as soon as possible to early adopters (akin to alpha testers).
**Upphovsrätt**

Detta dokument hålls tillgängligt på Internet – eller dess framtidiga ersättare – från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnt som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.


**Copyright**

The publishers will keep this document online on the Internet – or its possible replacement – from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: [http://www.ep.liu.se/](http://www.ep.liu.se/).

© Max Liedström Kvelland