Enhancing usefulness : connecting utility and usability

an explorative study towards integrating consumer/market research and usability testing

Valk, K.

Award date:
2007

APPENDICES

Enhancing Usefulness: Connecting Utility and Usability

-An explorative study towards integrating Consumer/Market research and Usability testing-

MSc Graduation Thesis
Eindhoven, July 2007

Krista Valk
Student of Industrial Engineering and Management Science
Eindhoven University of Technology
K.Valk@student.tue.nl
510916

Supervisors

1st Supervisor:	dr. ir. P.H. den Ouden Department Technology Management Quality and Reliability Engineering Group
2nd Supervisor:	prof. dr. ir. A.C. Brombacher Department Technology Management Quality and Reliability Engineering Group

Eindhoven University of Technology
Department of Technology Management
Sub-department Quality and Reliability Engineering

Table of Contents

I. Overview Usability methods

method name	lifecycle stage	users needed	evaluators needed	time required	required expertise	main advantage	main disadvantage
Webtesting software (logging actual use)	Final testing, follow-up studies.	20+	none	High	Low	Finds highly used (or unused) features. Can run continuously.	No content and aesthetics problems / satisfaction. Analysis programs needed for huge mass of data.
End user testing	Final testing	20+	1+	Medium	High	Assessing the system in action	Costly and time consuming
User feedback	Follow-up studies	Hundreds	None	High	Low	Tracks changes in user requirements and views.	Special organization needed to handle replies.
Interviews	Task analysis	5	1+	High	Medium	Flexible, in-depth attitude and experience probing.	Time consuming. Hard to analyse and compare.
Questionnaires	Task analysis, follow-up studies	30+	1	Low	Low	Finds subjective user preferences. Easy to repeat.	Pilot work needed (to prevent misunderstandings).
Observation	Task analysis, follow-up studies	3+	1+	Medium	Medium	Ecological validity; reveals users' real tasks. Suggests functions and features.	Appointments hard to set up. No experimenter control.
Focus groups	Task analysis, user involvement	6 – 9 per group	1+	Medium	High	Spontaneous reactions and group dynamics.	Hard to analyse. Low validity.
Online focus groups	Task analysis, user involvement	6 – 9 per group	1+	Medium	High	Diverse population, flexibility	Recruiting individuals is hard
Prototyping	After the core of the application is defined (iterative)	3+	1	Low, Medium & High	Low	Prototyping can validate customer requirements early in the product cycle	A prototype may give the wrong appearance of performance

GOMS	Requirements-specification phase	1+	1	High	High (Medium)	Usability information when the system has not been developed	Considerable time and effort need to be invested to develop a formal system model
Thinking aloud	Iterative design, formative evaluation	3 – 5	1	High	Medium	Pinpoints user misconceptions. Cheap test.	Unnatural for users. Hard to expert users to verbalize.
Heuristic evaluation	Early design, 'inner cycle' of iterative design	None	3+	Low	Medium	Finds individual usability problems. can address expert user issues	Does not involve real users, so does not find 'surprises' related to their needs.
Guideline reviews	Early design	None	1+	Low	High	Applied early in design, without user testing	1.000 guidelines and require high degree of expertise
Pluralistic walkthrough	Early	1+	2+	Medium	Low, Medium and High	Effective for evaluating the learnability	No evaluation of interfaces in daily use
Consistency and Standards inspections	Early	None	3+	Low	Medium	Relationship between cost and benefits	Subjective, due to heavy dependence on inspectors' skills
Cognitive walkthrough	All	None	3+	Medium	High	Focus on evaluating a design for ease of learning	Hard to learn and to apply
Formal usability inspection	Early	None	3+	Low	Medium	Effective and efficient	Subjective, due to heavy dependence on inspectors' skills
Feature inspections	Early	None	3+	Low	Medium	can also involve the design of the function	Subjective, due to heavy dependence on inspectors' skills

II. Uncertainty

In Chapter 2, two different types of uncertainty are presented: product technology uncertainty and market uncertainty. In the table below the different aspects of both types are elaborated.

Table II.1 Subdivision of product technology and market uncertainty [adapted from Kei 05]

Product technology uncertainty	**Market uncertainty**
Uncertainty about: 1. Ability of the (product) technology to meet the intended functions or a translation of those functions into lower level requirements. → **Ability** 2. Compatibility of the (product) technology with existing technological standards, use environment and/or knowledge of the consumer → **Compatibility** 3. Performance and attractiveness of the (product) technology compared to other (future) technology. → **Parity**	Uncertainty about: 1. What the target consumer and user group are. → **Target group** 2. What the consumer and the user want and need. → **Wants&Needs** 3. The changes in the consumer needs/attitudes environment during the product development. → **Changes** 4. Consumer use → **Use** 5. User acceptance of the product a. ease of use → **Ease Of Use** b. relative advantage → **Rel. Adv.** c. product communication → **Communication**

In order to describe the deliverables of the two concepts Consumer/market research and Usability testing, these two concepts are related to the different aspects of uncertainty as shown in the table above.
In this appendix, a profound assessment of the methods within the two methodologies is provided.

II.1 Consumer/market research

The effect of the four techniques of Consumer/market research; Concept testing, Product-use testing, Pre-Market testing, and Market testing in reducing uncertainty are elaborated in the following sub sections.

II.1.1 Concept testing

Kahn [Kah 05] defines Concept testing as follows:

Current and/or potential consumers evaluate a new product concept and give their opinions on whether the concept is something that they might have interest in and would likely buy. The aim of Concept testing is to prove the new product concept.

According to Ozer [Oze 99], commonly used methodologies are: analogies, expert opinions, intentions, multi-attribute models, focus groups, and scenario analysis/information acceleration.

Reduction of product technology uncertainty:
Ability, Compatibility, Parity: Concept testing is a test that is meant to prove the new product concept. This test is conducted in the concept phase. The translation into lower level requirements is not yet done, therefore the Ability, Compatibility and Parity uncertainty can not be reduced.

Reduction of market uncertainty
Target Group: In all methods for consumer testing that Ozer [Oze 99] mentions, the target group and customer are taken as a starting point and from that base, the product concept is evaluated. This means that the target consumer and user group is not directly defined by means of Concept testing. However it does become clear whether the intended target group is interested in the concept.

Wants&Needs: A survey [Dah 01, Ulr 00] can reveal needs of a consumer, but it is ineffective in revealing unanticipated needs [Kei 05]. In case of a focus group, the quality of information is limited by creativity [Kei 05]. Furthermore, lead users can identify needs for innovative products [Kei 05], but this method is still infancy [Zog 04] and for some innovative products, the lead users might be in another industry or not representative of the target market [Urb 98]. Lappin [Lap 94] adds that it is hard for consumers to respond to unfamiliar products, since they have no experience with the product. Therefore, in case of innovative products, the needs of the consumers can only partly be generated by means of a Concept test.
Changes: Since Concept testing takes place at the front end of a PCP [Kah 05, Kei 05, Oze 99], the changes in the consumer needs/attitudes environment during the product development can not be analyzed by means of concept testing.
Use: The aim of Concept testing is to prove the new product concept [Kah 05]. This means that the products functionality is defined and how the consumer will use that functionality is not directly the goal of concept testing. However focus groups might result in information about consumer opinions, purchase processes and usage situations [Oze 99]. In that sense there can be said that consumer use is partly covered in Concept testing.
User acceptance of the product

- *Ease Of Use.* Since the concept only covers the functionality, uncertainty concerning the ease of use of the product is not reduced.
- *Rel.Adv:* The relative advantage of the functionality of the product can partly be assessed by means of Concept testing since buyers are asked to evaluate the product concept and state their intentions to purchase it [Oze 99].
- *Communication:* Uncertainty concerning communication of the product advantages to the consumer is not yet determined, since the goal of Concept testing is to determine what the actual advantages could be.

II.1.2 Product-use testing

Kahn [Kah 05] defines Product-use testing as follows:

Current and/or potential consumers evaluate a product's functional characteristics and performance. The aim of Product-use testing is to prove the product's function.

Ozer has a slightly different definition and uses the expression: 'prototype testing' [Oze 99]. Ozer defines prototype testing as testing the prototype in order to determine (1) whether the product lives up to its promises, (2) comparison with other brands, (3) improvements; and (4) how consumers' preferences change after usage.

Product technology uncertainty
Ability: The ability of the (product) technology to meet the intended functions can be seen as what Ozer describes as determining whether the product lives up to its promises [Oze 99]. Ozer states that these results can be obtained by means of an alpha test. Since the focus still is on buying intention, the reduction-level is low.
Compatibility: Ozer mentions a beta test where people use the prototype in their own usage environment and then evaluate the experiences [Oze 99]. Gamma tests are slightly different in the sense that in that case people use the product indefinitely and report problems [Oze 99]. This can reveal issues concerning the compatibility of the (product) technology with existing technological standards, use environment and/or knowledge of the consumer. Issue is that these tests generally are focused on the buying intention, and therefore the phases after the

purchase in the consumer experience process are not represented exhaustively. As a consequence uncertainty concerning compatibility is not tackled.
Parity: Reducing uncertainty concerning performance and attractiveness of the (product) technology compared to other (future) technology is not directly the goal of product use testing. On the other hand an outcome can concern how the product compares with other brands in the market [Oze 99], but this will be on buying intention/functionality only.

Market uncertainty
Target Group: The target consumer and user group is considered as given in product use testing.
Wants&Needs: By means of Product-use testing, the consumer and user wants and needs can be uncovered by interviewing the user about the experienced interaction with the prototype. Still there is the drawback that the test group may not represent the whole population [Oze 99].
Changes: According to Ozer Product-use testing can assess how consumers' preferences change after usage [Oze 99]. Nevertheless, it is questionnable whether the test group is representative.
Use: Consumer use can be tested by alpha, beta and gamma testing [Oze 99]. Information about the problems and ways to correct them are provided, but the small sample may not represent the whole population [Oze 99].
User acceptance of the product

Ease of Use: A gamma test can reveal the ease of using the product very naturally, but it takes time. The tests that take less time have the drawback that they will become more task driven and less natural.

Rel.Adv.: Relative advantage compared to other products on the market can be evaluated by means of a Product-use test [Oze 99]. However, since there is no direct comparison with another product, the assessment of relative advantage depends on the participants' knowledge of similar products.

Communication: product communication is not covered in Product-use testing, since it is performed to proof the products function and not the perception of the product.

II.1.3 Pre-Market testing

A Pre-Market test is a procedure that uses syndicated data and primary consumer research to estimate the sales potential of new product initiatives in order to estimate the sales potential. A known model for pre-market testing is for example ASSESSOR, depicted in Figure II.2 [Kah 05, Mah 92, Sil 78, Urb 83], and BASES [Kah 05]. This model is used for the forecasting of sales and/or market share for a new brand, marketing strategies can be evaluated and diagnostics for improving the product are generated.

Design	*Procedure*	*Measurement*
O_1	Respondent screening and recruitment (personal interview)	Criteria for target group identification (e.g., product class usage)
O_2	Premeasurement for established brands (self-administered questionnaire)	Composition of "relevant set" of established brands, attribute weights and ratings, and preferences
X_1	Exposure to advertising for established brands *and* new brand	
$[O_3]$	Measurement of reactions to the advertising materials (self-administered questionnaire)	Optional, e.g., likability and believability ratings of advertising materials
X_2	Simulated shopping trip and exposure to display of new and established brands	
O_4	Purchase opportunity (choice recorded by research personnel)	Brand(s) purchased
X_3	Home use/consumption of new brand	
O_5	Post-usage measurement (telephone interview)	New brand usage rate, satisfaction ratings, and repeat purchase propensity; attribute ratings and preferences for "relevant set" of established brands plus the new brand

O = Measurement.
X = Advertising or product exposure.

Figure II.2 ASSESSOR research design and measurement [Sil 78]

Product technology uncertainty
Ability, Compatibility and Parity. All three attributes of product technology uncertainty are not covered in a pre-market test, since the aim of the test is to prove the proposed marketing plan and the final product.

Market Uncertainty
Target Group: The target consumer and user group are already defined and probably this group is used for representing the market. Dependent on the type of market used in the test, uncertainty concerning Target Group can be addressed. The more diverse the market is, the more surprising issues will come up.
Wants&Needs: What the consumer and the user want and need is defined earlier in the PDP.
Changes: The changes in the consumer needs/attitudes environment during the product development are not assessed, the needs are captured (O_2, Figure II.2) at the end of the PCP. Wants and needs are assessed in a questionnaire and the output is the weights and ratings of the attributes [Sil 78]. However this is only an input for the model. Therefore the reduction of Changes uncertainty is medium.
Use: In this case the goal is to assess the market success [Sil 78], [Oze 99], [Kah 05], the sales of the product. Consumer use can be addressed well, since the motivation for using the product is covered in a Pre-Market test.
User acceptance of the product

- *Ease Of use:* is not assessed, only the usage rate is predicted [Sil 78]
- *Rel.Adv.:* The ratings and preferences for established brands and the new brands are one of the outcomes of the pre-marketing test [Sil 78]
- *Communication*: product communication, the likeability and believability ratings of advertising materials can be assessed by means of pre-market testing. However not totally for innovative products. [Sil 78]

II.1.4 Market testing

Kahn [Kah 05] defines Market testing as follows:

Targeted consumers evaluate the market plan for a new product in a market setting in order to prove the proposed marketing plan and the 'final' new product

Ozer adds that Market testing is more convenient in case of high risk products and is a controlled experiment in a part of the target market. Lappin [Lap 94] describes a similar approach referred to as 'sales forecasting / predicting product diffusion' which is targeting lead users in order to analyze how quickly the product is adopted. This diffusion is predictive of how quickly the product will move through the following segments of the market. According to Ozer, a test market gives the answer on how to execute marketing strategies, and not on the question whether people will try a product, since this question should have been answered already [Cra 94]

Product technology uncertainty
Ability, Compatibility, Parity: Since the execution of marketing strategies is assessed in this case [Oze 99], Cra 94]. Product technology uncertainty is not reduced by means of market testing.

Market uncertainty
Target Group, Wants&Needs, Changes, Use, Ease Of Use, Communication: As mentioned before the only question answered by market testing is on how to

execute marketing strategies [Cra 94, Oze 99]. Since the product is provided to real users, the *Target Group, Changes, Use and Communication* can be evaluated well. The uncertainty related the *relative advantage* is partly addressed since the goal is not to compare the product itself to competitors' products. Furthermore, Ease of Use is not assessed, since a Test Market focuses on execution of marketing strategies.

The assessment of the four Consumer/market research methods, Concept testing, Product-use testing, Pre-Market testing and Market testing is summarized in the table below.

Table II.2 CMR and uncertainty-reduction (**H**igh, **M**edium, **L**ow)

	Uncertainty-category									
	Ability	Compatibility	Parity	Target Group	Wants&Needs	Changes	Use	Ease Of Use	Rel. Adv.	Communication
Concept testing				L	M		L		L	
Product-use testing	L					M	M	L	L	
Pre-Market testing				M		M	H		M	M
Market testing				H		H	H		M	H

II.2 Usability testing

A distinction is made between the two most common forms of Usability testing, namely, Analytical and Empirical Usability Evaluation [Kar 92, Nie 94, Wan 03].

II.2.1 Analytical Usability Evaluation

Product technology uncertainty

Ability: Nielsen mentions a technique called feature inspections that focuses on the function delivered in a system; whether the function meets the needs of the intended user for example. Feature inspection can involve not only evaluation of that function, but can also involve the design of that function [Nie 93].

Compatibility: Cognitive walkthroughs [Lew 90, Nie 94, Wha 94], consistency- and standards inspection [Nie 94, Wix 94], can eliminate uncertainty concerning Compatibility of the (product) technology. Since the focus is on the ease of use of the system and furthermore, generally no real users are involved, the reduction is low.

Parity: The adoption of a technology can not be predicted by a usability inspector [Nie 94]. Obviously the inspector has the knowledge of competitors' technologies, but this can only partly be useful.

Market uncertainty

Target Group: What the target consumer and user group are is specified before the usability inspection takes place.

Wants&Needs: What the consumer and the user want and need is specified before the tests, since the functionality is already defined in case of a usability test.

Changes: The changes in the consumer needs/attitudes environment during the product development are no serious issue for a usability inspector.

Use: In case of usability inspections where inspectors assess the product, the inspector assumes to be a specific user and assumes how the consumer would use the product. [Hix 93, Nie 94]. Guideline and heuristic reviews work with predefined guidelines and heuristics.

User acceptance of the product

- *Ease Of Use:* The main focus of usability testing is to assess the ease of use. Drawback in case of analytical evaluations is that the real user assumed by the inspector and in that case the assessment in not representative. In case of pluralistic walkthroughs, the real user can be involved, but the drawback remains that that user can not be representative for the whole target market. On the other hand Virzi [Vir 92] and Nielsen [Nie 93] argues that 5 usability inspectors can reveal 80% of the problems.
- *Rel. Adv.:* Usability testing does not specifically aim at comparisons with competitive products, although competitors' products can be evaluated and compared as well. Still, only usability is assessed and not the functionality.
- *Communication:* Product communication is not the scope of usability testing.

II.2.2 Empirical Usability Evaluation

Product technology uncertainty

Ability: According to Holzinger, empirical usability testing provides direct information about how people use systems and their exact problems with a specific interface. Therefore the Ability can be assessed well.
Compatibility: Compatibility of the (product) technology with existing technological standards, use environment and/or knowledge of the consumer is assessed when watching a user using a product (end user testing, interview and observing users in context). However it is questionable whether the user can really point out how innovative technology is compatible to his current situation.
Parity: Performance and attractiveness of the (product) technology compared to other (future) technology. This depends on the knowledge about other technology of the user that is observed or questioned, and obviously one user is not representative for the target market. [Nie 93]

Market uncertainty

Target Group: What the target consumer and user group are is specified before the usability inspection takes place.
Wants&Needs: What the consumer and the user want and need is specified before the tests, since the functionality is already defined in case of a usability test.
Changes: The changes in the consumer needs/attitudes environment during the product development can be slightly assessed by an end user test, although the test is focused on usability, not on utility.
Use: Consumer use is observed in end-user testing.
User acceptance of the product

- *Ease Of Use:* The main focus of usability testing is to assess the ease of use. Especially with the involvement of real end users the ease of use is assessed realistically.
- *Rel. Adv.:* Usability testing does not specifically aim at comparisons with competitive products, although competitors' products can be evaluated and compared as well. Still, only the usability is assessed and not the functionality.
- Communication: Product communication is not the scope of usability testing.

The assessment of the two Usability testing methods; Analytical and Empirical Usability Evaluation is summarized in Table II.2.

Table II.1 Usability testing and uncertainty-reduction (**H**igh, **M**edium, **L**ow)

	Uncertainty-category									
	Ability	Compatibility	Parity	Target Group	Wants&Needs	Changes	Use	Ease Of Use	Rel. Adv.	Communication
Analytical UE	**H**	**L**	**M**				**L**	**M**	**L**	
Empirical UE	**H**	**M**	**M**			**L**	**M**	**H**	**M**	

III. Standard model of the PDP

For describing the PDP in Chapter 2, a standard model of the PDP, presented by Keijzers [Kei 05] is used. Keijzers [Kei 05] has made this classification by combining the PDP of a company (which is the same as the investigated company in this research) with literature on product development. He describes the product design decisions as follows:

Table III.1 Product design decisions on PDP level [adapted from Kei 05]

PDP-level	**Product design decisions**
Product plan (PP)	Competitive dynamics, Segmentation of consumers, Outline of Value Propositions, User interface / usage concepts, Technologies, Key characteristics, Functionality
Product concept (PC)	Feasibility of the technology used in the product, Feasibility of the product concept (attractiveness of the product, concept, degree to which customer needs have been met, competitive environment, user needs/modes/wants.
Product design (PD)	Product platform definition, Product architecture definition, Interface specifications, Standard designs
Released product (RP)	All requirements changes, Verified and validated integrated product design ready for manufacturing
Commercially Released product (CRP)	Directions For Use, Launch plan, Marketing materials, Sales training

IV. Intake questionnaire Product X test

QUESTIONNAIRE BEFORE THE TEST

1 Please rate the following statements:
For each of the five statements, please tick the circle which is most appropriate

	strongly disagree 1	disagree 2	slightly disagree 3	neither agree nor disagree 4	slightly agree 5	agree 6	strongly agree 7
I know pretty much about product X.	○	○	○	○	○	○	○
I do not feel knowledgeable product X.	○	○	○	○	○	○	○
Among my circle of friends, I'm one of the "experts" on product X.	○	○	○	○	○	○	○
Compared to most other people, I know less about product X.	○	○	○	○	○	○	○
When it comes to product X, I really don't know a lot.	○	○	○	○	○	○	○

2 How much effort did you spent on searching information about product X? This information search may, among other things include: visiting consumer web sites, talking to a sales person, acquiring information through friends or colleagues, reading about it in magazines or on the internet.
Circle the appropriate number

None at all 1
Not much 2
Moderate 3
Fairly a lot 4
Very much 5

3 How would you characterize your amount of usage of product X?
This amount of usage may include the use of product X in both your professional, educational and personal life.
Circle the appropriate number

None at all	1
Not much	2
Moderate	3
Fairly a lot	4
Very high	5

4 Do you own a product X?

○ Yes, go to 5

○ No, go directly to 6

5 Of which model and brand is your product X?

6 How much would you pay for such a device?

7 What would you use such a device for?

8 What is your age?

☐ years old

9 What is your gender?

○ Male

○ Female

V. Tasklist OOB test

Because of confidentiality reasons, the OOB task list is not included in the appendices.

VI. Exit questionnaire OOB-test

Because of confidentiality reasons, the OOB exit-questionnaire is not included in the appendices.

VII. Exit questionnaire FT test

Because of confidentiality reasons, the FT exit questionnaire list is not included in the appendices.

VIII. Comparison results Empirical and Analytical Usability Evaluations

Because of confidentiality reasons, the comparison of the Empirical and Analytical Usability Evaluation is not included in the appendices.

IX. Uncertainty assessment in the five test-techniques

Table IX.1 Product technology and market uncertainty [adapted from Kei 05]

Product technology uncertainty	**Market uncertainty**
Uncertainty about: 1. Ability of the (product) technology to meet the intended functions or a translation of those functions into lower level requirements. → **Ability** 2. Compatibility of the (product) technology with existing technological standards, use environment and/or knowledge of the consumer → **Compatibility** 3. Performance and attractiveness of the (product) technology compared to other (future) technology. → **Parity**	Uncertainty about: 1. What the target consumer and user group are. → **Target group** 2. What the consumer and the user want and need. → **Wants&Needs** 3. The changes in the consumer needs/attitudes environment during the product development. → **Changes** 4. Consumer use → **Use** 5. User acceptance of the product a. ease of use → **Ease Of Use** b. relative advantage → **Rel. Adv.** c. product communication → **Communication**

In this section each of the techniques used for the usability evaluation of the PNS will be compared to the ten uncertainty categories presented in Chapter 2, Table IX.1.

The total number of 139 issues that have been addressed in both the analytical (75 issues) and the empirical (111 issues) test have been analyzed. All issues are counted once, this means that a number of 75 issues found in the analytical evaluation and 111 in the empirical will not lead to (75+111=) 186 issues, however, since there are 47 overlapping issues, it leads to a total amount of 139 issues. Similar application can be found in the following addition sums.

$$ST = \sum_{i=1}^{3} ST_i$$

$$OOBt = \sum_{j=1}^{10} OOBt_j$$

$$OOBq = \sum_{j=1}^{10} OOBg_j$$

$$DH = \sum_{k=1}^{6} DH_k$$

ST	Scenario test
OOBt	OOBt-outcome
OOBq	OOBq-outcome
FT_k	FT-outcome
i	scenario-number (i = 1, 2, 3)
j	OOB participant (j = 1,2, … , 10)
k	FT participant (k = 1,2, … , 6)

Since there was no indication about the impact and seriousness of the issues, each issue is handled equally. The assessment on each of the ten different categories of uncertainty by all of the tests together can be shown as follows:

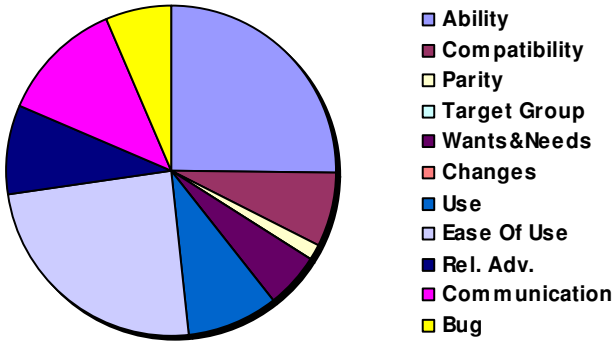


Figure IX.1 Uncertainty-category assessment all tests-techniques

As can be seen in Figure IX.1, the most addressed categories of uncertainty by all test-techniques together are Ability and Ease Of Use, the other user acceptance issues Rel. Adv. and Communication are well presented as well. Uncertainty concerning the Target Group is not addressed at all. The next sections discuss the different techniques that were used in the test separately.

Heuristic Evaluation

HE is used to quickly identify a large number of problems, especially in the User Interface (UI) design [Dou 97, Now 03]. This means that the system menu was examined thoroughly and input issues were addressed, as well as the total layering of the system, which turned out to be illogical. A better layering for the menu was proposed by the evaluators. As Doubleday et al. point out: HE is aimed at finding causes of issues by going deep into the system by means of a systematical assessment. Therefore this technique finds individual usability problems [Nie 93] and it can address expert user issues, although this depends on the evaluator [Nie 93]. Another positive point is the fact that this HE can be performed with a relatively small amount of resources [Gan 06, Nie 94]. In addition, HE can be in all phases of the PCP [Hel 06, Hol 05].

Disadvantage of HE is that it is very abstract since it is not performed in a real environment. Therefore a lot of hardware issues were not covered in the HE.

De Jong and Lentz [Jon 06] argue that HE is not able to reveal the true user-friendliness of the object that is studied. They point out that evaluations that that focus on an in-depth analysis of user problems in actual-use situations may be expected to shed more light on the usability [Jon 06]. Doubleday et al. find the following disadvantages of HE: (1) HE problems can be subjective and dependent on experience of the evaluator [Nie 93, Hol 05]; (2) HE problems are often not distinct; (3) observation vs. immersion; (4) imprecise terminology and (5) actual guidance of the heuristics checklist [Dou 97, Hol 05].

In summary, HE provides causal categories so it can help analyze observed usability problems. The difficulty in this technique is the fact that heuristic evaluators cannot place themselves in all users' shoes hence so they will miss errors [Dou 97] Furthermore, it does not involve real users, so does not find 'surprises' related to their needs [Hol 05, Nie 93].

Uncertainty: HE covers a lot of user acceptance issues: mostly Ease Of Use and Communication. This is not surprising, since HE tests aim at uncertainty concerning user-product interaction and the ease of use by thoroughly and systematically going through the UI. Since the analysis is not supposed to be subjective, facts concerning the other user acceptance class; Rel. Adv., are only slightly addressed.

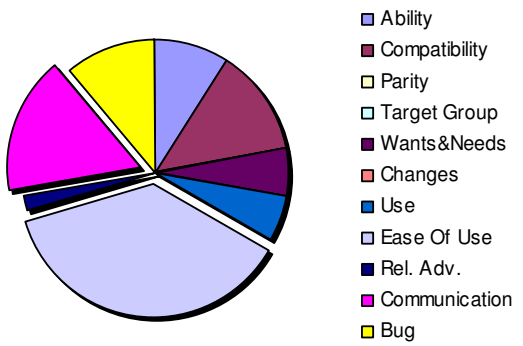


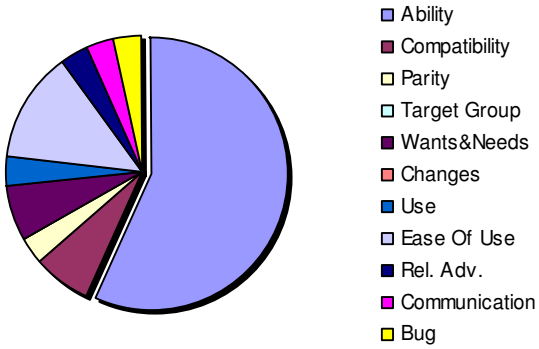
Figure IX.2 Uncertainty-types in HE

Scenario Testing

To perform the ST, a considerable amount of resources was needed to perform the scenarios. When looking at the issues found uniquely by ST, interesting conclusions can be made. In this case, the scenario testers were able to put themselves in uncommon situations, and can therefore (1) try the UI in a real use situation; and (2) just play around and see the reaction of the system In comparison to the FT test, the scenario evaluators can test more focused, since they know in advance what they want to assess. Important in choosing which scenario to test, the focus of the scenarios needs to be realistic and verified.

According to De Jong and Lentz [Jon 06], scenario testing is a remedy for three weaknesses of expert evaluation: (1) a lack of empathy; (2) the danger of experts assuming a superficial review mode and (3) the danger of mixing up different quality criteria. The use of scenarios may be expected to considerably improve on expert reviewers' ability to represent the user and point out the user problems [Gru 02, Jon 06, Sad 05]. These authors take up the use of scenarios not only as performing a task-scenario but they suggest using personal-scenarios, like Cooper's personas [Coo 99] as well. According to Grudin and Pruit [Gru 02], scenarios are less effective when not built on personas.

Uncertainty: As can be seen in Figure IX.3, more than 50% of the issues addressed by the ST, can be labeled as Ability. Since the main goal of the ST test scenarios where the product is tested in common and uncommon situations, al lot of Ability issues were tested by the ST.



Out-of-Box test

Observing users provides insight in the product-user interaction in several use-cases. Nieminen and Vaananen [Nie 03] argue that usability tests in laboratory settings emphasise the instrumental nature of interaction and represent humans as information processors. Therefore it is better to observe (and interview) people in their real settings [Nie 03, Row 94], in this case the setting was the participants' own environment. The users were in their own environment, although there were two observers present. However, some participants pointed out that they would have endeavoured their new system not in the setting of the test for the first time.

The issues observed are highly dependent on the tasks that are performed by the user [Dou 97, Nie 93], so these tasks have to be chosen carefully. The execution of the OOB test was time consuming [Dou 97, Nie 94], since participants had to be recruited, appointments needed to be made and were subject to rescheduling, the outcome needed to be analyzed and translated and the video-material had to be processed. UI issues in the OOB test incline to be more subjective than those found by means of analytical evaluation. Drawback is the fact that the observer always has to translate the observations in words, which might in some case give a wrong impression about the real user experience [Kah 05]. Hornbaek gives some examples of objective measures for measuring effectiveness, efficiency, and satisfaction [Hor 06].

Table IX.2 Objective measures for usability [Hor 06]

Usability aspects	**Objective measures**
Outcomes (effectiveness)	*Expert assessment, comprehension*
Interaction process (efficiency)	*Time, usage patterns, learnability*
Users' attitudes and experiences (satisfaction)	*Psychological usability, reflex responses*

Uncertainty: Figure IX.4 shows that the Ease of Use of the product is the most occurring uncertainty-category observed in the OOB test. The reason is, when observing a test; the observer wants to find out how different persons are able to use the product with efficiency, effectiveness and satisfaction.

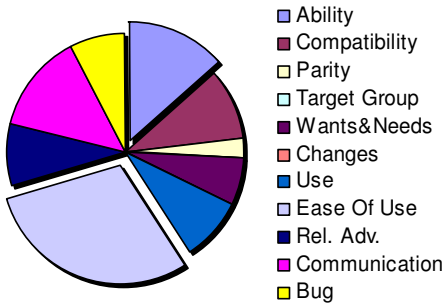


Figure IX.4 Uncertainty-types in OOBt

Furthermore, the observer noticed that participants tend to make comments about the system comparing it to the system that they are used to. Therefore issues concerning the comparison to other technology are addressed by the OOBt. In case of the OOBx, no Parity issues have been addressed. The reason for this might be that the participants might say they like another product (technology) better while they are performing the tests, but they do not feel the need to actually write it down in the questionnaire.

Out-of-Box experience

In this case study, only the additions to the various questions about what the user likes and does not like were taken into account for the comparison in the previous section. However, OOB-experience, measured by means of a five-point Likert scale can reveal useful information as well. The overall usability of the system is normally given by the mean value of each of the attributes that have been measured [Dum 93, Nie 93]. Nielsen adds that this mean should be above a previously specified minimum. A more specific way to measure the overall system usability is to look at the entire distribution of the usability measures. For example, a criterion on the mean value should be at least 4 on a 1-5 scale; that at least 50% of the users should have given the system the top rating; and that no more than 5% of the users gave the system the bottom rating. [Nie 93]. A similar method is given by Hix and Hartson, who define for each usability attribute: a measuring instrument, a value to be measured, a current level, a worst acceptable level, a planned target level and a best possible level. The observed results per attribute are averaged [Hix 93]. Although the value is not statistically correct, this does give a very good impression about the participants' opinion about the system.

Hornbaek gives some examples of subjective measures for measuring effectiveness, efficiency, and satisfaction [Hor 06].

Table IX.3 Subjective measures for usability [Hor 06]

Usability aspects	**Subjective measures**
Outcomes (effectiveness)	*Users' perception of outcome*
Interaction process (efficiency)	*Subjectively experienced duration, mental workload, perception of task difficulty*
Users' attitudes and experiences (satisfaction)	*Validated questionnaires*

According to Holzinger [Hol 05], issues related to the subjective satisfaction of the users and their possible anxieties can best be studied by querying the users, since these kinds of issues are difficult to measure objectively. Questionnaires give information about how end users use the system and their preferred features, but they are an indirect method [Hol 05]. Holzinger adds that one should pay attention to the fact that this technique does not study the actual user interface: it only collects the opinions of the users about the interface. Therefore the user statements should be related to the actual behavior [Hol 05].

Uncertainty: Figure IX.5 shows that uncertainty categories concerning user acceptance; Ease Of Use, Rel. Adv. and Communication are well represented in OOBx.

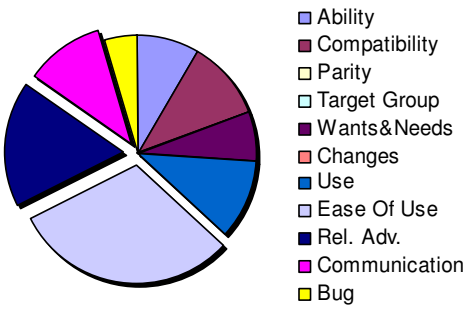


Figure IX.5 Uncertainty-types in OOBx

As can be seen in Table IX.3, Hornbaek [Hor 06] indicates that validated questionnaires are a good means for revealing users' attitudes and experiences. Furthermore, the interaction process was assessed as well in the questionnaire. Since the main focus of the OOB test was to assess the UI, the Ease Of Use of the system was represented well in the questionnaire afterwards.

Field Test experience

Of all techniques mentioned and used, FT gives the user the best opportunity to place the system in his own environment and therefore more practical issues are covered. No extreme situations have been tested and therefore the results might be not very surprising, but on the other hand it gives a proper reflection on the product-user interaction in a real context.

The FT itself was not observed, and the video-material could not be used in some of the cases, therefore the only matter to evaluate the FT was by means of a questionnaire. As was the case in the OOB-experience, only the comments on the questions were taken into account for the comparison. The elaboration on the five point Likert scales, as was mentioned before in case of the OOB-experience, holds for the FT experience as well.

The issues obtained from the FTx were valuable since the user pinpoints what he or she really experienced and therefore what he or she thinks is important in using product X.

Uncertainty: As can be seen in Figure IX.6, the main focus of FT is the assessing Ability-uncertainty. The assessment of this particular category of uncertainty is realized by testing the ability to function of the system in real life-context. The relative amount of Ease Of Use issues is larger that the other technique that has a main focus on Ability; Scenario Testing. The reason for this is the fact that the real user is not mainly focused on the performance of the system, but on the ease of use as well, as opposed to the scenario testers. According to Rosenbaum et al. results from laboratory usability tests tend to be immediately implemental and focused on specific changes to improve ease of use or effectiveness of the product. In comparison, field studies often result in descriptive data that requires greater interpretation and is more subjective [Ros 00].

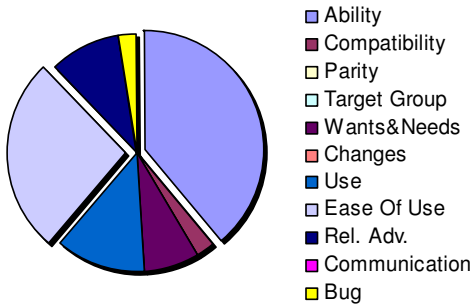


Figure IX.6 Uncertainty-types in FTx

X. Participants Product Y-test

Table X.1 Participants Product Y test

	Gender	Age	Nationality	Industry	Current relation to Product Y	Product Y Set up
1	M	30	Sporean	Insurance	Intender	Never but confident
2	M	32	Msian	Manufacturing (Cable)	Own & Use	Set up with some help
3	M	35	American	Education	Own & Use	Set up without help
4	M	35	Sporean	Building Maintenance & Management	Own & Use	Set up without help
5	F	33	American	Medical Rehabilitation	Intender	Never but confident
6	M	29	American	Education	Own & Use	Set up with some help
7	M	34	American	Education	Own & Use	Never but confident
8	M	46	American	Finance	Intender	Never but interested
9	M	34	Sporean	Medical	Intender	Never but confident
10	F	44	Mexican	Language	Own & Use	Never but confident

Due to confidentially and relevancy issues, the following columns of the table are left out of Table X.1:

- Date
- Panel ID
- Name
- Race
- Designation
- Marital Status
- Number of Kids
- Product Y-brand

-

XI. Task list Product Y-test

Because of confidentiality reasons, the Product Y task list is not included in the appendices.

XII. Function Y_1

This Appendix covers the execution of Function Y_2.
Function Y_1 consists of three tasks: TYa, TYb and TYc

Session summary and critical episodes are left out of the table, because of confidentiality issues. The exection of Function Y_1 consists of six modes (A,B,C,D,E,F) and five actions (1,2,3,4,5). In the table, the modes/actions that either resulted in a referral to the Manual or Quick Start Guide (QSG) or in a hint are elaborated.

Table XII.1 Function Y_1 Execution

test person	**duration**			**manual / QSG**	**hint**
	TYa	TYb	TYc		
1	1:12	4:19	9:07		5
2	10:48	3:50	0:14		5
3	11:16	2:24	5:02	2 3	C E 5
4	0:43	0:14	0:28	4	2 5
5	5:02	2:24	2:09		5
6	10:33	1:12	2:52		2
7	4:48	0:57	8:38		2
8	10:33	10:04	2:09	3	2 3 5
9	10:04	7:55	2:24		5
10	0:57	1:12	3:36	3 5	2 5
sum	18:00	10:33	12:43		
av. duration	1,48	1,03	1,16		

XIII. Function Y_2

This Appendix covers Function Y_2 in six sub-tasks. This action was observed in the Product Y-test (Case Study 2) for all 10 test persons. The most efficient way is given in the EFFICIENT column and the percentage (%) of persons who performed the action confirm the EFFICIENT way, right and right without help is given. Furthermore, the average of the rating of the task by the participants is provided.

Legend:

M	way 1 to assess Function Y_2
S	way 2 to assess Function Y_2
T	way 3 to assess Function Y_2
V	Manual/hint needed to execute
na	not available

Table XIII.1 Execution Function Y_2

test person	1	2	3	4	5	6	7	8	9	10	EFFICIENT	% right average	% right without help
Male/Female	M	M	M	M	F	M	M	M	M	F			
Age	30	32	35	35	33	29	34	46	34	44			
1	T	T	M	T	T	T	T	T	T	M	**T**	80%	
manual/hint	V				V		V	V					
rating	**3**	**4**	**5**	**3**	**5**	**2**	**3**	**4**	**2**	**3**		3,4	
2	T	S - T	T	T	T	T	T	M	M	M	**T**	60%	50%
manual/hint					V								
rating	**4**	**5**	**5**	**3**	**3**	**2**	**4**	**3**	**4**	**4**		3,7	
3	S	S	S	S	S	S	S	M - S	O - S	M - S	**S**	70%	70%
manual/hint													
rating	na												
4	S	S	S	T - S	S	T	S	S	S	S	**T**	10%	10%
manual/hint													
rating	**5**	**3**	**5**	**3**	**1**	**1**	**3**	**1**	**2**	**4**		2,8	
5	T	T	T	T	T	T	T	T	T - M	T	**T**	90%	90%
manual/hint													
rating	na												
6	T	T	T	T	T	T	T	T	T	T	**T**	100%	80%
manual/hint		V					V						
rating	na												

XIV. Delfts Innovatiestappen Model

Den Ouden [Oud 06] identified the iterative Delfste Innovatiestappen Model (DIM) of Buijs and Valkenburg [Bui 05] as the most complete model for a product innovation cycle. This Appendix covers all theories that from the basis of het Delfste Innovatiestappen Model (DIM). After explaining all theories, the DIM is visualized.
Furthermore, since the focus of this research is testing during the PDP, the Product Development Phase of the DIM is highlighted and further explained in Section XII.2.

XIV.1 Basis of the DIM

Integral thinking
In order to emphasize the cooperation between the market, product and manufacturing [And 85, Roo 95], the DIM incorporates a structure based on these three elements. In the model this is adopted as follows: In vertical mode, the internal aspects are represented on the left-hand side of the model (like manufacturing), the innovation project element in the middle (all the product elements) and the external elements on the right-hand side (market) [Bui 03].

Five phases
Kolb's Experimental Learning Model [Car 76, Kol 76] forms the basis for the five main phases of the DIM [Bui 03]. This lead to a number of five main phases: strategy formulation, design brief formulation, product development, market introduction and product in use.

Circular
The five phases as mentioned above succeed each other in a circular model. Inspired by several circular models for creative problem solving [Cou 95, Isa 93], new concept development [Koe 01] and product innovation, Buijs and Valkenburg visualized the process as a circular model. This suggests that there is neither beginning nor end since each product in use leads to a new strategic position of the company [Bui 03]. Since the overall objective of a product innovation process is to have a commercially successful and easy-to-use product in the market place, Buijs and Valkenburg [Bui 03] placed the 'product-in-use' phase at the top (Figure XIV.1)

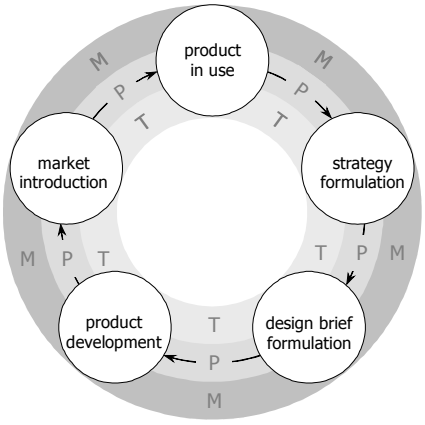


Figure XIV.1 Structure of "het Delftse Innovatiestappen Model (DIM) [Bui 05]

Divergence and convergence
Roozenburg and Eekels [Roo 95] provided with the idea that each phase of the innovation process is consists of a divergent activity followed by a convergent activity. Therefore each phase in the DIM is firstly oriented as getting as many alternatives as possible (divergence) and afterwards a convergent idea screening step follows [Bui 03].

The Basic Design Cycle
Roozenburg [Roo 77] analyzed different models of the product design process and provided with a design method with an analyse-synthese-simulate-evaluate sequence. In the DIM model this basic design cycle is represented in every design stage and every design step.

A visual representation of the model is provided in the figure below.

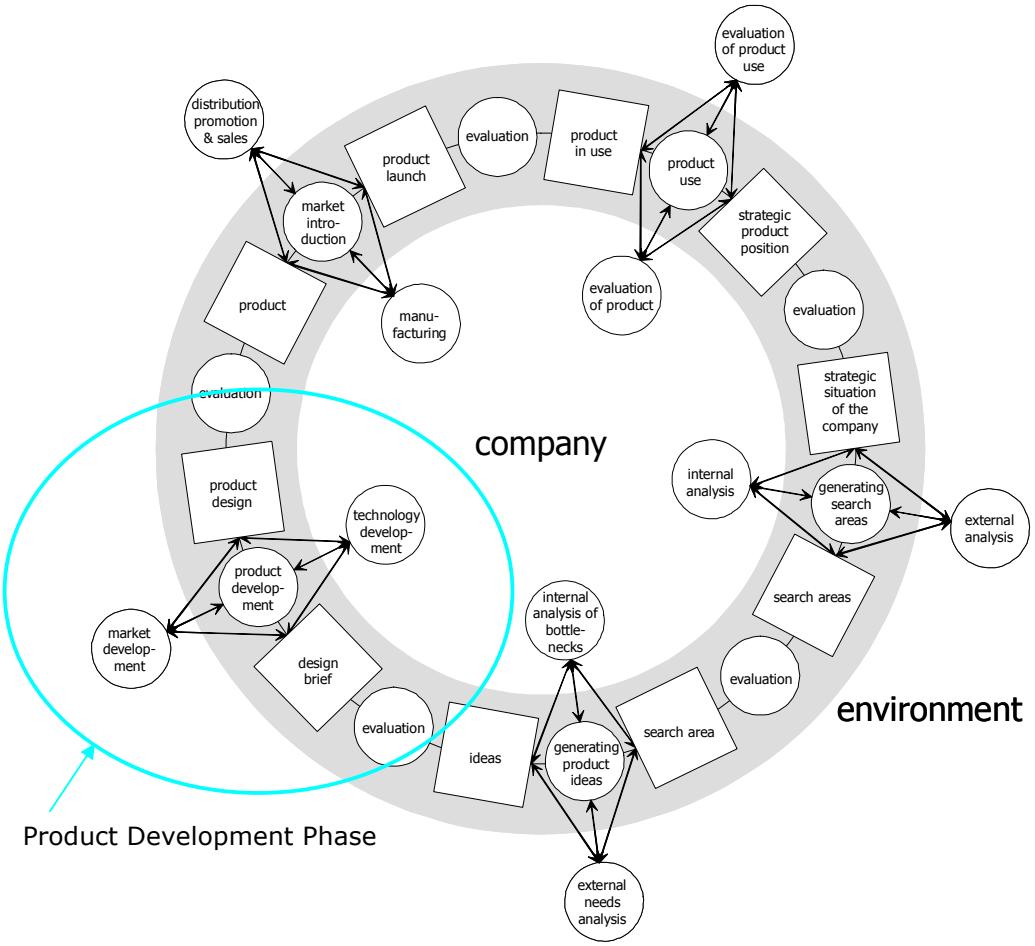


Figure XIV.2 Het Delfts Innovatiestappen Model [Bui 05]

XIV. 2 Product Development Phase

As was mentioned in the introduction of this Appendix, this research focuses on the Product Development Phase and therefore this part of the DIM (the circle in Figure XIV.2) is highlighted. The Product Development Phase consists of four sub-phases: analysis, synthesis, materialization and optimization. In each sub-phase, several steps have to be taken and the three knowledge areas of the DIM are present in the sub phases as well. Following Andreasen and Hein [And 85], the areas production (left), product development (middle) and marketing (right) can be distinguished. In Table XIV.1, considerations during the process and tasks and responsibilities of the diverse disciplines are elaborated [Bui 05]. In the table, the activities in the DIM that relate to the topic of this thesis are underlined and numbered (**1**,**2**,**3**,**4**). These activities are further elaborated in the rest of this Appendix.

Table XIV.1 Realization of the Product Development Phase of the DIM [translated from Bui 05].

	Production	**Product development**	**Marketing**	**Other disciplines**
	The instruction to the development team: the design brief			
Analysis phase	**Technical analysis** • What are the production possibilities? • Which technology are we going to use • Does a new technology need to be developed or bought? • Patent research. • Are there existent parts that need to be used in the product?	**Function Analysis** • Improving an existent or develop a whole new product? • What are the principal and additional functions? • What is the line and sight of the current products? • Analysis of competitors' products.	**Market Research** • What is the size of the market? • Are there different market segments? • What is the potential? • Who are the lead users • What are the demands, expectations and needs of the consumers?	**Finance:** • What is the hedge of the project? **Legal:** • Patent research • Legislation **Supply:** • Overview current suppliers and alliances.
	Formulation of the design specifications by means of a problem definition and a demand program in which all these aspects are represented (for example by means of checklists or a product life-cycle analysis).			
Synthesis phase	**Technical opportunities** • Examine the concepts on technical opportunity. • Find suppliers for important parts.	**Concept Development** • Generate ideas, initial concepts and concepts. • Build and test experimental models. • Feasibility study on concepts.	**Consumer Test** • Choose market segment. • What is the possible application of the product? • Possible observation-test on product-use **1** • Define product options or product line.	**Finance:** • Analysis different initial concepts **Legal:** • Bringing in important service-aspects. **Supply:** • Standard parts.
	Allocate feasible directions in concept and a make choice from these **2**			
Materialization Phase	**Production Process Development** • Allocate final materials and technologies. • Allocate production of all parts • Make technical	**Detail Design** • Elaborate and detail the concepts. • Allocate features.	**Concept testing** • Possibility for doing a concept test with potential buyers or users **3** • Observe shortcomings in	**Finance:** • Make investment plan **Planning:** • Planning allocation people and

	drawings and assembly plan. • Design and buy tools. • Plan the application of people in the production.		product. • Compose marketing plan for this product. • Choose and allocate distribution channels.	machines. **Quality:** • Set up test procedures. **Legal:** • Contracts suppliers.
	Test a working prototype on for example extreme circumstances, feasibility, functions, lifecycle, side effects and critical users **4**			
Optimization Phase	**Production Preparation** • Detail production- and assembly plan • Training production employees. • Produce and evaluate null series.	**Optimize product** • Evaluate test-outcomes. • Adapt the product if necessary.	**Market Preparation** • Execute marketing plan. • Set up market introduction plan. • Approach lead users.	**Sales:** • Set up sales plan. • Develop promotional activities. **HRM:** • Educate mechanics, service employees, sales persons, and etcetera.
	The final product design described in a production plan and a marketing plan			
Market Introduction	**Manufacturing** • Manufacturing • Evaluate production • Acquire part and parts stock control	**Evaluate product** • Evaluate product and feedback information to the succeeding product development.	**Market Introduction** • Evaluation sales. • Evaluation customer satisfaction.	**Sales** **Service** **Logistics**

As mentioned before, the highlighted activities in the table below are further elaborated in the upcoming sections.

XIV.2.1 Observation test on product use

As pointed out by Buijs and Valkenburg [Bui 05], there is a possibility to perform a test with consumers to observe the product use in the synthesis phase. As can be seen in the figure below, the output of this consumer test (together with concept development and technical opportunities) should be a number of concepts, given a problem formulation. In this phase, the market segment is chosen and the possible use is determined. The authors identify a possibility to conduct an observation test that is focused on product-use [Bui 05].

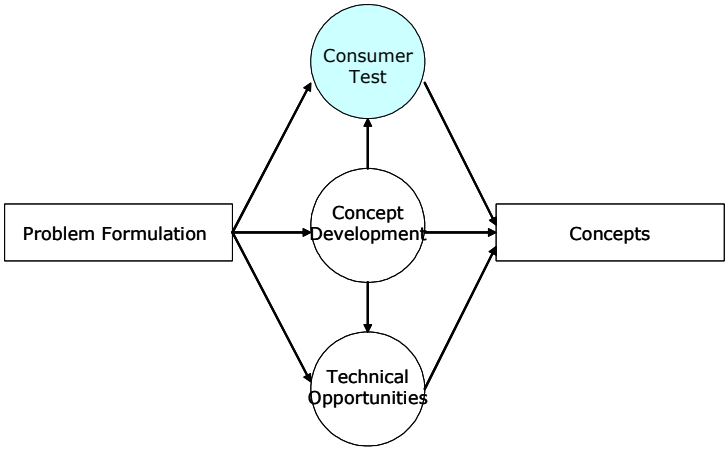


Figure XIV.3 DIM Consumer Test

Input-Output
As can be seen in the figure above, the input for the consumer test is the problem formulation and the concept development. Output is a set of concepts.

Available information
The available information is the problem formulation and a concept under development. Dependent on the kind of problem, a competitors product can be used or a similar product.

Basic Design Cycle

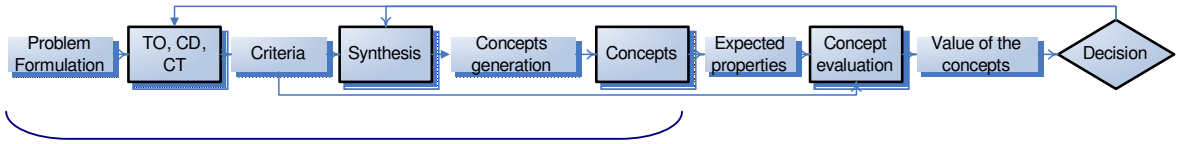


Figure XIV.4 BDC Synthesis Phase (CST)

When the synthesis phase is depicted as a Basic Design Cycle (BDC) [Roo 77], the consumer test (CST) together with the technical opportunities (TO) and the concept development (CD) form the analysis phase. The outcome of this phase are criteria that form the basis of the synthesis, furthermore, these criteria are used to evaluate the concept. The evaluation of the concepts is described in the next subsection.
Jordan et al. point out that including designers and engineers in the requirements capture has positive effects [Jor 96].

XIV.2.2 Concept evaluation

When a number of concepts is generated by allocating feasible directions in the concepts, these concepts are evaluated and a choice is made [Bui 05].

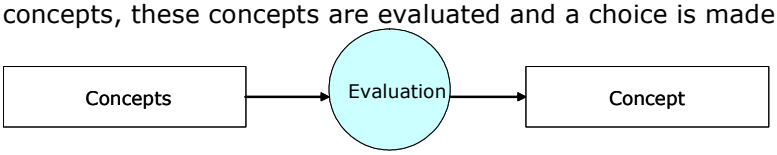


Figure XIV.5 DIM Concept Evaluation

Input-Output
As can be seen in the figure above, the input of the concepts evaluation is a set of concepts and the output is one concept.

Available information
The concepts are the available information together with the criteria formulated in the analysis of the technical opportunities, the concept design and the consumer test. The concepts can have several physical appearances. Dahan and Srivasan developed an Internet-based product concept testing method that incorporates virtual prototypes of new product concepts, substituting them for physical prototypes [Dah 00]. In this method representations of the products can be static or dynamic or with dynamic representations that demonstrate how the product works through a simulated video clip of its operation. Other examples of concepts low fidelity paper-and –pencil mock-up [Bus 03].

Basic Design Cycle

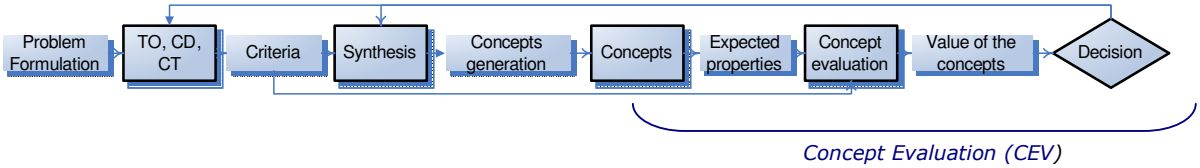


Figure XIV.6 BDC Synthesis Phase (CEV)

As can be seen in the BDC above, the concept evaluation should be able to indicate the value of the concepts which is afterwards used to make a decision. As was mentioned earlier, the criteria that result from the analysis are used to evaluate the concepts as well.

XIV.2.3 Concept test with potential buyers and users

Buijs and Valkenburg point out that there is a possibility for doing a concept test with potential buyers or users during the materialization phase [Bui 05].

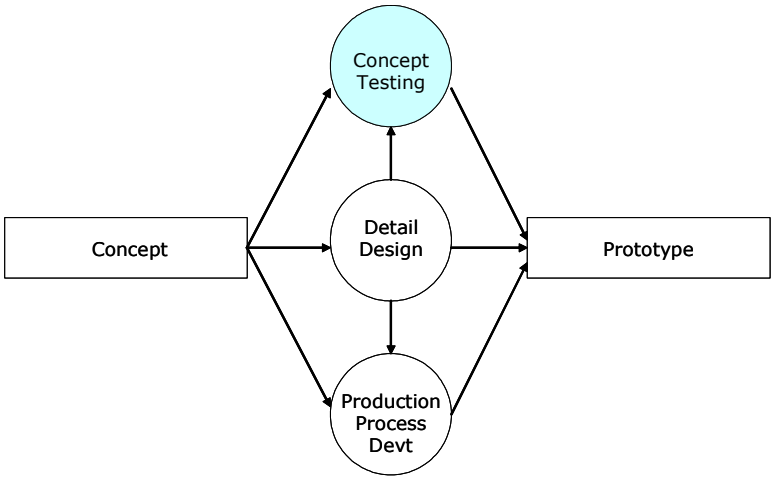


Figure XIV.7 DIM Materialization phase

Input-Output

From the figure above, it becomes clear that the input of the concept testing is the concept and a detailed design, which is a drawn up and detailed concept where the features are established.

Available information

Kaulio [Kau 98] recommends the use of stimulus materials, such as paper-and-pencil sketches, models, mock-ups and prototypes of the product-to-be, are recommended, in addition to verbal communication and he claims that ideally, the presentation of a concept should offer a realistic description of the proposed product(s), in order to facilitate specific responses from customers. To support the conversation between user and designer about how to modify the proposed system to fit the users work better, Beyer and Holzblatt recommend a paper prototype. The prototype must be easy to build, represent the user interface well enough to communicate it to a user, and be easy to modify in the field to support the design conversation [Bey 98]. Another type of prototype is an interactive (software-based) prototype. Dumas and Redish [Dum 93], point out the possibilities to simulate the look and feel for a complex product.

Basic Design Cycle

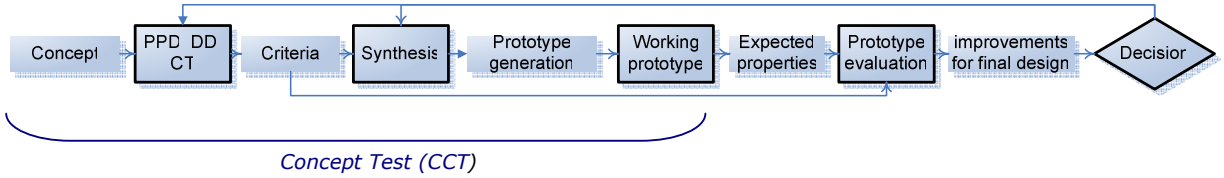


Figure XIV.8 BDC Materialization Phase (CCT)

The BDC shows that the concept test (CCT) together with the production process development (PPD) and the detail design (DD) forms the analysis. The outcome of this phase are criteria that are the basis of the synthesis, furthermore, these criteria are used to evaluate the prototype. The evaluation of the prototype is described in the next subsection.

Conclusion

Kaulio defines concept testing as an approach that aims to involve customers in the conceptual design phase. [Aci 81, Moo 82, Pag 92].

The outcome of the concept test should be able to answer the question whether the product development is on the right track [Bui 05]. In this phase the concept is made tangible and the product team needs to gather information about how the concept is being received by the target market, and on the other hand whether the product/technology is able to meet the determined functionality.

XIV.2.4 Prototype evaluation

When a working prototype has been generated, it needs to be tested on for example usage under extreme circumstances, feasibility, functions, lifecycle, side effects and critical users. This evaluation should evolve in the final design, as can be seen in the figure below.

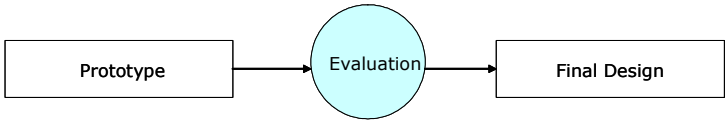


Figure XIV.9 DIM Evaluation prototype

Input-Output

As can be seen in the figure above, the input of the prototype evaluation is a working prototype and the output is the final design.

Available information

According to Buijs and Valkenburg the prototype evaluation assumes a working prototype [Bui 05].

Basic Design Cycle

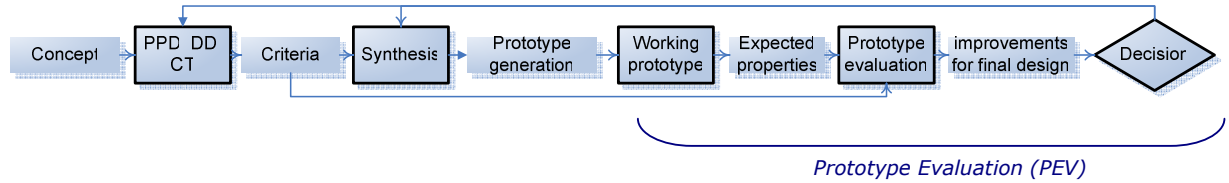


Figure XIV.10 BDC Materialization Phase (PEV)

As can be seen in the Figure XIV.10, the goal of the prototype evaluation is to indicate the value of the prototype and together with some improvements should lead to the final design. As was mentioned earlier, the criteria that result from the analysis (concept test) are used to evaluate the prototype as well.
As mentioned before, in the DIM model this basic design cycle is represented in every design stage and every design step. This suggests that not only the Materialization phase can be depicted as a BDC, however a step in this phase can be depicted as a BCD as well. In aggregation 3 of the Conceptual Model (Section 4.3.3), the BDC for the prototype has been elaborated for a test strategy where a CMR Product-use test and usability testing have been integrated. The structure of the BDC is as follows:

Figure XIV.11 Prototype evaluation based on BDC

XV. Techniques

This Appendix covers the elaboration of all CMR and Usability testing techniques that serve as examples in the conceptual model.

Analytical Usability Evaluation
Analytical Usability Evaluation has been described extensively in the report, Section 2.7 and Appendix II.2.1

Benchmark performance
Olshavsky and Miller perform a benchmark test on the best product available and worst product available to set the parameters for the high and low performance of the features of a product [Ols 72].

Cultural research or ethnography
Investigation of the values, habits and thinking structures that organize the users' everyday activities to understand customer needs. Rosenthal and Capper point out that ethnographic research, carefully planned and implemented, is an effective method for providing user-centered perspectives early in the product innovation cycle [Nie 03, Ros 06].

CMR Concept testing
CMR Concept testing is described extensively in the report, Section 2.6 and Appendix II.1.1
Furthermore, Ozer [Oze 99] defines six sub-groups in CMR concept testing:

Table XV.1 Sub-groups concept testing [derived from Oze 99]

Analogies	Predicting performance by historical sales data of similar products
Expert Opinions	Predicting performance by opinions of different experts
Intentions	Predicting performance by consumer survey data
Multi-attribute models	Predicting a product's relative market position and designing it's features by consumer survey data
Focus Group	Understanding a new product's usage and relevant purchase processes/ Designing new products by opinions of consumers and/or experts
Scenario Analysis	Understanding future market conditions/ Designing new product and predicting their performance by consumer data, historical sales data of similar products, managerial input and data for production constraints

CMR performance evaluation

In order to investigate the relation between performance and expectation of the consumer, Olshavsky and Miller performed a benchmark sound-test. The part where the performance was tested was a sound test on several recorders available in the market [Ols 72]. For the Olshavsky and Miller test, the brand names were visible on the recorders, however if the brand names would be invisible, a test participant is able to evaluate the performance of each system unprejudiced.

CMR Product-use test

CMR Product-use testing has been described extensively in the report, section 2.6 and Appendix II.1.2

Ozer distinguishes between alpha, beta and gamma testing:

- Alpha testing is observing an end-user interacting with a product in a laboratory.
- Ozer mentions a beta test where people use the prototype in their own usage environment and then evaluate the experiences [Oze 99].
- Gamma tests are slightly different in the sense that in that case people use the product indefinitely and report problems [Oze 99].

Co-Discovery Exploration/ Learning

A method to formulate a requirements specification is co-discovery [Kem 96]. Kemp and Van Gelderen describe a co-discovery session as a pair of subjects exploring a product on subjective and emotional aspects. This approach is similar to co-discovery learning [Ken 89], with the difference that the last approach is more focused on usability related to objective product attributes.

Figure XV.1 Co-discovery session [Kem 96]

Consumer idealized design

An example of an approach similar to focus groups is Ciccantelli & Magdison's consumer idealized design: "a process for involving consumers in the actual design of new manufactured goods or services" [Cic 93]. The process focuses on involving users that represent the target market in the early phases of the product design process. The basic idea behind the approach is to get the customers to forget existing products and ignore the feasibility of new designs [Kau 98]. Kaulio points out that the outcome of the session is

- a (new) design
- a list of articulated requirements
- a record of the underlying reasons for the design choices [Kau 98]

Contextual user needs studies

Contextual research is needed to understand the full range of situations in which users must be able to use the product in a satisfying or even enlightening way. Beyer and Holtzblatt describe a methodology to understand customer needs that is called 'Contextual Design' that provides explicit steps and deliverables for the front-end of a development process. [Bey 98, Nie 03].

Data Logging

Logging actual use [Nie 93] is collecting statistics of the detailed use of a system. Normally this is done in the field to collect information about the actual consumer use, but according to Nielsen [Nie 93] it can be used in a testing environment as well. This technique is very common in software development and gives a good indication about the efficiency and effectiveness of the system. Therefore this technique might be less effective in assessing all quality-elements of consumer electronics [Kui 06].

Empirical Usability Evaluation

Empirical usability Evaluation has been described extensively in the report, section 2.7 and Appendix II.2.2.

Feature Checklists

Edgerton focuses on Feature Checklist to obtain real life information because of their ability to identify usability issues like information flood, guessability and reminding in a low cost manner [Edg 96]. Furthermore, Feature Checklists offer information concerning utility like 1) what features are used, understood and reminded, 2) how frequently they are used and 3) the attitude towards the features.

Focus Groups

Nielsen explains focus groups as groups where six to nine persons that represent the potential consumers of the product are brought together to discuss new concepts and identify issues over a period of about two hours [Nie 93], Dumas and Redish point out that most studies that rely on focus groups include 3 tot 10 groups [Dum 93]. Focus groups solicit ideas and feedback through group discussion [Mor 98]. They are moderated by a discussion leader, a skilled moderator, who asks questions and prompts for elaboration, as described in advance by those sponsoring the session [Rob 05]. Focus groups are a popular technique for gathering qualitative data about particular issues, by means of interviewing the users about products or simulations [Mor 98, Via 96].
Several authors address the established focus group method and evaluate its success in online applications [Blo 01, Ste 05, Wan 03].
As pointed out by Stewart and Williams, the online method has several advantages:

- More diverse populations
- Temporal and spatial flexibility for the researcher and researched
- Participants can be questioned over longer periods of time
- Larger numbers can be managed in asynchronous settings
- More considered narratives as result of asynchronous settings

The method is still not used on a big scale due to the following disadvantages:

- Ethical considerations can complicate the research process.
- Identifying and recruiting individuals for online focus groups is problematic.

Hierarchical Task Analysis (HTA)

Hierarchical Task Analysis (HTA) [Ann 67, She 89, Sta 96] is a general approach to analyze tasks. First a hierarchical diagram is made for a particular task and supplementary information is accommodated in a table. According to Stammers [Sta 96], the key features are a hierarchical representation of a task, flexibility in the level of detail of information collected and a tailoring of the analysis to the purpose in hand.

High Contrast Consumer Test (HCCT)

A High Contrast Consumer Test (HCCT) [Boe 03] aims at dealing with unexpected consumer behavior by maximizing variability in the interaction between product and consumer [Kei 05]. HCCT focuses on the Out-of-Box phase of the consumer experience process. Critical and extreme customers are observed using a product in realistic operating conditions. The purpose is to accelerate failures and expose product usage issues as soon as possible [Boe 03]. Applicable in PDP phases: Design, Released Product Design, Commercially Released Products [Kei 05]. is a High Contrast Consumer Test (HCCT) [Boe 03]. According to Keijzers [Kei 05] HCCT test is a promising method, but Van Zoggel [Zog 05] mentions that there still are drawbacks since the test is not very familiar and HCCT is positioned too late in the PDP to be able to anticipate early on NTF.

High/low performance expectations

Already mentioned in case of CMR performance evaluation, Olshavsky and Miller [Ols 72] performed a benchmark sound-test to investigate the relation between performance and expectation of the consumer. Two systems that were already on the market with a total different performance were used and the functionality was described in parameters. One of the systems was the best in class and the other was a reasonable system that the average person could afford. In the test they varied the quality of the functions parameters between these two borders and asked for a judgment by the participants.

Hybrid method

Acito and Hurstad point out that in typical concept testing research, customers are asked to indicate their likelihood of purchase using a summary measure of acceptance [Aci 81]. They describe a hybrid method based on conjoint analysis and conventional concept testing. Conjoint testing measures preferences of individual consumers and therefore yields a more diagnostic approach into regular concept testing. The strategic marketing questions that can be answered concern market share, however the issues related to uncertainty regarding soft failures are:

- In what situations will the new product be most used?
- What is the relative importance of the product's features?

Interviews

Interviewing a consumer is a means of getting information about for example tasks, preferences and the use environment. In the table below a formal definition of interview is provided.

An interview is a conversation between two or more people (The interviewer and the interviewee) where questions are asked by the interviewer to obtain information from the interviewee source: http://en.wikipedia.org/wiki/Interview

Several possibilities for interviewing are known. A well-know techniques is a semi-structured interview [Edg 96]. These semi-structured interviews are conducted with a fairly open framework and which allow for focused, conversational, two-way communication and therefore can be used both to give and receive information [F].

Longitudinal acceptability analysis

The product is analyzed during the whole lifetime to understand customer needs, for example the learning curve and the boredom curve [Nie 03]

Mystery Shopping

Mystery shopping provides insights about the service or product directly from the consumer place [Mon 03]. Mystery shoppers visit selected retail points to gather information and observations about staff responsiveness, attitudes towards customers or products, staff quality and competence, their appearance (and other related behavioral attributes), the aesthetics and functionality of inspected site, i.e. overall perception of the shopping experience [Mon 03]. Mystery shopping originally aims at determining the service level of a particular retail point [Low 02], however if the focus is on the product, interesting information for product development can be obtained.

Personas

Another method that is convenient in the definition of tasks and goals is the relatively new interaction design technique 'Personas'. Cooper describes his invention *persona* as a precise description of out user and what he whishes to accomplish [Coo 99]. The emphasis in the persona-methodology of Cooper is the importance of goals over tasks. In recent literature personas are described as concrete representations of fictional users [Gru 02], or for example Kujala and Kauppinen [Kuj 04] note that the persona is a precise description of a hypothetical user and his or her goals, and it represents a group of users throughout the entire design process.

Questionnaire

A questionnaire can be used to collect subjective data. Nielsen points out that questionnaire are generally given to a test participant after performing tests to measure the subjective satisfaction [Nie 93]. Nielsen adds that a questionnaire can also be used for a system that is already in the market, in this case there is no direct interaction between the test participant and the product (design) [Kei 05].

Realistic user scenarios

De Jong and Lentz use a middle course between Analytical Usability Evaluation and Empirical Usability Evaluation to evaluate a municipal web site, which is presenting experts with realistic usage scenarios, combined with limited sets of user characteristics and evaluation criteria [Jon 06], so the real user is presented and a the cost of the evaluation is narrowed.

Repertory Grid Theory

Baber discusses an evaluation method based on repertory grid theory [Bab 96] to predict the actions of people based on hypothesis and expectations. Baber [Bab 96] provides with two situations that serve as examples of the possible application of this theory in product evaluation. First of all he mentions the factors influencing the decisions of a consumer in selecting a CE product and secondly he uses the method for comparing different solution-concepts.

Rough usability field test

Houde and Hill [Hou 97] describe an experiment with a look&feel prototype of a laptop which only represents the form and weight. This prototype is carried along with a user and the form and weight is evaluated in everyday life.

Survey
Dahan and Hauser define a survey as a written or electronic questionnaire on consumer needs and preferences, the usage situation, etc. [Dah 01]. A survey is a technique that uses one-way interaction; it is performed while the test facilitator and the test participant do not directly interact with each other [Kei 05]. Keijzers adds that there is no direct interaction between the test participant and the product (design) and he exposes the possibility of questionnaires through mail or Internet [Kei 05].

TAFEI
The TAFEI method by Stanton and Baber [Sta 96a] combines Hierarchical Task Analysis (HTA) [Ann 67, She 89, Sta 96] with State Space Diagrams (SSD) [Ang 68]) to provide a useful picture of interactions between human operators and machine components within a system in respect of possible actions and errors [Gle 94]. Stanton and Baber point out that TAFEI can evaluate a product before it is brought into physical existence and it can be used in conceptual stages of product design, reiterative design and evaluation of existing products [Sta 96a].

Understand and Specify Context of Use
In order to obtain criteria for the concepts form consumers several strategies are known. One of these is User Centered Design (UCD), where Jokela et al. [Jok 03] describe the UCD activity 'Understand and Specify Context of Use'. Three main objectives are 1) Know the user, 2) Environment of use, and 3) Tasks that he or she uses the product for. As mentioned before, these issues concern both usability and utility. When a competitor's product or similar product is available it is possible to observe the product-user interaction is a laboratory or real context to obtain criteria on usability and utility.

Wizard of Oz
The Wizard of Oz (WOz) technique is an experimental evaluation mechanism [Sal 93]. A user is observed using an system that may look fully functional in the eyes of the user, but missing services are supplemented by a hidden wizard [Hou 97, Sal 93].

Workshops
Next to focus groups, user representatives can be interviewed in workshops as well [Via 96] in order to generate the criteria for product use. Vianen et al. point out that user workshops to discuss generic usability and utility questions by means of interviewing and observing the user whereas focus groups are carried out to discuss particular issues by means of interviewing the users about products or simulations [Via 96].

XVI. Footnotes Sheet_a and Sheet_b

* If the product under development is not a radical innovation, but already available in another industry and/or market and/or firm [Gar 02, Figure 2.3], the available artifact can be used in the CST. This way a broader spectrum of uncertainty might be covered than in case of a radical innovation. Since the available artifact can have different resemblances to the product under development (same industry and/or market and/or firm), the category of uncertainty reduction of the similar product can not be elaborated in this sheet.

** In literature, there were no applicable techniques found for field observation of real users in the concept testing phase. This is not very surprisingly since unstable prototypes like the ones available for a CEV are generally not used in field studies with real users. Bly [Bly 97] mentions that these prototypes might require constant technical support or when particular elements of the design might need rigorous test measurements. Exception in this case is the look&feel aspect of the artifact which can be evaluated by means of a rough usability field test [Hou 97]. The F–O/I–role and F–O/I–implementation are not represented for CEV in the framework.

*** In literature, there were no applicable techniques found for field observation of real users in the concept testing phase. This is not very surprisingly since unstable prototypes like the ones available for a CCT are generally not used in field studies with real users. Bly [Bly 97] mentions that these prototypes might require constant technical support or when particular elements of the design might need rigorous test measurements. Therefore the F–O/I–role, F–O/I – look&feel and F–O/I–implementation are not represented for CCT in the framework.

XVII. Elaboration Table 4.3

This Appendix elaborates further on the meaning of Table 4.3. This is the first aggregation in Section 4.2 and gives an overview of the category and level of uncertainty that can be reduced in each of the four DIM phases.

Table 4.3 (In this appendix depicted as Table XVII.1) was constructed as follows: As can be seen in Sheet_a, there are several realistic combinations in the CST phase. Each combination has certain categories and accompanying levels of uncertainty reduction. The green row in Sheet_a depicts the highest level of uncertainty reduction for each Uncertainty-type over all CST combinations.

Table XVII.1 represents this summary for each of the four phases.

The goal of the CST is to gain criteria to develop concepts. The criteria cover issues concerning Parity, Target Group, Wants&Needs, Use and Ease Of Use. These criteria can be used to evaluate the concepts (CEV), however there are more uncertainty-categories that can be reduced in CEV: Compatibility and Rel. Adv.. The concept that is chosen in the CEV is further designed and role, look&feel and implementation prototypes can be assessed. In this case the performance (Ability, Compliance and Parity) of the artifact can be assessed fully, while the market uncertainty-categories can only be reduced partly. Furthermore, Target Group is taken as a given and only the Changes are assessed and the earlier posed Wants&Needs can be concerned frozen. Issues found in the CCT and criteria formulated in the test can be used for the detailed design of a prototype. This prototype is evaluated and all uncertainties (except for Target Group and Wants&Needs which were already frozen in the CCT) can be reduced by means of the evaluation of this prototype.

Table XVII.1 (Table XVII.2) Uncertainty reduced in the DIM phases (**H**igh, **M**edium, **L**ow)

	Uncertainty-category									
	Ability	Compatibility	Parity	Target Group	Wants&Needs	Changes	Use	Ease Of Use	Rel. Adv.	Communication
CST			H	H	H		H	H		
CEV		H	H	M	M		M	H	H	
CCT	H	H	H			M	M	M	M	M
PEV	H	H	H			H	H	H	H	H

Interpretational issues

Table x might give the reader the idea that half of the uncertainty categories might be fully reduced in the CST already. Unfortunately this is not the case. In each phase of the PDP, the product to be developed gets more concrete, and evolves by going through the several phases. Therefore high-reduced Use uncertainty in the CST phase does not mean that there will be no more uncertainty concerning Use in the remaining phases of the PDP.

Furthermore a reader might think that, since the final step in the PDP (PEV) is able to reduce all uncertainty in a high level manner, testing only at the PEV phase might be sufficient. Unfortunately this is not a beneficial strategy, because it is important to reduce uncertainties as early as possible [Lu 02, Ver 99], since the cost of an alteration is lower early in the process. Moreover, problems found in PEV tests are generally found too late to solve in the product under development [Lu 02, Lui 03].

In summary, to reduce as much uncertainty as possible, the best manner is to go through all steps.

XVIII. Issues Product X Usability test adapted to the Conceptual Model

Because of confidentiality reasons, the issues of Product X are not included in the appendices.

XIX. Techniques in Figure 4.10

In Sheet_b several techniques can be found in one particular combination of the three dimensions. In order to define a proper test strategy, these several techniques can be integrated as was shown in aggregation 3, Section 4.2.3. Furthermore, due to multiple options in the HOW and WHAT dimensions of the CCT phase, there are two options that have the same effect. These two options are

1) <CCT, L, O/I, role> & <CCT, L, O/I, look&feel>
2) <CCT, F, B, role> & <CCT, F, B, look&feel>

In Table XVIII.1 these two options are highlighted light yellow.

Table XVIII.1 Test strategy Product X adapted

phase	L F	B O/I	prototype	name	Ability	Compatibility	Parity	Target Group	Wants&Needs	Changes	Use	Ease Of Use	Rel. Adv.	Communication
CST*	F	O/I	comp	Survey (I) Feature Checklists	H	H	H	H	H	H	H	H	H	H
CEV	L	O/I	role	Hybrid Method CMR Concept Testing, Multi-attribute Models (I) Repertory Grid Theory				M	M		M			
CEV	L	O/I	impl	CMR Performance Evaluation		H	H						M	
CEV	F	O/I	l/fl	Rough Usability Field Test		H					M	H	H	
CCT	L	O/I	role	CMR Concept testing, Intentions (I) CMR Concept testing, Multiattribute models Repertory Grid Theory						M	M			M
CCT	L	O/I	l/f	Wizard of Oz		M						M	M	
CCT	L	O/I	impl	Co-Discovery Exploration/Learning Empirical Usability Evaluation, Lab CMR Performance Evaluation	M	H	H						M	
CCT	F	B	role	Realistic User Scenarios						M	M			M
CCT	F	B	l/f	Analytical Usability Evaluation, Field		M						M	M	
CCT	F	B	impl	Analytical Usability Evaluation, Field Realistic User Scenarios	H	M	M						M	
PEV	F	O/I	int	CMR Product-use Test, Beta Testing (I) Empirical Usability Evaluation, Field CMR Product-use Test, Gamma Testing (I) HCCT	H	H	H			H	H	H	H	H

* If the product under development is not a radical innovation, but already available in another industry and/or market and/or firm [Gar 02, Figure 2.3], the available artifact can be used in the CST. This way a broader spectrum of uncertainty might be covered than in case of a radical innovation. Since the available artifact can have different resemblances to the product under development (same industry and/or market and/or firm), the category of uncertainty reduction of the similar product can not be elaborated in this sheet. → **In this case the competitors' product has many similarities and therefore all categories of uncertainty are expected to be addressed well.**

Legend:

	combinations in the proposed test strategy
	options in the test strategy
comp.	competitors' product
l/f	look&feel
impl.	implementation
int.	integration
(I)	Categorized as O/I, but covers only I (interview)

XX. Combinations in adaptation to Case Study 2

This Appendix visualizes the promising combinations in the framework, belonging to the adaptation of Function Y_1 (Table XIX.1) and Function Y_2 (Table XIX.2) in Case Study 2.

Table XIX.1 Possible combinations to assess Function Y_1

phase	L F	B O/I	prototype	name	Ability	Compatibility	Parity	Target Group	Wants&Needs	Changes	Use	Ease Of Use	Rel. Adv.	Communication
CCT	L	B	impl.	Informal Usability Evaluation, Lab TAFEI	M	M	M						L	
CCT	L	O/I	impl.	Co-Discovery Exploration/Learning Empirical Usability Evaluation, Lab CMR performance evaluation	M	H	H						M	
CCT	F	B	impl.	Informal Usability Evaluation, Field Realistic User Scenarios	H	M	M						M	
PEV	L	B	int.	Informal Usability Evaluation, Lab workshop TAFEI	M	M	M			L	L	L	L	L
PEV	L	O/I	int.	Co-Discovery Exploration/Learning Empirical Usability Evaluation, Lab CMR Product-use testing, alpha testing	M	H	H			M	M	M	M	M
PEV	F	B	int.	Informal Usability Evaluation, Field	H	M	M			M	M	M	M	M
PEV	F	O/I	int.	CMR Product-use test, beta testing (I) Empirical Usability Evaluation, field CMR Product-use test, gamma testing (I)	H	H	H			H	H	H	H	H

Table XIX.2 Combinations to assess Function Y_2

phase	L F	B O/I	prototype	name	Ability	Compatibility	Parity	Target Group	Wants&Needs	Changes	Use	Ease Of Use	Rel. Adv.	Communication
CCT	L	O/I	l/f	Wizard of Oz		M						M	M	
CCT	F	B	l/f	Informal Usability Evaluation, Field		M						M	M	
PEV	L	O/I	int.	Co-Discovery Exploration/Learning Empirical Usability Evaluation, Lab CMR Product-use testing, alpha testing	M	H	H			M	M	M	M	M
PEV	F	B	int.	Informal Usability Evaluation, Field	H	M	M			M	M	M	M	M
PEV	F	O/I	int.	CMR Product-use test, beta testing (I) Empirical Usability Evaluation, field CMR Product-use test, gamma testing (I)	H	H	H			H	H	H	H	H

Legend:

comp.	competitors' product
l/f	look&feel
impl.	implementation
int.	integration
(I)	Categorized as O/I, but covers only I (interview)