Analyzing User Perceived Failure Severity in Consumer Electronics Products

Incorporating the User Perspective into the Development Process
Analyzing User Perceived Failure Severity in Consumer Electronics Products
Incorporating the User Perspective into the Development Process

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op dinsdag 13 mei 2008 om 16.00 uur

door

Ilse Maria de Visser

geboren te Eindhoven
Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. A.C. Brombacher

Copromotoren:
dr. Y. Lu
en
dr.ir. P.H. den Ouden

Copyright © 2008 by I.M. de Visser

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the copyright owner.

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Visser, Ilse M. de

Predicting User Perceived Failure Severity in Consumer Electronics Products – Incorporating the user perspective into the development process / By Ilse M. de Visser. – Eindhoven: Technische Universiteit Eindhoven, 2008. – Proefschrift –


NUR 964

Keywords: Product development / Reliability optimization / User perspective / Consumer complaints / Failure attribution / Consumer electronics

Printed by: University Printing Office, Eindhoven

Cover design by: Sofie den Ouden (VissenCom)

This PhD-study has been carried out as part of the Trader project under the responsibility of the Embedded Systems Institute. This project is partially supported by the Dutch Government under the Bsik program.
Acknowledgements

Doing a PhD is quite a journey. After four years of “travelling” the academic world, I have reached my destination: the completion of this thesis. However, this thesis would have never been completed successfully without the directions and good company of all my travelling companions. Therefore, I would like to take this opportunity to thank a number of them in particular.

First, and foremost, I would like to thank my supervisors. After doing my Masters it was prof. Aarnout Brombacher, my first supervisor, who convinced me to do a PhD. But also during these four years of doing research, he has been my strongest motivator. His enthusiasm, encouragement and good advice helped me to develop myself as a researcher and business professional. I am very grateful for the freedom he gave me in doing my work, but also for the guidance at crucial moments in my research project. I also owe much gratitude to my daily supervisor, dr. Lu Yuan. During the past four years, she has always been there for me. Although she is a “busy bee”, she always found time to help me out. I don’t think that there is another person who has read and commented this thesis more than she did. Her patience, advice, humor and friendship pulled me through the difficult last phase of writing this thesis. Thank you, Lu Yuan! Furthermore, I would like to thank my third supervisor, dr. Elke Den Ouden. Although her involvement in this project was in the later stages, it was very fruitful and pleasant. Especially her knowledge of industry greatly improved the business perspective of this thesis. Her sharp remarks and ideas about product development have helped me a lot.

Secondly, I would also like to thank prof. Aksit, prof. Eggen and prof. Van Gemund for their role as additional supervisors and their help with my thesis. They thoroughly reviewed the draft of this thesis and provided me with many useful comments in the final stage. I really appreciate their contributions.

I would like to convey my appreciation to the members of the Trader project and the industry partners. Without their support, this thesis would have never been published. In particular, I have to mention Frans Beenker (ESI), David Watts (ESI) and Ben Pronk (NXP). With their help, I had the opportunity to participate in an interesting multidisciplinary project and was able to connect to different industry partners.

Several students contributed to the research reported in thesis. I would like to thank all of them. They are Martin Kools, Erik Schrijnemaekers, Bas Gielen and Rick Bergmans.

My colleagues at Industrial Design in the subdepartment Business Process Design at TU/e also gave me much support during my period as a PhD student. I would like to thank them all for helping me in one way or the other. They include Aravindan Balasubramanian, Hanneke Driessen, Wim Geudens, Christelle Harkema, Maurits Houben, Kostas Kevrekidis, Aylin Koca, Joël Luyk, Jan Rouvroye and Peter Sonnemans. Jeroen Keijzers, as my former roommate and fellow Trader project member, deserves a special ‘Thank You’ for listening to me during difficult moments of this research.
I would like to thank my parents for their unlimited support and attention. During all these years, they have helped me out with so many things. My well-being has always been their primary concern. I am very grateful for this. Mom and dad, thanks for everything!

Lastly, I want thank Joël for always being there for me. His support, trust and humor mean the world to me. Joël, I couldn’t have done it without you!

Ilse de Visser
2008
# Table of Contents

Acknowledgements..........................................................................................................................................v

Table of Contents ............................................................................................................................................. vii

Summary ........................................................................................................................................................... xi

Samenvatting ................................................................................................................................................... xiii

List of Abbreviations....................................................................................................................................... xv

1 Introduction ......................................................................................................................................................1

1.1 Consumer electronics in the 21\textsuperscript{st} century; status and trends ...........................................1

1.2 The Trader project .....................................................................................................................................5

1.3 Definition of concepts.............................................................................................................................5

1.4 Thesis structure .......................................................................................................................................7

2 Quality Improvement in Consumer Electronics .........................................................................................9

2.1 Implications of trends on quality improvement.....................................................................................9

2.2 Quality improvement methods in literature ..........................................................................................16

2.3 Exploring industrial best practices with respect to customer dissatisfaction.................................27

2.4 Problem definition and research question............................................................................................35

2.5 Research Approach.............................................................................................................................37

3 User Perceived Failure Severity (UPFS) .....................................................................................................41

3.1 Consumer (dis)satisfaction and complaints .........................................................................................41

3.2 Emotion, personality and irritation.........................................................................................................43

3.3 Overall theoretical model ....................................................................................................................47

3.4 User Perceived Failure Severity (UPFS) ...............................................................................................48
4  An UPFS Model based on Best Practices from Industry .........................................49
4.1  The expert-based UPFS prediction model .............................................................49
4.2  Experimental design .............................................................................................51
4.3  Experimental results .............................................................................................56
4.4  Conclusions ............................................................................................................61
5  An UPFS Model with Theoretical Aspects .............................................................63
5.1  The theory-based UPFS prediction model ..............................................................63
5.2  Influence of Function Importance on UPFS ..........................................................67
5.3  Conclusions ............................................................................................................72
6  Measurement of the Experimental Variables .........................................................73
6.1  Independent variable: FUI measurement ...............................................................74
6.2  Dependent variable: UPFS measurement ...............................................................80
6.3  Extraneous variables: control for users and their environments .........................97
6.4  Conclusions ............................................................................................................100
7  Function Importance Experiment .........................................................................103
7.1  Overview of the FUI experiment ...........................................................................103
7.2  Hypotheses ............................................................................................................104
7.3  Experimental variables .........................................................................................104
7.4  Experimental tasks: failure scenario design ..........................................................107
7.5  Experimental protocol ...........................................................................................109
7.6  Participant selection ..............................................................................................111
7.7  Pilot experiment ....................................................................................................112
7.8  Experimental results .............................................................................................113
7.9  Conclusions ............................................................................................................121
Summary

Currently, the field of consumer electronics is one of the most challenging environments with respect to product design. Due to the combination of the continuous influx of new technology (Moore's law) and the economic effects of globalization, it is now possible to create products with a functionality that was unimaginable even one generation ago against a price level that opens huge markets on a global level in a very short time. However, the combination of technically sophisticated products and global markets is no guarantee for customer satisfaction.

Not too long ago, the field was far more predictable. The development process of the simple and analog hardware-based consumer electronics products of the 1980's/1990's was mainly driven by static product roadmaps that were cost oriented and had a strong focus on product manufacturing optimization. The main goal was to make products that complied with technical specifications. Therefore, in this period, product defects were mainly specification violations and the customer dissatisfaction level could be defined in terms of the required number of product repairs in the after-market. However, the complexity of current software-based consumer electronics products and the increasing customer expectations result in increasing numbers of consumer complaints on new products in consumer electronics industry [OUD06]. Analysis of these complaints indicates that to an increasing degree the cause of the complaint cannot be retrieved [BRO04]. Current product defects do not only represent violations of the product specifications, but also unconsidered customer requirements and unexpected product behavior. Consequently, consumer electronics products nowadays require a design oriented development process in which the focus is on the user of the product. The focus in the product development process has shifted from a specification focused manufacturing approach to a user focused design approach.

This shift in focus of the product development process has, as a logic consequence, resulted in a simultaneous increase in the level of market uncertainty over the nature and extent of customer's need for new products. This higher market uncertainty results in increased information requirements of the current product development processes. Literature review was conducted to investigate the potential contribution of existing quality methods on market uncertainty reduction in the development process of consumer electronics products. It was revealed that current methods are incapable of dealing with market uncertainty due to the combination of lacking user-orientation and confined completeness and specificity. Particularly, the unknown impact of quality problems on user dissatisfaction limits the potential contribution of these methods to the quality improvement decision process.

In order to evaluate the potential contribution of quality feedback information to market uncertainty reduction, three case studies were performed at consumer electronics service centres. The case studies investigated whether the feedback processes have met the increased information requirements of the current product development processes. The case studies revealed some important fundamental constraints of the current feedback processes for uncertainty reduction. Particularly, the strong logistical orientation of the service process hinders information collection for improved design decision making. Also, the available service information does not provide the required insight into customer dissatisfaction levels.
The combination of all the above mentioned limitations makes the service feedback process structurally unsuitable for uncertainty reduction in the design process of consumer electronics.

In order to deal with the increased market uncertainty, existing quality methods from literature and practice will have to be complemented with a user-oriented impact prediction. The first step in developing this impact prediction model is the definition of a measure for user dissatisfaction that expresses the impact of quality problems in consumer electronics: User Perceived Failure Severity (UPFS). UPFS is the level of irritation experienced by the user caused by a product failure. Insight into the expected/predicted UPFS resulting from a certain design decision would reduce uncertainty in the product development decision process and decrease the number of product complaints.

Subsequently, the best practices in industry with respect to customer dissatisfaction prediction were analyzed. The starting point for this analysis is an approach that is currently applied in the consumer electronics industry for user dissatisfaction prediction. A first hypothetical UPFS model was further tested in a consumer experiment. With the results of this experiment, this practical prediction approach is proven to be invalid. Several other recent sources in literature also dispute the ability of experts to predict user preferences and behavior. In short, there is lack of theoretical foundation for the application of this first hypothetical UPFS prediction model in practice.

In the following, a theoretical UPFS prediction model was developed. This adjusted UPFS prediction model is based on the results of the first UPFS consumer experiment, literature from different scientific fields and an expert validation session with people from academia. In two consumer experiments, the validity of parts of this UPFS prediction model was evaluated. The experimental results confirm the validity of these parts of the UPFS prediction model, namely:

- The influence of Function Importance on UPFS
- The influence of Failure Attribution on UPFS
- The influence of Irritableness on UPFS

In addition, based on the results of these three consumer experiments, an UPFS research approach has been developed. By applying this validated approach, future UPFS research can be performed reliably. These UPFS research results give insight into the impact of product quality problems on user dissatisfaction in consumer electronics. Although the experimental results and the proposed UPFS experimental protocol are already useful, additional research is recommended to further validate the UPFS prediction model.
Samenvatting

Het beheersen van productkwaliteit in de consumenten-elektronica industrie is aanzienlijk ingewikkelder geworden als gevolg van een aantal trends in deze sector. Voornamelijk de toegenomen tijdsdruk op het productontwikkelingsproces, de globalisering van zowel de afzetmarkt als de ontwikkelingsprocessen, de technische complexiteit van de producten en een afnemend inzicht van klanten in de technische werking van producten dragen hiertoe bij. Aan de ene kant stellen klanten hogere eisen aan het technische vernuft en de gebruiksvriendelijkheid van producten. Aan de andere kant worden de product-ontwikkelingsprocessen korter en ingewikkelder als gevolg van de verhoogde druk op de doorlooptijd en de globalisering.

Het ontwikkelingsproces van de simpele, voornamelijk uit analoge hardware bestaande consumenten-elektronica producten uit de jaren '80 en '90 was voornamelijk kostprijs gedreven en vereiste daardoor een sterke (prijs)optimalisatie van het productieproces. Het belangrijkste doel van dit proces was om producten voort te brengen die voldoen aan de technische specificaties. Klantenklachten werden voornamelijk veroorzaakt door producten die buiten de productspecificaties vielen, waardoor het aantal productreparaties in het serviceproces een goede indicatie van het algemene ontevredenheidsniveau van de totale klantengroep was. Echter, de complexiteit van de huidige, op software gebaseerde consumenten-elektronica producten en de hogere verwachtingen van de klanten resulteren in een toename van de aantal klantinklachten over nieuwe producten in de consumenten-elektronica [OUD06]. Analyse van deze klachten geeft aan dat in toenemende mate de oorzaak van de klachten niet kan worden achterhaald [BRO04]. Daarom zal in het ontwikkelingsproces van huidige consumenten-elektronica producten de nadruk moeten liggen op de gebruiker. In het ontwikkelingsproces is de focus verschoven van productspecificaties en productie naar productgebruikers en ontwerp. Huidige klantenklachten worden niet alleen veroorzaakt door producten die buiten de productspecificaties vallen, maar ook door onvoorzien klanteneisen.

Deze verandering in focus van het productontwikkelingsproces wordt gekenmerkt door een gelijktijdige toename in de mate van marktonzekerheid en een groeiende informatiebehoefté in het productontwikkelingsproces. Door middel van literatuuronderzoek is de potentiële bijdrage van bestaande kwaliteitsmethoden aan de benodigde onzekerheidsreductie in het ontwikkelingsproces onderzocht. Dit literatuuronderzoek toont aan dat de huidige kwaliteitsbeheersingsmethoden ongeschikt zijn voor het verlagen van deze marktonzekerheid doordat in de methodes een duidelijke klantoriëntatie ontbreekt en ze vaak incompleet en onspecificisch zijn. Voornamelijk de onbekende invloed van kwaliteitsproblemen op de ontevredenheid van de gebruiker beperkt de potentiële bijdrage van deze methodes aan het beslisproces tijdens de productontwikkeling.

Drie casestudies zijn uitgevoerd om de potentiële bijdrage van feedback informatie aan de onzekerheidsreductie in het ontwikkelingsproces te onderzoeken. De casestudies zijn uitgevoerd in drie servicecentra voor consumenten-elektronica producten. Door middel van de casestudies is onderzocht in hoeverre de gebruikte feedbackprocessen voorzien in de toegenomen informatiebehoefté van het product ontwikkelingsproces.

xiii

Om met deze toegenomen marktonzekerheid om te kunnen gaan moeten bestaande kwaliteitsbeheersingsmethoden uitgebreid worden met een model dat de invloed van product kwaliteitsproblemen op de klantentevredenheid voorspelt. De eerste stap in de ontwikkeling van een dergelijk voorspellingsmodel is de definitie van een maat voor gebruikersentevredenheid waarin de invloed van product kwaliteitsproblemen uitgedrukt kan worden: User Perceived Failure Severity (UPFS). UPFS is de mate van irritatie die een gebruiker ervaart als gevolg van het falen van een product. Inzicht in de verwachte/voorspelde UPFS als gevolg van een bepaalde ontwerpbeslissing zou de onzekerheid in het product ontwikkelingsproces kunnen verlagen en zou het aantal klantenklachten kunnen verminderen.

Vervolgens is er een UPFS voorspellingsmodel ontwikkeld gebaseerd op een praktijkmethode voor het voorspellen van klantentevredenheid in consumenten-elektronica. Dit eerste hypothetische UPFS model is getoetst in een consumentenexperiment. Aan de hand van de resultaten van het experiment is aangetoond dat deze huidige praktijkmethode voor het voorspellen van klantentevredenheid ongeschikt is. Een aantal recente literatuurbronnen beschrijven ook het onvoldoende van experts om de voorkeuren en het gedrag van klanten te voorspellen. Al met al is er een onvoldoende theoretische onderbouwing voor de implementatie van het hypothetische UPFS voorspellingsmodel in de praktijk.

Hierna is een theoretisch UPFS voorspellingsmodel ontwikkeld. Dit aangepaste UPFS model is gebaseerd op de resultaten van het eerste UPFS consumenten experiment, literatuur uit verschillende wetenschappelijke onderzoeksgebieden en een validatie-sessie met een academisch panel. Vervolgens is de validiteit van delen van het model getoetst in twee consumentenexperimenten. De experimentele resultaten bevestigen de validiteit van deze delen van het UPFS voorspellingsmodel, namelijk:

- De invloed van Functie Belang op UPFS
- De invloed van Fout Attributie op UPFS
- De invloed van Geïrriteerdheid op UPFS

Daarnaast is aan de hand van deze experimentele resultaten een UPFS onderzoeksaanpak ontwikkeld. Door het toepassen van deze aanpak kan toekomstig UPFS onderzoek op een betrouwbare manier worden uitgevoerd. De UPFS onderzoeksresultaten geven inzicht in de invloed van product kwaliteitsproblemen op de klantentevredenheid voor consumentenelektronica producten. Hoewel de experimentele resultaten en de voorgestelde onderzoeksaanpak op dit moment al bruikbaar zijn, is toekomstig onderzoek nodig voor de verdere validatie van het UPFS voorspellingsmodel.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis Of Variance</td>
</tr>
<tr>
<td>DE</td>
<td>Direct Elicitation</td>
</tr>
<tr>
<td>DOE</td>
<td>Design Of Experiments</td>
</tr>
<tr>
<td>DR</td>
<td>Direct Rating</td>
</tr>
<tr>
<td>FA</td>
<td>Failure Attribution</td>
</tr>
<tr>
<td>FF</td>
<td>Failure Frequency</td>
</tr>
<tr>
<td>FI</td>
<td>Failure Impact</td>
</tr>
<tr>
<td>FMEA</td>
<td>Failure Mode and Effect Analysis</td>
</tr>
<tr>
<td>FMUP</td>
<td>Failure Moment in Use Process</td>
</tr>
<tr>
<td>FR</td>
<td>Failure Reproducibility</td>
</tr>
<tr>
<td>FS</td>
<td>Failure Solvability</td>
</tr>
<tr>
<td>FUI</td>
<td>Function Importance</td>
</tr>
<tr>
<td>FWA</td>
<td>Failure Workaround</td>
</tr>
<tr>
<td>HCCT</td>
<td>High Contrast Consumer Test</td>
</tr>
<tr>
<td>IDB</td>
<td>Information Display Board</td>
</tr>
<tr>
<td>IFP</td>
<td>Introduction Factory Process</td>
</tr>
<tr>
<td>IRP</td>
<td>Introduction Repair Process</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>MSA</td>
<td>Measure of Sampling Adequacy</td>
</tr>
<tr>
<td>NA</td>
<td>Negative Affect</td>
</tr>
<tr>
<td>NFF</td>
<td>No Fault Found</td>
</tr>
<tr>
<td>NPD</td>
<td>New Product Development</td>
</tr>
<tr>
<td>PA</td>
<td>Positive Affect</td>
</tr>
<tr>
<td>PAL</td>
<td>Point Allocation</td>
</tr>
<tr>
<td>PANAS</td>
<td>Positive And Negative Affect Schedule</td>
</tr>
<tr>
<td>PCP</td>
<td>Product Creation Process</td>
</tr>
<tr>
<td>QFD</td>
<td>Quality Function Deployment</td>
</tr>
<tr>
<td>QRE</td>
<td>Quality and Reliability Engineering</td>
</tr>
<tr>
<td>ROE</td>
<td>Rank Order Elicitation</td>
</tr>
<tr>
<td>RPL</td>
<td>Rating on a Pre-established List</td>
</tr>
<tr>
<td>SC</td>
<td>Service Centre</td>
</tr>
<tr>
<td>TQM</td>
<td>Total Quality Management</td>
</tr>
<tr>
<td>TRADER</td>
<td>Television Related Architecture and Design to Enhance Reliability</td>
</tr>
<tr>
<td>TS</td>
<td>Triadic Sorting</td>
</tr>
<tr>
<td>TTM</td>
<td>Time-To-Market</td>
</tr>
<tr>
<td>TV</td>
<td>Television</td>
</tr>
<tr>
<td>UCD</td>
<td>User Centered Design</td>
</tr>
<tr>
<td>UPFS</td>
<td>User Perceived Failure Severity</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Consumer electronics in the 21st century; status and trends

Consumer electronics is a commonly used term to specify the category of electronic equipment intended for everyday use by end-users, the consumers. Although this description is at least 50 years old, the technical capabilities of these systems have changed tremendously over the past two decades [GRI94].

Complexity
A good illustration of this development is the transformation of the mono-functional black and white television (commonly available from the 1960's) into the network connected, multi-functional LCD television (introduced in the 2000's). The increasing technical complexity of consumer electronics products can also be illustrated by the increase in software content of these products. In the late 1980's, a consumer electronics product contained around 10 000 lines of code. Around 1995, a similar product already contained around 100 000 lines of code [ROO96]. In 2000 this number had already increased to an average of 1 000 000 lines of code. This implies a tenfold increase in software content within 5 years. It is expected that this exponential growth will continue for the entire field of consumer electronics products [BER04], [GEN07].

The increase in software content of consumer electronics products has resulted in a corresponding increase in internal complexity or increasing state space of these products. This increasing internal complexity of consumer electronics products results in an explosive growth in the number of possible state transitions within the product. This huge amount of possible interactions between different software and hardware components makes it almost impossible to predict overall product behavior and to determine all possible failures of these products [PET03]. Organizations in the professional electronics industry have discovered how difficult it is to put a product with high software content on the market without any defects [ROO96]. Due to the inherent complexity of consumer electronics products, it is practically impossible to specify the interfaces between a company's self-developed software and its 3rd party software with hundred percent coverage. Outsourcing parts of the software development process may also result in the deterioration of the company's knowledge about the performance of the software. Ultimately, the lack of knowledge about the quality and reliability of new products will result in unexpected field failures and high levels of warranty costs.

Furthermore, the context in which current consumer electronics products operate has changed tremendously over the past years. Currently, these products are embedded in a complete network of related digital products in their use environment. Interconnectivity and compatibility are basic requirements for current consumer electronics products. Nevertheless, users have unaltered expectations regarding the performance of these products.

Product Understanding
Related to, but not necessarily resulting from the increasing technical complexity of consumer electronics products, is the growing lack of product understanding by consumers. Research indicates that the number of product features is an important buying criterion for consumers; the more the product “can do”, the better [THO05].
But at the same time, consumer electronics service centres are triggered by the increasing number of returned products caused by ignorant users [OUD06]. Thomson summarizes this problem as follows: “Because consumers give more weight to capability and less weight to usability during pre-use evaluations than during post-use evaluations, consumers tend to choose overly complex products that do not maximize their satisfaction, which results in feature fatigue” [THO05].

Supplementary to the growing lack of product understanding by consumers is the decrease in tolerance of consumers for undesired product behavior. Customers often do not realize the complexity behind the systems they use, therefore they do not see the difficulties that come with complex systems and just expect them to work [BRO01]. The trend of this increase in customer requirements is also expressed in the warranty period and coverage. Nowadays, a warranty period of two years is quite common. In the past, warranties only covered the products that did not comply with the technical product specifications, the so-called “technical quality problems”. These days, most manufacturers tend to follow a “no questions asked policy”. This means that a product is considered to have failed if customers are simply dissatisfied with its performance.

Oliver defines customer satisfaction as “a judgment that a product feature, or the product itself, provided (or is providing) a pleasurable level of consumption-related fulfillment, including levels of under- or overfulfillment” [OLI96]. This definition implies that currently the development of “non-failing” products implies more than providing a product that complies with technical product specifications; consumer products should create sufficient consumption-related fulfillment. This additional group of failures is named “non-technical quality problems” and is formally defined as “a situation where in spite of meeting explicit product specifications, a customer explicitly complains on the (lack of) functionality of the product” [BRO04].

Globalization
In this 21st century’s world of global consumption, consumer electronics companies need to bring about this consumption fulfillment for many different user groups distributed around the globe. The worldwide availability of the internet, combined with cheap and easy travel and transport of people and goods around the world, result in an open market situation for consumer products in which country borders are absent. Nevertheless, use conditions and preferences may differ considerably among these distributed user groups. Moreover, research indicates that an important condition for successful global market penetration is the adaptation of the products to local markets [RUG01].

In addition to this global market orientation, increased international competition in the 1990s made it necessary for many consumer electronics companies to also manage their product creation activities globally [LU02]. On the one hand, consumer electronics companies opened factories in so-called “low wage” countries in order to reduce production costs. On the other hand, many companies now focus on their core-business and outsource the other activities globally. Nowadays, an increasing number of companies do not only outsource production but also parts of the development process [DAV05]. The immense growth of India’s information technology (IT) industry is a visible result of this trend to outsource software development activities to third parties. Several leading consumer electronics companies have outsourced the development of their embedded software to local companies in India. From 2002 till 2003, worldwide revenue of IT services grew less than 2 percent.
However, India's IT service industry experienced 22 percent revenue growth [NAS03]. The outsourcing of IT development by multinational corporations is driving this rapid growth [BHA05].

An important advantage of outsourcing is the increased flexibility of the outsourcing company and a reduction in overhead costs [PET03]. Nevertheless, there are some serious risks attached to the outsourcing process. Outsourcing can cause unwanted dependencies on suppliers [PET03]. A survey among large U.S. companies that internationally outsource IT work revealed that a majority of the outsourcing companies experiences communication difficulties and planning issues [ENG03]. Effectively managing outsourced relationships has become a core competence itself. Insufficient communication and unclear interfaces between the products of outsourcing and supplying companies may result in serious product failures. Communication should be considered as one of the biggest challenges in outsourcing product development. Especially the difference in time zone, culture and language makes communication between the outsourcing and supplying company very difficult [CHR07].

Moreover, it is often assumed that product development projects can be specified unambiguously. Subsequently, based on this specification, (parts) of the development process can be outsourced. But requirement specifications contain a huge amount of implicit knowledge. It takes time to make this implicit knowledge explicit and to transfer it. In addition, specifying requirements to a level that is useable for the third party company is often considered a demanding and very time consuming task [CHR07].

**Time-to-Market**

Another relevant characteristic of the consumer electronics industry is the ever increasing pressure on time-to-market. Development time determines how responsive a company can react to competitive forces and to technological developments, as well as how quickly a company will receive returns from the development effort [ULR00]. In order to keep up with competitors, it has become essential for companies to introduce more products to the market faster. Several years ago, the development process of a new television (TV) set used to take almost four years. Nowadays, a new TV set is developed within a timeframe of six months or even shorter. Brombacher states that “the challenge for manufacturers has become to maximize products profitability by minimizing time to profit” [BRO00]. Most companies have translated this challenge in the approach of bringing as many products to the market as possible in the shortest possible time. The general rule is that for each month that is cut from the development process of a product, up to one month is added to its sales life [SMI98].

This increased pressure on time-to-market has many implications for the product development process of consumer electronics companies. For these projects less time is available to perform quality tests before market introduction and due to time constraints the implementation of the test results is not a matter of course [MIN05]. Furthermore, incremental improvements to next product generations is troublesome because field feedback information about the previous product generation is not yet available at the concept release date of the next generation of the consumer electronics product [BRO04A].

**Quality of Consumer Electronics Products**

Recent research has demonstrated the increasing number of customer complaints on new products in consumer electronics industry [OUD06]. Moreover, analysis of these complaints indicates that to an increasing degree the cause of the complaint can not be retrieved [BRO04], see Figure 1.1.
This group of customer complaints for which no cause can be determined is denoted as No Failure Found (NFF). Research into this increasing number of customer complaints by Den Ouden indicates that 85% of the complaints can be traced back to decisions made in the product creation process [OUD06]. In other words, most of the customer complaints in consumer electronics are predominantly caused by a wrong decision in the product creation process.

![Graph showing percentage NFF in modern high-volume consumer electronics](image)

Figure 1.1: Percentage NFF in modern high-volume consumer electronics [BRO04]

In order to reduce the number of future problems with consumer electronics products, Den Ouden suggests improving the decision processes in the product creation process by supporting it with up-to-date and rich information about customer use preferences. One approach to acquire this information is to collect high quality feedback information in the aftermarket/service processes. Subsequently, this information can be used for deciding on the different product improvement options based on the impact on consumer satisfaction and improvement implementation effort [OUD06]. The selection of product improvement options does not only relate to next generation product improvements. In the phase before the market introduction of an innovative consumer electronics product, more product improvement options are available than time to implement them [KET02]. In order to overcome this problem, several implementation decisions should be made based on expected consumer satisfaction impact and implementation effort. Wrong decisions in this phase of the new product development (NPD) are likely to cause customer complaints after market introduction.

Clearly, in the context of consumer electronics more insight is required into these product improvement decision processes. Moreover, a better understanding of customer satisfaction and complaining behavior is needed to predict the impact of product design decisions on the number of customer complaints [KET02], [OUD06]. Based on these earlier findings, the ultimate goal of this thesis is to contribute to the overall quality improvement of consumer electronics products. By gaining insight into and improving the product improvement decision process and the underlying customer satisfaction models in the NPD processes, the quality of the product improvement decisions can be improved.

The next chapter illustrates the problem of the growing market uncertainty in consumer electronics industry. This problem implicates that the combination of the earlier mentioned business trends (complexity, product understanding, globalization and time-to-market) makes it difficult to make correct product design decisions from a user perspective. Chapter 2 demonstrates that in order to deal with this increased market uncertainty in consumer electronics, existing quality methods from literature and practice should be complemented with a user-oriented impact assessment of the identified quality problems (and the
accompanying quality improvement decisions). The combination of the existing quality methods with an accompanying user-oriented impact assessment would substantially lower the market uncertainty in the design process. Eventually, this approach should contribute to a decreasing number of customer complaints in consumer electronics. The development of this user-oriented impact assessment method is the main contribution of this research.

1.2 The Trader project

The research work presented in this thesis has been carried out as a part of the Trader project under the responsibility of the Embedded Systems Institute. This project is partially supported by the Dutch government under the Bsik program. Trader is a collaboration of several industrial and academic partners.

The previous subsection illustrated the explosive increase in complexity of consumer electronics products. Organizations in the professional electronics industry have found how difficult it is to put a product with high software content on the market without any defects [ROO96]. Nowadays, best in class software companies produce software code with roughly one bug in every thousand lines of code. Commercial software typically has 20 to 30 bugs for every 1000 lines of code, according to Carnegie Mellon University's Cylab Sustainable Computing Consortium. Given the current software content of consumer electronics of around 1 000 000 lines of code, undesired product behavior seems unavoidable.

The overall goal of the Trader project is to minimize product failures that are exposed to the user in consumer electronics, in particular in digital television. Given the inevitability of faults in the complete software content of consumer electronics products, the Trader approach to achieve this overall goal is to aim for:

- A technical restriction of the impact of a software fault on the overall operation of the consumer electronics product
- An user-centered approach to identify most important product failures on which product improvement efforts can be concentrated

The focus in this thesis is on this second approach. The apparent link between the goals of the Trader project and this thesis is the contribution to better consumer electronics products. However, in this context better implies an improved product design in terms of failures according to the actual users of the product.

Before explaining the general structure of this thesis, it is important to define some of the key concepts with respect to quality improvement in consumer electronics.

1.3 Definition of concepts

The complexity of current consumer electronics development processes requires the adoption of a multidisciplinary research approach [VIS02]. The research approach taken here is drawn from different disciplines such as industrial design and engineering, information technology and technology management. In order to prevent misapprehensions with respect to the meaning of different theoretical concepts, this section presents some important definitions of the concepts that are used in this thesis.

In order to contribute to the overall quality improvement of consumer electronics products, it is important to agree upon the definition of product quality.
It is clear that over the last decades, businesses have placed increasing emphasis on product quality improvements. Nevertheless, no full agreement exists on the meaning of quality in practice. Historically, manufacturers defined quality as “the degree to which a specific product conforms to a design or specification” [GIL74]. However, as a result of the increasing internal complexity and growing customer requirements of consumer electronics products, a complete coverage of the product specifications is unfeasible. Therefore, the traditional quality definition has been extended. Most consumer electronics companies have adopted this extended quality definition which defines quality as “being able to design and deliver a product to the customer that is conforming to his/her expectations” [PDA04]. In this thesis, the extended quality definition is applied.

In this extended quality definition, the role of the customer is crucial. In literature, often the distinction is made between the customers, consumers and users of a product. In this distinction, customers are the people that buy a certain product; consumers are the individuals that purchase and use the product; and users are the people who actually use the product [WEB05]. However, in this thesis, unless stated differently, the words “customer” and “consumer” are used as synonyms for user.

Lastly, it is important to stress the difference between the concepts of product fault, product error, product failure and customer complaint. The definitions of these four concepts originate from information technology literature [LAP85], but are just as valuable in the consumer electronics context. A product fault is a mistake made during the development of a product. Such a mistake is a potential cause of a product error. Faults are dormant and can become active. When a fault becomes active, it causes an error [DEC06]. An error is the manifestation of something actively wrong in the product. Often errors lead to new errors, which eventually may lead to product failure. A product failure is the situation where the customer may decide to report the fact that the product is not conforming to his/her expectations [PET03], resulting in a customer complaint. Summarizing, the distinction between a product error and failure is the observability of the problem by the user. The possible escalation of a product fault into a product complaint is illustrated in Figure 1.2.

![Figure 1.2: Escalation of product fault into product complaint [DEC06]](image)

Given the technical origin of this model, it is important to emphasize this thesis' broad definition of a product failure. In this definition, a product failure is a situation in which a customer is using a product that is not conforming to his/her expectations. In the case of a non-technical failure/quality problem the meaning of the preceding product fault and error becomes artificial.
1.4 Thesis structure

In this first chapter, the general goal of this dissertation is formulated as contributing to the overall quality improvement of consumer electronics products. By gaining insight into the product improvement decision process and the underlying customer satisfaction models in the product development processes, this thesis aims to improve the quality of the product improvement decisions by incorporating the user perspective into this decision process.

In order to achieve this objective, the remainder of this thesis is organized as follows. In Chapter 2, the current situation with respect to quality improvement in consumer electronics is illustrated based on the results of a literature study and a case study in consumer electronics industry. Based on these results, the general research goal is translated into a problem definition and corresponding research questions. Subsequently, the research design and approach are presented.

Based on the findings in Chapter 2, Chapter 3 introduces a new concept to support the product improvement decisions process; the so-called User Perceived Failure Severity (UPFS). The theoretical and practical foundations of this concept are presented together with its intended implications on the product improvement decision process.

Chapter 4 introduces a first conceptual UPFS model based on insights from actual product improvement decision processes in practice. This model is tested in an experimental setting resulting in several adjustments to the conceptual UPFS model.

This adjusted model is presented in Chapter 5. In this chapter the first experimental results are combined with different theoretical aspects into a new conceptual UPFS model. In order to stress the scientific contribution of this part of the research, the identification and validation of the measurement approaches for the different variables of this experimental model are presented in the next chapter; Chapter 6. In Chapter 7 and 8 parts of this model are validated in an experimental set up. Based on the results of these experiments, a validated, partial UPFS model is presented in Chapter 8.

Lastly, in Chapter 9 the implications of this validated UPFS model on the product improvement process in consumer electronics industry is explained. The main conclusions of this thesis are drawn and some suggestions for future research are given.
2 Quality Improvement in Consumer Electronics

The previous chapter introduced recent business trends in the consumer electronics industry. The increasing time-to-market pressure, globalization and technical complexity of the products combined with decreasing product understanding of users has resulted in an increasing number of previously unknown consumer complaints on especially new products in this industry. This chapter explains that relationship by illustrating the implications of the business trends on the quality improvement processes in consumer electronics industry. Moreover, these implications also result in some specific requirements for successful implementation of quality improvement methods in consumer electronics.

Subsequently, based on these requirements, the advantages and disadvantages of existing quality improvement methods as mentioned in literature are explored. This theoretical literature review is followed by three explorative case studies investigating the quality improvement process in current industrial practice. The combination of the identified requirements for successful quality improvement adoption in consumer electronics and the theoretical and practical limitations of the current methods result in the problem statement and research questions of this dissertation.

2.1 Implications of trends on quality improvement

The combination of the earlier described business trends in consumer electronics makes it increasingly difficult to adequately manage product quality [BRO04]. On the one hand, customers demand more sophisticated and user-friendly products that have to be realized with decreasing development times. On the other hand, product development processes are getting more complicated as a result of the globalization and outsourcing trend. The combination of the earlier presented business trends has drastically changed the main objectives of the consumer electronics product development process, see Figure 2.1.

![Product Development Process](image-url)

*Figure 2.1: Trends in the product development process of consumer electronics products*
The development process of the simple hardware-based products of the 1980's/1990's was mainly cost oriented and consequently required a strong focus on product manufacturing optimization. In this period, the market requirements were known and resulted in stable, long term (10 year) product roadmaps. The main goal was to make products that comply with the technical specifications. Therefore, in this period, product defects always implied specification violations and the overall customer dissatisfaction level could be expressed by the required number of product repairs in the after-market. However, the increasing complexity of current software-based consumer electronics products and the increasing customer expectations result in unknown market requirements and an exponential growth in product functionality. Moreover, the development process has become more design oriented in which the main focus is on the product user. In other words, based on the trends in consumer electronics, the focus in the product development process has shifted from a specification focused manufacturing approach to a user focused design approach. Moreover, current reported product defects do not only represent violations of the product specifications but also uncovered customer requirements and unexpected product behavior. Consequently, the number of actual product repairs in the after-market (to deal with specification violations) is no longer a representative measure of the overall customer dissatisfaction level.

This gradual shift in focus of the product development process in consumer electronics can be illustrated by the simultaneous increase in the level of uncertainty in the product development process. Galbraith defined uncertainty as “the difference between the amount of information required to perform a task and the amount of information already possessed by the organization” [GAL73]. Historically, consumer electronics companies have been dealing with three different types of uncertainty in their product development process [MUL98]:

- **Industrial chain uncertainty**: uncertainty over the level of resources which should be committed to achieve success with the new product in the market
- **Product technology uncertainty**: uncertainty as to whether the firm has chosen the best technological and market paths to lead to prosperity in the market
- **Market uncertainty**: uncertainty over the nature and extent of customers’ needs for new products made possible by the application of new technologies

In the 1980's, the main emphasis of consumer electronics companies was on the reduction of industrial chain uncertainty. By this time, technological development was gradual and product diversity was low resulting in low market- and product technology uncertainty. In this period, companies were mainly concerned with optimizing their profit/investment ratio by answering the question: “Are we making our products the optimal way?” In this decennium, product development uncertainty can be characterized as “marginal uncertainty” and was mainly managed by different quality control methods like statistical process control.

In the 1990's, the speed of technological development increased resulting in a wider variety of technological solutions and shorter product development times. Nevertheless, the product preferences of the user were known and product-user interaction was fairly predictable. However, these developments increased the product technology uncertainty in the development process. In this period, companies were mainly concerned with optimizing their product by answering the question: “Are we making the product with the right settings?” In this decennium, product development uncertainty can be characterized as “parametric uncertainty” and was mainly managed by different quality methods like Design-Of-Experiments, Taguchi, Total Quality Management (TQM) and Six Sigma.
In the 2000's consumer electronics companies have to deal with all the earlier mentioned trends. Moreover, many businesses in the consumer electronics industry in the western world have chosen the strategy to bring more innovative products and functions to the market [OUD06]. The decision of these companies to develop “new-to-the-world” products can be best explained through the introduction of the product life cycle model. Product life cycles follow a generic pattern of introduction, growth, volume, enhancement, decline, replacement and obsolescence. This product life cycle can be divided into three general phases: embryonic, growth and maturity, see Figure 2.2 [MIN05].

![Product Life Cycle Phases](MIN05)

This product life cycle has been portrayed as a “technology adoption life cycle” in which the behavior of customers designated as “innovators”, “early adopters”, “early majority”, “late majority” and “laggards” critically affect the uptake of a new product [ROG62], [MIN05]. Along these phases of the product life cycle model, from the embryonic till the maturity phase, the product adoption numbers increase but the expected profit per product decreases. A successful business of mature products requires high product sales numbers combined with low product costs. On the other hand, a successful embryonic product requires high profits per product and distinguishing product features.

Over the last decades, many Asian and East European companies have optimized the production of mature consumer electronics products. As a result of the low production costs of consumer electronics products in these countries combined with improved technical know-how of the involved companies, it is difficult for western consumer electronics companies to compete with these companies on mature products. However, the development of innovative or embryonic products requires more innovative business processes and technical knowledge [COO99]. As a result, many western consumer electronics companies focus on the development of new or “embryonic” products with a high level of innovativeness. This product newness and higher levels of innovativeness are linked to higher levels of uncertainty and risk, to new product development difficulty and performance and to the required resources for undertaking new product development activities [BRE00]. Moreover, this strategy requires the introduction of new products before the maturity phase of the previous product generation and reduces the possibilities to learn from mistakes made in the development process.

Also, by the introduction of the internet, the global market is emerging. Consumers explore foreign markets via the internet and buy products from different geographical regions.
As soon as a product becomes available in one region, consumers from other regions acquire the product via the internet. This makes the product introduction for a restricted target market unfeasible. Consumer electronics businesses also realize that the local introduction of their product makes it available for many different user groups worldwide. Given the increased time-to-market pressure, most companies can spend only limited time to adapt products to specific target groups. As a result, consumer electronics products should fit to a large extent to the complete population of user groups.

Furthermore, the new product should be embedded in a complete network of existing products in the use environment. A good illustration of this trend is the (incomplete) description of compatible products for a new TV set as presented in Figure 2.3. For each of the compatible products, a similar compatibility picture can be shown. The complexity of this picture increases if the different types and brands of the compatible products (with different specifications and characteristics) are added to the picture. This is illustrated by the dotted lines in Figure 2.3.

![Figure 2.3: Complexity of the product environment; products, types and brands](image)

The combination of the technological complexity of the product itself, the complex environment in which the product should operate and the diversity in user groups that may possibly adopt the product, makes it very difficult to predict the user requirements of a new consumer electronics product. Moreover, it is impossible to foresee all future product-user interactions already in the product development process. In other words, the current trends in consumer electronics industry have resulted in an increasing level of market uncertainty.
Companies these days are mainly questioning themselves: “Are we making the right products for our customers?” Therefore, product development uncertainty can presently be characterized as “structural” uncertainty. This shift from industrial chain uncertainty to market uncertainty in consumer electronics product development over the past 30 years is correlated with the trend in consumer complaints in this industry. Resulting from the successful application of quality (control) methods in the 1980’s and 1990’s, overall product quality was having its all time high in the middle of the 1990’s, measured in terms of consumer complaints (see Figure 2.4). But soon after that, the number of complaints started to increase again. Moreover, analysis of these complaints indicates that to an increasing degree the cause of the complaint can not be retrieved [BRO04] (see Figure 1.1). Research indicates that the available approaches are no longer sufficient to design products that meet the consumers' expectations [OUD06].

The causal chain from business trends till increasing number of customer complaints is represented in Figure 2.5. The combination of the earlier mentioned business trends cause an increase in market uncertainty over the nature and extent of customer's need for new products. This uncertainty makes it very difficult for designers to assess the impact of design decisions on customer dissatisfaction. This is the case for early decisions in the concept selection process as well as for later quality improvement decisions in the product validation phase. This uncertainty in the decision process eventually causes “incorrect” design decisions; decisions that have a negative impact on the customer dissatisfaction level. In the end, these wrong design decisions result in an increasing number of customer complaints. This conclusion is confirmed by Den Ouden, who states that 85% of the complaints on consumer electronics products can be traced back to decisions made in the product creation process [OUD06].
Based on this combination of findings from earlier research, it can be concluded that in order to improve quality in consumer electronics, this market uncertainty should be better dealt with. In other words, the gap between the amount of information needed to make a design decision and the amount of information available in the company should be reduced. In the context of product quality control, the main denotation of market uncertainty is that product designers lack insight into what the users actually want. Accordingly, the required information should contribute to the impact evaluation of a certain design decision on the customer dissatisfaction level and consequently to a decreasing number of customer complaints.

This analysis of the current situation in consumer electronics results in several requirements for potential information collection methods. Based on the above mentioned market trends, it can be summarized that in order to reduce market uncertainty in a consumer electronics quality context, the method should be:

- **User-oriented**: In order to reduce market uncertainty, the expected impact of a design decision on customer (dis)satisfaction should be determined. Evidently, this requires a user-oriented approach.
- **Specific**: The methods results should found concrete design decisions. Therefore the method should not be too general or high-level.
- **All-round**: The method should cover all aspects of market uncertainty in the context of product quality. All the product design decisions that may result in customer complaints should be supported by the information collection method.
- **Rapid**: In order to support fast time-to-market, this information should be collected and distributed in a rapid way. Furthermore, the collected information should be available in time at the right phase of the NPD process and should also support rapid decision making.
- **Complexity proof**: The method should be able to collect this user-oriented, specific but broad information on innovative (software based) products in a constant changing networked environment.

Several methods that explicitly aim to decrease market uncertainty in the product development process are presented in literature, [MOR89], [KAU98], [MUL98], [MAY99], [OZE99], [PET03], [CLA05], [FOL07] and others. All these methods can be grouped into the following three categories based on their application moment in the product development process [LUI03]:

- **Consumer expectation analysis**: Early in the product development process, consumer needs and expectations are analyzed by the marketing department. Although the methods in this category have proven their usefulness in the development process of relatively simple and mature products, their value for innovative consumer electronics products is restricted. This limited value is caused by the “fuzzy front-end”; the first steps of product development processes of these innovative products are often highly uncertain and fast evolving or chaotic by nature [CLA05]. Furthermore, it is very difficult for customers to predict their requirements for innovative products [MOR89], [MUL98]. Although the contribution of these methods to a business' market intelligence knowledge may be quite valuable, the total contribution to market uncertainty reduction in consumer electronics development is very limited. As a result, in this research, the consumer expectation analysis methods are left out of consideration.
- **Concept/product testing**: These tests are performed during product development. The main goal of these tests is to evaluate a concept or product on consumer requirements. In this category, a broad scope of methods is available from literature [KAU98], [OZE99], [DAH02]. For example, within consumer electronics industry, the interest for usability testing is growing [ROS02], [NIE03], [VAL07]. But also other methods like focus groups, beta tests and High Contrast Consumer Tests are commonly applied in this field.

- **Field feedback information**: Field feedback is “the information related to the performance of products in interaction with customers” [PET03]. Industry has two direct interfaces with customers: marketing/sales and service. Marketing/sales are mainly success oriented: “What will sell our products”. Service, on the other hand, is concerned with product performance and customer dissatisfaction. Therefore service may provide useful field feedback information. Service information is collected and analyzed after the product's market introduction. Possible service feedback information sources are product service centres, helpdesks, internet forums and focus group tests [PET03], [OUD06]. In literature, the results of general evaluations of the suitability of field feedback information for quality improvement are conflicting. Petkova emphasizes the important limitations for the application of field feedback information, like the lack of root-cause information and the time delays [PET03]. However, Koca indicates that field feedback is currently collected more effectively and quality improvement opportunities are promising [KOC07]. Nevertheless, the amount of literature evaluating the suitability of feedback information for quality improvement is very limited.

Unlike most methods, the User Centered Design (UCD) approach can be applied in all phases of the product development process [FOL07]. The ISO 13407 standard defines UCD as a multidisciplinary activity, which incorporates human factors and ergonomics knowledge and techniques with the objective of enhancing effectiveness and productivity and improving human working conditions [ISO99]. This adoption of UCD in different phases of the development process is also expressed by the different phases of the usability engineering lifecycle: requirements analysis, design/testing/development and installation [MAY99]. However, the UCD approach is mainly success-oriented (with the focus on user satisfaction) and the increasing market uncertainty in consumer electronics requires a failure-oriented approach (with the focus on user dissatisfaction and complaint behavior). Nevertheless, the different usability (testing) goals that are applied within UCD could be helpful to decrease market uncertainty in the product development process.

Notwithstanding the promising contribution of all the earlier mentioned methods to product quality improvement, it is unclear whether these methods directly contribute to the impact evaluation of design decisions in the product development process in consumer electronics. Currently, it is unclear whether the possible outcomes of these methods can guide the product design decision process in consumer electronics industry by somehow predicting the impact of conceivable decisions on the overall customer satisfaction level. In other words, to what extent are these methods conforming to the above formulated requirements? The term “to what extent” is explicitly chosen in this sentence because it is important to mention that the complexity of current (embryonic) consumer electronics products and development processes makes it almost impossible to fully comply with all these above mentioned requirements for information collection methods.
Moreover, it is unclear how other quality oriented improvement methods (that are not explicitly focused on market uncertainty reduction) may contribute to the impact prediction of design decisions in consumer electronics industry. Resulting from this indistinctness of the potential contribution of existing methods to market uncertainty reduction, the next section presents a literature review evaluating the capability of these methods. This evaluation is based on the above mentioned requirements for potential information collection methods.

The potential contribution of quality feedback information to market uncertainty reduction is difficult to determine from the existing literature. The amount of literature dealing with feedback information is limited, conflicting, and unclear about the methods capacity for uncertainty reduction in consumer electronics. In order to evaluate the suitability of these feedback methods, three case studies were performed at feedback information providers, namely consumer electronics service centres. In the third section, the evaluation results of these case studies are presented. Based on the results of the literature review and case studies, some conclusions are drawn on the advantages and disadvantages of these methods in their ability to reduce market uncertainty in the product development decision process. These results are summarized in section four in the final problem definition of this research and the corresponding research questions.

2.2 Quality improvement methods in literature

As indicated in the previous section, several methods for market uncertainty reduction in the product development process are described in literature. Complementary to these methods are some quality oriented methods that could possibly contribute to this same uncertainty reduction. In general, these methods are elaborately dealt with in academic literature. However, it is unclear to what extent these methods conform to the above formulated requirements for market uncertainty reduction in a consumer electronics quality context. To what extent are these methods able to manage the recent trends in consumer electronics industry? In the next subsections, the general approach and characteristics of several relevant methods are described. Subsequently, their potential contribution to market uncertainty reduction in the consumer electronics industry is evaluated based on the above mentioned requirements for these methods; user-oriented, specific, all-round, rapid and complexity proof.

It is not the purpose of this section to give an exhaustive literature overview of existing quality methods and their application areas. The main goal of this section is to demonstrate the general strengths and weaknesses of existing methods with respect to market uncertainty reduction in consumer electronics. For that purpose, only a selection of relevant methods is presented.

2.2.1 Concept and product testing methods

A large number of methods and tools is available in academic literature for the evaluation of product concepts and products. In this subsection, two commonly applied evaluation tests are presented, namely usability testing and concept testing. These methods are complemented with the description of a relatively new approach, the so-called High Contrast Consumer Test.

Usability testing

The formal definition of usability is “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” [ISO99].
Usability testing is a widely used technique to evaluate user performance and acceptance of products and systems. Originally, the usability testing approach was applied to the design and evaluation of software applications and user interfaces [NIE93], [HAN01]. The usability concept evolved when software companies realized the market opportunity of “user friendly” products. Resulting from the increased customer requirements and product complexity, the consumer electronics industry has adopted the usability concept as well.

Usability has been applied to consumer electronics products and companies now consider it an important key to their success [HAN01]. However, Han et al. illustrate that there is a difference in the concept of usability between the software and consumer electronics applications [HAN01]. The main differences between these two usability application areas is that for consumer electronics products physical product aspects are important as well. Additionally, consumer electronics products are used in a broader environment than software products. Therefore, Han et al. define the concept of usability to include two aspects of usability: the performance and the image and impression aspects [HAN01] (see Figure 2.6).

![Figure 2.6: Usability of consumer electronics products consists of two aspects: performance and image/impression [HAN01]](image)

Nielsen stresses the importance of realizing that usability is not a single, one-dimensional property of a product [NIE93]. According to Nielsen, usability has multiple components and is traditionally associated with five usability attributes:

- **Learnability:** The product should be easy to learn so that the user can rapidly start using the product.
- **Efficiency:** The product should be efficient to use so that once the user is acquainted with the product, a high level of productivity is possible.
- **Memorability:** The product use process should be easy to remember, so that the user is able to return to the product after some period of not having used it, without having to learn everything all over again.
- **Errors:** The product should have a low error rate so that users make few errors during the use of the product. Moreover, if users do make errors they should be able to recover easily from them. Further, catastrophic failures must not occur.
- **Satisfaction:** The product should be pleasant to use so that users are subjectively satisfied when using it.
Traditionally, usability is measured by having a number of test users (selected to be as representative as possible of the intended users) operate the system to perform a prespecified set of tasks [NIE93]. In current usability literature, several methods for usability testing are presented. Based on their testing approach, these methods can be classified as being analytical (evaluation without real users) or empirical (evaluation with real users). Table 2.1 on the next page gives an overview of the most common usability evaluation methods.

Usability testing is a reliable way to quantitatively estimate users’ performance and subjective satisfaction with products [WIC00]. Jokela argues that usability may remarkably impact customer satisfaction and product reputation. Based on the well-known Kano model it is stated that meeting must-have usability is the minimum requirement for today’s consumer electronics products [JOK04]. However, achieving this requirement is not straightforward in current consumer electronics industry.

<table>
<thead>
<tr>
<th>Analytical Evaluation Methods</th>
<th>Empirical Evaluation Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic evaluation</td>
<td>User testing</td>
</tr>
<tr>
<td>Guideline reviews</td>
<td>User feedback</td>
</tr>
<tr>
<td>Plurastic walkthroughs</td>
<td>Focus groups</td>
</tr>
<tr>
<td>Consistency inspection</td>
<td>Online focus groups</td>
</tr>
<tr>
<td>Standards inspection</td>
<td>Prototyping</td>
</tr>
<tr>
<td>Cognitive walkthroughs</td>
<td>Thinking aloud</td>
</tr>
<tr>
<td>Formal usability inspection</td>
<td></td>
</tr>
<tr>
<td>Feature inspection</td>
<td></td>
</tr>
</tbody>
</table>

The main difficulty with usability testing is the fact that the target market and functionalities of the product under development are specified before the usability test is performed [VRI94], [VAL07]. So market uncertainty reduction is only partly attainable. Furthermore, the complexity of consumer electronics products and the increasing interactions between these products via wired and wireless connections make the traditional approaches of ensuring usability (laboratory-based testing) largely meaningless in this context [VRI94], [ZIR94], [THO02].

The potential contribution of usability testing to market uncertainty reduction in consumer electronics is evaluated as follows: the user-orientation is present especially in the empirical usability approach. However, the target market and product functionalities are largely fixed before performing the usability evaluation. Nonetheless, actual users are evaluating product performance and product image before market introduction.

Although the execution of a usability evaluation test requires extensive preparation, in general usability testing is not characterized as being too time-consuming. The traditional focus of the usability testing approach on user-friendliness results in the identification of mainly non-technical quality problems.
The main contribution of the usability testing method to the product design decision process is an extensive overview of non-technical product quality problems causing restricted product usability. However, the impact of these individual problems on customer satisfaction can only be estimated based on users' remarks and testers' appraisal. As indicated before, the increasing complexity of consumer electronics products and their environment makes the traditional laboratory-based usability approach less meaningful. Nevertheless, in this context many researchers propose useful supplements and adjustments to this traditional usability approach [ZIR94], [WIC00], [THO02], [VAL07].

Concept testing
Even though usability testing and concept testing are both concerned with market uncertainty reduction by means of testing, their approach is rather different. On the one hand, usability testing is concerned with the performance and image evaluation of a specific product design. On the other hand, concept testing estimates the consumer reactions to a product idea before further developing this idea into a product [MOO82]. In general, concept testing can be defined as any number of marketing research procedures which test some sort of stimulus as a proxy for a new, revised or remarkeleted product. So concept testing is supposed to provide a market evaluation of the ideational content itself rather than any particular execution(s) of that content [IUS75].

The main goal of concept testing is to estimate consumer reactions to a product idea before committing substantial funds to it. Furthermore, concept tests are used to determine the potential target market, purchase interest and how the concept might be improved [ACI81], [MOO82], [PAG92]. The general approach for concept testing is a combination of concept presentation and consumer evaluation. The concept evaluation starts with the presentation of the concept to a (group of) potential user(s). This presentation can be a written description, a collage of related pictures, a physical product or even a concept demo movie. Subsequently, users are questioned about their opinion about the concept and their possible buying intention. These questions can be posted via direct questioning, in (web) questionnaires or during discussions in focus groups. Several statistical approaches are used to summarize these results in a description of the market potential of the proposed concept(s).

The difficulty in concept testing is that tests can evaluate the concept only as interpreted through its specific executions (e.g. the product, the description or another presentation). This makes it very difficult to determine if the attempted evaluation of the concept is mainly a function of the way in which it is executed or if it is an evaluation of the concept itself [HAL71]. Furthermore, the main focus of concept testing is the evaluation of the market potential of a product concept. So, although concept testing can greatly contribute to the development of promising product concepts, methods for concept quality improvement are still relatively underexposed in this approach [DUK94]. Besides that, concept tests are generally unable to predict market failures, because of the changing conditions in the product concept or its environment [TAU75], [DUK94]. Moreover, it is very difficult for consumers to predict their buying intention or product acceptance especially for innovative and complex products [TAU75].

Based on this analysis, the contribution of the concept testing approach to market uncertainty reduction in consumer electronics is summarized as follows: The user-orientation is certainly present in this approach. Contrary to the usability approach, product functionalities are not yet fixed at the moment of performing a concept test. Any possible concept adjustments are still implemental in this phase of product development.
Like the usability testing approach, concept testing requires extensive preparations but in general usability testing is not characterized as being too time-consuming. The focus of the concept testing approach is on concept utility and customers’ buying intention. Concept quality improvement is largely ignored in this approach. Nevertheless, concept testing specifically supports the concept selection decision by providing quantitative information about consumer acceptance and buying intentions of different concepts. Unfortunately, this information becomes highly unreliable when the degree of product innovation and complexity increases.

**High Contrast Consumer Test**

High Contrast Consumer Test (HCCT) is a consumer testing concept which tries to deal with unexpected customer behavior. HCCT is based on the assumption that it is not likely that “normal” products fail while being used by “normal” consumers [BOE03]. Therefore, these HCCTs are designed to maximize the variability in the interaction between product and consumer in order to provoke, early in the product development process, failures that would normally occur in the field. In a HCCT, extreme consumers are observed using products in realistic operating conditions. The purpose is to accelerate failures and expose product-usage issues as early as possible [BOE03]. An extreme consumer is a consumer that uses the product in an extreme manner. During the HCCT, these extreme consumers are using extreme products as well. Extreme products are fully functional products but are close to the technical specification limits. The HCCT approach focuses on the first product use phases, namely unpacking, installation and first use. The extended use phase is not included in this approach.

This new approach of consumer testing generates some advantages compared to the classical consumer testing approach [BAS03]. The HCCT approach may improve the understanding of how consumers use the product. Moreover, a relatively wider range of potential design issues may be identified earlier in the product development process.

However, in the context of this research, the application of the HCCT approach also raises some difficulties. The focus of the HCCT method on the first product use phases results in the identification of mainly non-technical quality problems. Most non-technical problems come up in the first phases of the product use process when the product is unpacked, installed and used for the first time. That is why the application of HCCT for the identification of technical quality problems is probably more difficult. Most likely, this would require substantial longer test times and more resources.

With respect to HCCT’s potential contribution to market uncertainty reduction, the following can be stated: User-orientation is certainly present in the HCCT approach since actual (extreme) users are testing the product. Moreover, the combination of extreme users with extreme products results in shorter test times and consequently supports fast time-to-market. Conversely, increasing product/environment complexity results in longer test times or lower test coverage. The design decisions that can be supported by means of the HCCT approach are mainly related to non-technical quality problems. The main contribution of the HCCT method to the product design decision process is an extensive overview of non-technical quality problems present in the product. However, the impact of the presence of these individual problems on consumer satisfaction is indistinct.
2.2.2 Related quality methods

The number of quality improvement methods described in academic literature is almost uncountable. However, since these methods originated in different industrial periods their approach to improve quality can be very dissimilar. As presented in Section 1.1, in the 1980's the main emphasis of consumer electronics companies was on the reduction of industrial chain uncertainty. Consequently, the corresponding quality methods were mainly concerned with quality (process) control.

In the 1990's, the main focus was on product technology uncertainty reduction in the development process. In this period, consumer electronics companies were mainly concerned with optimizing their product and finding the right production settings. Corresponding quality methods are for example Design Of Experiments (DOE), Six Sigma and Total Quality Management (TQM). Because of this difference in focus, most of these earlier quality methods are unsuitable for structural uncertainty reduction. Nevertheless, some (parts) of these methods seem promising with respect to their potential contribution to market uncertainty reduction. This subsection gives an overview of three of these quality improvement methods, namely Quality Function Deployment, Failure Mode and Effects Analysis and Software Reliability Engineering.

Quality Function Deployment (QFD)

QFD is “an overall concept that provides a means of translating customer requirements into the appropriate technical requirements for each stage of product development and production (i.e. marketing strategies, planning, product design and engineering, prototype evaluation, product process development, production and sales)” [SUL86]. Since its original development in Japan in the late 1960's the QFD method has gained considerable popularity. Currently, QFD is the most frequently mentioned tool for achieving customer-oriented quality and consequently, a vast literature basis on QFD has evolved [CHA02], [OUD06].

The main principle of the QFD method is the belief that products should be designed to reflect customers' desires and tastes [HAU88]. The method provides simple tools to translate customer requirements into technical specifications. However, the massive adoption of the QFD method in business contrasts sharply with the disappointing results of the method as published in academic literature. Where industry practitioners claim that the QFD method decreases the development time and costs of new products, Griffin demonstrates that only 27% of the QFD projects bring the expected results [GRI92]. Successful application results of the QFD method are most probable when the method deals with simple (not complex) service projects (not physical goods) that aim for incremental changes (no clean sheet design) [GRI92].

Although the communication between different functions (like marketing, manufacturing, engineering, etc.) improves by applying QFD the tool soon becomes an administrative burden especially when the QFD method is applied to clean sheet designs of complex physical products [GR193], [OUD06]. Nevertheless, currently many managers in industry promote the use of the QFD method in order to tackle the increasing number of consumer complaints [OUD06]. The potential contribution of the QFD method to market uncertainty reduction in consumer electronics is evaluated next. The user-orientation is at the basis of the QFD method. As mentioned earlier, the main principle of the QFD method is that products should be designed to reflect customers' desires and tastes.
Nonetheless, the actual translation of these customer needs into product specifications remains ambiguous and arbitrary. Regularly, in this step, valuable information about customer needs is lost.

The fact that the QFD method is often considered as an administrative burden in product development processes underlines its lacking support for fast time-to-market. Nevertheless, QFD information from previous projects could be used to speed up the development process of the current project. However, the successfulness of such an incremental approach within the ever changing context of consumer electronics is at least doubtful. The main focus of QFD is on the product utility decision process by translating customer requirements into product specifications. The QFD method can only provide support for this decision process for simple products. Consequently, the suitability of the QFD method for innovative consumer electronics products is questionable.

**Failure Mode and Effects Analysis (FMEA)**

FMEA is a traditional quality and safety analysis technique which has enjoyed extensive application to diverse products over several decades. In the late 1960's, the practice of using FMEA as a way to improve product design began to surface. The FMEA method is a systemized series of activities intended to discover failures and recommend corrective actions for design improvements [LEV03]. According to Statitis, “the FMEA is a specific methodology to evaluate a system, design, process or service for possible ways in which failures can occur” [STA03]. This evaluation can be done according to two approaches. The first is by using historical data: there may be data available of similar products and/or services, warranty data, customer complaints, and other appropriate information to define failures. The second approach, often resulting from the absence of historical data, makes use of mathematical modeling or simulations to identify and define failures. In essence, the FMEA provides a systematic method of examining all the ways in which a failure can occur in a product. The FMEA includes an overview of failure modes, possible causes of each failure, effects of the failure and their seriousness, occurrence (frequency) and likelihood of detection [BLI00].

The emphasis of the classical FMEA is on the identification and prevention of known and potential problems from reaching the customer. One important assumption that forms the foundation of the FMEA method is that problems have different solution priorities. The purpose of the method lies in finding that priority. The main advantage of the FMEA method is the revelation and prioritization of hidden problems that otherwise would not have been caught until the product had been in production for some time [LEV03]. The prioritization helps to identify critical failures that need to be solved before market introduction. Although most literature stresses the advantages of applying the FMEA method in current business processes, some critical comments can also be made regarding FMEA in the light of current trends in the consumer electronics industry.

With respect to FMEA's potential contribution to market uncertainty reduction, several remarks can be made. The first remark is concerned with the team performing the FMEA. Actual product users are not involved in the FMEA process. In general, an FMEA team is composed of the following disciplines: a design engineer, a manufacturing engineer, a test engineer and a quality and reliability engineer. These people are familiar with the product and its technologies. Based on their experience with earlier products, these engineers draw up a list of expected failure modes together with the prioritization of the failures. The failure prioritization is made regarding the expected consequence of a failure for the customer.
However, this complete prioritization is performed by engineers. These engineers can be biased by their technical background and historical knowledge of the product. In other words, the user-orientation fails because these engineers may not be able to prioritize product failures on behalf of customers.

Moreover, performing an FMEA requires a lot of preparation. For the FMEA process, the required people should be available within the company and a clear product description should be made. Furthermore, the actual FMEA process together with the implementation of the FMEA results can be quite time-consuming and labor-intensive. In other words, the FMEA method does not explicitly support short time-to-market. Moreover, the FMEA method mainly concentrates on the occurrence of technical failures which are product specification violations. The consequences of a product that does not comply with the expectations of a user are generally not considered in the FMEA method. On the other hand, the FMEA gives specific support to the quality improvement decision process. It not only describes possible product quality problems, it also results in a problem ranking based on expected failure severity. The application of FMEA to complex software based products can be quite problematic [GOD00]. Although several papers describe the application of the FMEA method for software based products [BOW01], [YAC02], [OZA04], the usefulness of the FMEA method for these products is often doubted. There is no general accepted approach for performing an FMEA on software designs [OZA04]. The general FMEA method is an approach for prioritizing product failures in the development process by mentally walking through the different product states and possible failures modes. Although this approach can be very helpful for pure hardware products, it is unfeasible for software based products like consumer electronics products. The state space of these products is immense even for software components with a limited size. Moreover, not all software states can be foreseen by the designers as a result of unexpected interactions between different parts of the software code or between parts of the software code and hardware modules.

Software Reliability Engineering (SRE)

SRE is “the quantitative study of the operational behavior of software-based systems with respect to user requirements concerning reliability” [LYU96]. According to the SRE approach making, a good reliability estimate depends on testing the product as if it were used in the field. In SRE, operational profiles are quantitative characterizations of how a system is used [MUS93]. In addition, the operational profile indicates how to improve quality by allocating development and testing resources to functions on the basis of use. In practice, an operational profile is a set of alternative product operations each with a probability of occurrence. Developing an operational profile involves five 'break-down' steps [MUS93], namely:

- Find the customer profile
- Establish the user profile
- Define the system-mode profile
- Determine the functional profile
- Determine the operational profile itself

These five steps are illustrated in Figure 2.7. In the SRE context, a customer is the person or group that acquires the system. The customer profile is the complete set of customer groups and their associated occurrence probabilities. System users are not necessarily identical to customers. A user is a person or group that adopts and not acquires the system.
A user group is a set of users who will employ the system in the same way. Subsequently, a user profile is the set of user groups and their occurrence probabilities. If there are multiple customer groups, each user-group probability should be multiplied by its customer-group to obtain its overall probability [MUS93]. A system mode is a set of functions or operations that are grouped for convenience in analyzing execution behavior. A system-mode profile is the set of system modes and their associated occurrence probabilities. The next step is the breakdown of each system mode into the function it needs, resulting in a function list. Subsequently, for each function the occurrence probability should be determined.

**Figure 2.7: Development of an operational profile**

A function is a task or part of the overall work to be done by the system as viewed by the system engineer and high-level designer [LYU96]. The functional profile gives a quantitative overview of the relative use of different functions. Lastly, an operation represents a task being accomplished by the system as viewed by the people who will run the system (also viewed by the testers, who try to put themselves in this position). The operational profile gives a quantitative overview of the relative use of different operations.

Subsequently, this operational profile should be used to allocate testing effort, to select tests and to determine the order in which the tests should be performed. Testing driven by an operational profile is efficient because it identifies failures, on average, based on their order of occurrence frequency [MUS93]. The occurrence frequency of a failure is primarily used to prioritize the failures.

This operational profile concept is especially concerned with the detection of software failures. However, this concept can also be employed for hardware testing. Customer profiles, user profiles, system-mode failures, functional profiles and operational profiles can also be formulated for hardware systems. The application of the operational profiles concept is therefore practical with respect to current complex products consisting of both hardware and software components.

With respect to SRE's potential contribution to market uncertainty reduction, several remarks can be made. The first remark is concerned with the limited user-orientation in the SRE method. Only the software test coverage is determined based on the operational profiles of the different user groups. The identification and evaluation of product failures is performed by software/hardware engineers.
Nevertheless, the SRE approach results in efficient test protocols based on the predefined operational profiles. Consequently, the SRE approach does support the trend of short time-to-market. Like FMEA, the SRE approach is only concerned with the identification of technical product failures. Although a severity prioritization is made based on the frequency of the failures found, a user perspective is lacking in the determination of the failure severity. In that sense, the SRE approach does not specifically support the quality improvement decision process. The software background of the SRE method makes it well-suitable for application in the complex consumer electronics context.

2.2.3 Conclusions

The purpose of this section was to demonstrate the general strengths and weaknesses of existing methods with respect to market uncertainty reduction in consumer electronics. Table 2.2 on the next page gives an overview of all the described methods and their potential contribution to market uncertainty reduction. Based on these results, the following conclusions can be drawn.

In general, modern product- and concept testing methods can be considered as being user-oriented. They impute important test contributions to current and future product users. Actual users are taking part in the different product/concept tests. However, the user-orientation is much more limited in the related quality methods. In these methods the actual users are generally not participating in the product evaluation and the opinion of the user is formulated by product engineers. Overall, most product testing and quality methods require extensive preparation- and execution time. This time can be reduced by the application of some related methods like HCCT or SRE. However, on the average, rapidness is not considered a hampering factor for the application of these methods in the consumer electronics industry.

The methods have a different focus with respect to their contribution to quality improvement. Several methods focus on non-technical quality problems (e.g. HCCT and usability testing), other concentrate on technical quality problems (e.g. FMEA and SRE) and some mainly consider utility problems (e.g. concept testing and QFD). The increasing complexity of the consumer electronics products and their environment is hindering the successful application of some methods within this industry (see Table 2.2). The increasing complexity results in less reliable quality predictions and longer test times of these methods. However, the adoption of some (parts of the) software improvement tools (e.g. SRE) could support the application possibilities of these methods for complex products in complex environments.

A general drawback of almost all methods is the lacking impact prediction of quality problems and the related improvement decisions. The application of these methods results in the identification of various quality problems (technical and non-technical) and therefore in a similar number of quality improvement possibilities. But the impact of these problems on user dissatisfaction is undetermined. Moreover, the methods that do predict the impact of the identified quality problems are not using a user-oriented approach. In these methods, the impact prediction is based on the general appraisal of the involved (product) engineer(s).

Summarizing, it can be stated that the combination of lacking user-orientation and confined specificity is restricting the potential uncertainty reduction of current quality improvement methods. Particularly, the unknown impact of identified quality problems on user dissatisfaction is limiting the potential contribution of these methods to the quality improvement decision process. Obviously, this indistinctness of impact eventually results in “incorrect” design decisions and consequently in an increasing number of product complaints.
### Table 2.2: Overview of potential market uncertainty reduction methods

<table>
<thead>
<tr>
<th>Methods</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User-Orientation</strong></td>
<td><strong>Rapid</strong></td>
</tr>
<tr>
<td><strong>Usability testing</strong></td>
<td>[NIE93], [ZIR94], [WIC00], [HAN01], [THO02], [VAL07]</td>
</tr>
<tr>
<td><strong>Concept testing</strong></td>
<td>[HAL71], [TAU75], [MOO82], [PAG92], [DUC94]</td>
</tr>
<tr>
<td><strong>HCCT</strong></td>
<td>[BAS03], [BOE03]</td>
</tr>
<tr>
<td><strong>QFD</strong></td>
<td>[SUL86], [GRB92], [CHA02], [OUD06]</td>
</tr>
<tr>
<td><strong>FMEA</strong></td>
<td>[BLI00], [GOD00], [BOW01], [YAC02], [STA03], [OZA04]</td>
</tr>
<tr>
<td><strong>SRE</strong></td>
<td>[MUS93], [LYU96]</td>
</tr>
</tbody>
</table>
2.3 Exploring industrial best practices with respect to customer dissatisfaction

Another category of methods aiming to decrease market uncertainty in the product development process are information feedback methods. As mentioned earlier, the amount of academic literature concerning this group of methods is very limited. Consequently, the potential contribution of quality feedback information to market uncertainty reduction is difficult to determine from this literature. In order to evaluate the contribution of these methods, three case studies were performed at feedback information providers, namely consumer electronics service centres.

Industry has two direct interfaces with customers: marketing/sales and service. Marketing/sales are mainly success oriented but service is concerned with product performance and customer dissatisfaction. Therefore service may provide useful field feedback information. As mentioned in the previous section, historically product defects implied specification violations and the overall customer dissatisfaction level could be expressed by the required number of product repairs in the after-market. In this perspective, service centre information provided useful feedback for quality improvement decision making. However, the increasing complexity of current software-based consumer electronics products and the increasing customer expectations result in a design oriented development process in which the main focus is on the product user. Consequently, current reported product defects do not only represent violations of the product specifications but also uncovered customer requirements and unexpected product behavior. Consequently, the number of actual product repairs in the after-market (to deal with specification violations) is no longer a representative measure of the overall customer dissatisfaction level. The case studies presented in this section investigate whether the service centre feedback processes have been upgraded to the information requirements of the current product development processes. In other words, this section investigates industrials best practices with respect to customer dissatisfaction information collection.

In this third section, the evaluation results of these case studies are presented. The section starts with an overview of the (limited) available literature on information feedback methods followed by a short introduction of the case studies and a description of the case study approach. Subsequently, the service processes and feedback processes of these companies are presented. Lastly, some conclusions are drawn about the potential contribution of quality feedback information to market uncertainty reduction.

2.3.1 Feedback Methods in Literature

In consumer electronics, most business processes mainly rely on feedback information from service centres to collect information about field problems [OUD06]. The main job of a service centre is to repair failed consumer electronics products.

Earlier research into these information feedback processes by Petkova and Den Ouden has resulted in the following conclusions [PET99], [PET03], [PET05], [OUD06]:

---

* The findings presented in this section are reproduced from "Understanding failure severity in new product development processes of consumer electronics products", by I.M. de Visser, Y. Lu and G. Nagappan in the proceedings of the 2006 IEEE International Conference on Management of Innovation Technology. For more details on these case studies, please refer to Appendix A.
The collected field information in the service centres is incomplete and not root-cause related.

Information about customer experience is usually not collected. Moreover, the collected service data only covers information on technical product problems.

The collected information becomes available too late to be useful for product quality improvement.

Products where no failure can be identified end up in the category “Failure Not Found”, and no further analysis is performed to discover the consumer's motivation for complaining about the product.

In addition to these results, the case studies in this section explicitly investigate the potential contribution of current information feedback methods to market uncertainty reduction. To what extent can service centres provide information about customer dissatisfaction with respect to consumer electronics products? Again, this evaluation is based on the earlier formulated requirements for these methods; user-oriented, specific, all-round, rapid and complexity proof.

Furthermore, the above presented conclusions with respect to information feedback methods are based on research performed at least four years ago. The case studies presented in this section investigate whether, nowadays, the (improved) service and feedback processes are more useful for product quality improvement in general.

2.3.2 Introduction to the Case Studies

The case studies described in this section are all performed in the field of high volume consumer electronics. In order to get a complete overview of current feedback processes in this field, three third party European service centres of a large consumer electronics company (company X*) were visited. These service centres are located in the Netherlands, Germany and France, and are responsible for most of company X's repairs. However, these three service centres work independently of each other. Therefore, the investigation of the service process at the service centres gives better insights into the feedback information processes in these countries.

The service process is one of the many business/consumer processes that connects the consumer to consumer electronics companies. After market introduction, consumers buy company X's products and start to use them. When the products do not meet their requirements or expectations, products are sent to the service centres via a number of different channels.

After repair or other alternative activities, products are returned to the customers and quality information with various levels of detail on the action performed is sent back to company X. In these case studies, two processes are reviewed:

- **The service process**: This is the product repair process of the service centres. A description of this process illustrates the general repair approach of the service centres which makes it easier to understand the different information feedback processes.

---

* For confidentiality reasons, the company name is fictitious
The information feedback processes: This is a description of the collected information in the service process and its transfer from the service centres to company X. On the one hand, it illustrates how the collected information by the service centres is actually obtained/used by company X. On the other hand, this description also indicates the potential contribution of these processes to market uncertainty reduction in consumer electronics. In other words, what more could company X do with collected service information.

2.3.3 Case Study Approach
The case studies consisted of the following activities:

- **Case study preparation**
  - Selection of carrier product. This carrier product is followed during the service process in the different service centres. The selection of one carrier product makes it easier to trace the different service activities and to compare the different service centres.
  - Definition of relevant key persons, operational and managerial, who represent the different activities within the service process.
  - Acquisition of commitment of the involved service centres. This commitment was acquired with the help of company X.

- **Interview and cross check**
  - Operational process mapping by interviewing people involved in the service process. This operational process mapping consisted of a structural analysis of the existing processes via the Maturity Index on Reliability (MIR) approach. For more details on this approach, please refer to [SAN99]. For analyzing the service process, this analysis focuses on the information structure behind the field returns.
  - Operational process verifying by cross checking with on site observations in the service process.

- **Report generation and feedback**
  - Description of the case study results.
  - Presentation of the results to company X.

Although the three service centres worked independently of each other, the service processes of the service centres appeared to be quite similar. In order to prevent repetition, one general description of the service processes is given. When these processes diverge at certain points, these distinctions are mentioned explicitly.

2.3.4 The Service Process
An overview of the general service process is given in Figure 2.8. The service process consists of five phases, namely:

- Reception, unpacking and identification
- Diagnosis, repair and correction
- Quality check
- Packaging
Dispatching

The German service centre in some cases applied, in addition, the so-called “swap procedure”. Subsequently, an explanation of the different process phases is given.

Reception, unpacking and identification

As part of the service agreement between company X and the service centres, the three service centres collect field returns of company X's products from various dealers and distributors. The service centres can also receive direct returns from end users. After reception, all returned products are unpacked and visually inspected. The identification is based on product type and serial number. At the request of company X, for all products basic customer information (e.g. name, address, date of purchase) is logged in a computer program. Subsequently, the different product types are sorted out and forwarded to the involved repair station.

Diagnosis, repair and correction

In Germany and the Netherlands, a restricted number of product types are assigned to the same people. As a result of that, service people can be considered as experts with regard to these assigned product types. For several product failures, the repair engineers have executed complete root-cause analyses and designed repair processes or work-around solutions. Next to that, the repair engineers have full access to customer information. In the case of a product entering the service process for the second time within three months without a clear problem, the repair engineer will contact the customer to ask for extra information about the failure behavior of the device. Some failures are solved by giving some extra information on usage to the customers. This information is provided by the service centres themselves. Sometimes instructions are written down and in other cases pieces of the user manual are copied. For other failures, the repair engineer will start a root-cause analysis based on the customer information of the failure, in order to solve it.

In France, two levels of repair processes are defined by the service centre. The first level of repair is concerned with common failures in the product. These repairs can be executed by non-technical repair personnel according to a standard repair protocol. The second level of repair is similar to the standard diagnosis, repair and correction process of the Dutch and German service centres. In this second level of repairs, all less common failures are dealt with by technically educated repair personnel. Again, these repair people can be considered as experts with regard to the involved product types. However, the French service centre does not contact the customers in case of repeating repairs.
Quality Check
In the Quality Check phase, a number of tests are performed. The test process is defined by the service centres with the requirements of company X. About fifty percent of the products that have passed the repair process will be tested at the quality check. Products that do not pass the quality check test are sent back to the repair process. Products that initially did not pass the quality check are always tested a second time after the second repair process.

Packaging
All repaired products are packed properly in this phase. All necessary documents will be attached.

Dispatching
All repaired products are sent to the customers directly or via distributors and dealers.

Swap Procedure
In Germany, some product types are immediately replaced by other products. For these products no repair takes place. The customers of these products always receive an alternative product instead of their own product. This process is called the swap procedure. This swap procedure is also used for some immature products for which no root-cause analysis has been performed to identify the main causes of failure. Important information about failure mechanisms in these products is discarded together with the products. In future product generations, these failures will not be prevented due to this loss of information about the root-causes of these failures.

2.3.5 Feedback Process
In order to describe the feedback process from the service centres to company X, a distinction is made between standard repair information, introduction repair information and introduction factory information. This subsection discusses the feedback processes for these different types of information.

Introduction Repair and Factory Process
For each new product introduction, company X defines a number of first returned sets that need to be analyzed in detail and repaired in specialized service centres. All three service centres described in this research are indicated as specialized services centres for new product introduction. In these specialized service centres, the technical department repairs products with the direct support from company X. Subsequently, detailed root-cause information is sent back to company X. This information collection process is called Introduction Repair Process (IRP).

For some of these newly introduced products, company X requires that a certain number of these products are to be swapped and directly sent back to company X for further analysis. This process is called Introduction Factory Process (IFP). Company X decides for every new product type whether the sets are to be swapped or repaired. The IRP and IFP approaches are part of a new procedure within company X to improve product quality.

In other words, based on the agreement between the service centres and Company X with regard to the IRP approach, the service centres send detailed root-cause information to company X. This root-cause information is given for a number of field returned products that is defined by company X. In case of the IFP, the returned products are sent back to company X for further analysis. This is done for sets that returned immediately after market introduction.
Standard repair information

For other repairs, the repair centres send repair job sheets back to company X via the central service organization of company X. From these job sheets, company X can learn what failure symptoms the returned products have and what repair actions were taken. Based on this information, company X calculates the amount of service costs that should be paid to the service centre. As a result of that, the content of this information source is logistically oriented. This logistical orientation of the service information makes it hardly suitable for quality analysis and reporting. Moreover, company X requires from the service centres that they use so-called IRIS codes to describe the failures. The IRIS code is a standard failure description code for consumer electronics. However, the IRIS code does not specify the failure symptoms specific enough. As a result, the service centres internally use their own failure description to classify the failure symptoms more specifically. Nevertheless, this information is not sent back to company X because it is not requested for.

One important observation is that the service centres often classify the failure symptoms on the basis of the repair actions that were taken. This is quite understandable considering the logistical orientation of this feedback information. However, the repair actions aim to remove the failure symptoms instead of solving the failures from their root-cause level. That means that the failure classification does not provide any information about the root-causes of the different failures.

This formal feedback process is often supplemented with ongoing or ad hoc communication. Ongoing communication is mostly the result of the collaboration between company X and the service centres. People from all kinds of departments and business units within company X visit the service centres on a regular basis. During these visits, information about the service process and the products is exchanged from the service centres to company X. During the outbreak of epidemic failures, ad hoc information transfer takes place between the service centre and company X. However, the service centres do not have fixed communication channels with company X. Therefore, most ad-hoc communication is the result of company X asking for information.

Historically, the No Fault Found (NFF) process was considered a special case of the repair process. This describes the situation in which no clear failure can be identified during the repair process in the service centre. In the past, these NFF's were mostly “hard to identify” failures that were solved by swapping the product instead of repairing it. This approach makes sense from an efficiency perspective, since identification of some failures may take a long time and through that the repair may become too costly. However, these days a NFF in the repair process is mostly caused by one of the following causes:

- The failure identified by the customer is an intermittent failure. This means that the failure is not always visible to the user but only under certain conditions. In the repair process, this failure could not be reproduced.
- The failure identified by the customer is caused by an interaction between the product and its environment (e.g. with other consumer electronics products). Since the service centres only receive the product from the customer, the failure can not be reproduced at the service centre.
- The failure identified by the customer is the result of customer's incomplete understanding of the functionality of the product. Since these failures are customer dependent, the failure can not be reproduced at the service centre.
The failure identified by the customer is the result of customer's dissatisfaction with the functionality of the product. In other words, the failure is a non-technical quality problem. Since these failures are also customer dependent, the failure can not be reproduced at the service centre.

When a product enters the service centre for the first time and the repair engineers can not find a problem, the device software will be updated. From a historical perspective, this approach makes sense. By updating the software, the service centre hopes to solve the deep-routed failure, intermittent failure or the interaction failure that was possibly present in the product. However, the third and fourth category of failures can not be solved by this software update. After this software update, the product is packaged and dispatched to the customer. All these product failures are reported as being software failures to company X.

The service centres receive a fixed amount of money per repair for every product type. When a product enters the service process for the second time within three months, company X will only pay material costs for this second repair. The procedure for the second NFF-service for the first three months differs for the same product between the three service centres. In France, the procedure is exactly the same as for the first NFF service. After a new software update the product is dispatched to the customer. Only at the third request for service for the same product, this service centre will ask company X to offer this customer a so-called “commercial solution”. This means that company X will offer this customer another product or (part of) their money back.

In the Netherlands and Germany, the procedure for the second NFF service is different. Both service centres will do another software update, but also provide extra information on usage to the customer. As mentioned before, sometimes instructions are written down and in other cases pieces of the user manual are copied. By providing this information to the customers, the service centres try to solve the third category of failures that is caused by incomplete customer understanding of the product.

However, the fourth category of non-technical failures, caused by customer's dissatisfaction with the functionality of the product is not given any attention and consequently, these problems are not solved by the service process.

2.3.6 Conclusions

Earlier research has already demonstrated some limitations in service feedback information from service centres [PET99], [PET03], [OUD06]. However, the main purpose of these case studies was to evaluate the potential contribution of this feedback information method to market uncertainty reduction in consumer electronics. To what extent can service centres provide information about customer dissatisfaction with respect to consumer electronics products? This evaluation is based on the earlier formulated requirements for these methods, namely user-oriented, specific, all-round, rapid and complexity-proof.

The complete service process is designed for managing technical quality problems. The only non-technical quality problems that are dealt with are the issues with so-called ignorant customers. Even these actions are initiated by the service centres and not by company X. Company X does not pay these service centres for the activities. As a result, the customers depend on the goodwill of the service centre whether they will provide this extra information.
Consequently, the collected information with respect to non-technical failures is very limited. Furthermore, the transfer of this information to company X is unusual. This service information does not support design decisions with respect to non-technical product quality.

As a result of the fixed allowance for every repair, the main performance characteristic of the service centre is service time. This is the time a product is present in the service process. The performance characteristic is the result of the cost-driven design of the complete service organization within company X. Within this service process, no distinction is made between commonality products and high-end products. For commonality products, which are mainly confronted with product technology uncertainty, this cost-driven organization process is very suitable since the main causes of failure and the repair process of these products are well known and controlled. For new high-tech products, this cost-driven service organization results in the outbreak of epidemic failures for which substantial root-cause analyses should be performed. The competence level within the service centres is certainly high enough to be able to perform these root-cause analyses, but the cost-driven instructions from company X often prevent this from taking place. Increasing product complexity and increasing uncertainty in product use result in a growing number of unknown failure symptoms. Because of the cost-driven nature within the service processes, it is unlikely that root-cause information related to those unknown failure symptoms can be obtained using these processes. In other words, the information collected in current product service processes is far from specific enough to potentially contribute to uncertainty reduction in the product design process.

The user-orientation in the service process is only about informing the customer on product use. Customer's incomplete understanding of product functionality results in reoccurring customer complaints at the service centres. In order to solve this problem, some service centres offer a supplemental product explanation. Nonetheless, the actual user-oriented failure description is discarded by the adoption of the IRIS codes.

Earlier research has demonstrated the absolute lateness of the service information feedback process [PET03]. In these case studies, the timeliness analysis of the feedback process was not repeated. However, the complex organizational structure of a third party service centre combined with a large consumer electronics company most probably results in an unresponsive feedback process.

Increasing product complexity requires more elaborate root-cause analyses for product quality improvement. However, in cost-oriented service processes the time to perform these root-cause analyses is lacking. Generally, the root-cause analysis process is substituted by a more elaborate component replacement procedure. Complete product components are swapped in order to solve the increasing number of customer complaints, but this procedure does not assure improved product quality.

In summary, these case studies confirm the results of earlier research. The earlier identified limitations of feedback information processes have been diagnosed. However, within the service centres some progress has been made with respect to the management of non technical product failures. All service centres recognize the significance of this group of failures and some service centres even offer additional information to customers to solve these problems. Despite this progress, all the actions related to non technical failures are based on the initiative of the service centres and still no information about non technical failures is transferred to the involved consumer electronics company.
Furthermore, the case studies revealed some important constraints of the current feedback processes for uncertainty reduction in the design process of consumer electronics. Particularly, the logistical orientation of the service process hinders the required information collection for improved design decision making. The service information does not provide additional insight into customer dissatisfaction levels with respect to consumer electronics products. Consequently, the combination of all the above mentioned limitations makes the service feedback process unsuitable for uncertainty reduction in the design process of consumer electronics. For that reason, in this thesis the service feedback process is no longer considered an option for uncertainty reduction in the consumer electronics design process.

2.4 Problem definition and research question

The earlier literature review shows the potential of current quality improvement methods for user-oriented data collection. However, the specificity of the collected data is for almost all methods far from satisfactory. Furthermore, in several quality improvement methods the user-orientation is limited or even completely lacking. Particularly, the unknown impact of identified quality problems on user dissatisfaction is limiting the potential contribution of these methods to the quality improvement decision process. This indistinctness of impact eventually results in “incorrect” design decisions and consequently in an increasing number of product complaints.

In consumer electronics industry, all the quality improvement methods are applied to identify immense numbers of user-oriented quality problems. However, as a result of the limited specificity and user-orientation of these methods, only a limited user-oriented severity assessment of these problems takes place. The direct consequence of this situation is an increasing market uncertainty for consumer electronics companies.

In other words, the earlier mentioned quality methods provide many suggestions for product quality improvement in the development process. They cover the areas of technical as well as non-technical quality problems in an often user-oriented way. However, most methods lack an impact evaluation of these identified quality problems on customer dissatisfaction. In the methods that do perform an impact evaluation of the identified problems, the user-perspective is lacking. Consequently, all the identified quality problems are considered of equal impact on customer dissatisfaction.

This situation would not lead to incorrect design decisions in case of simple products and indefinite product development time where customer requirements could be fully understood; in that case all identified quality problems could be solved anyhow. However, in the consumer electronics industry with its increasing complex products and strong time-to-market pressure, this situation results in unfounded decisions for certain quality improvement actions. Which quality problem should be solved first is based on the non-user-oriented assessment of the design team. Thus, despite the contribution of existing quality improvement methods to the identification of product improvement possibilities, no method specifically supports the product improvement decision process by determining the impact of a certain product improvement decision on the customer dissatisfaction level. It is not difficult to recognize that the absence of this impact assessment in the design process irrevocably results in incorrect design decisions. Unimportant (from a user perspective) product quality problems are settled before market introduction and important problems are still present in the final product. Apparently, these incorrect design decisions negatively influence the number of customer complaints, see Figure 2.4.
At this point, it is important to mention that this thesis particularly considers the impact of design decisions on user dissatisfaction. As indicated in Chapter 1, the goal of this thesis is to contribute to the overall quality improvement of consumer electronics products. And also in Chapter 1, quality is defined as being able to design and deliver a product to the customer that is conforming to his/her expectations. Products that do not conform to the expectations of the user result in dissatisfaction of this user and conceivably in user complaints. Therefore dissatisfaction is considered a mediating variable in the relationship between quality problems and user complaints. This relationship is presented in Figure 2.9.

![Figure 2.9: Relationship between quality problems and user complaints](image)

The results of the service centre case studies confirm the absence of proper user complaints in service feedback information. Moreover, the presence of user dissatisfaction as mediating variable in the relationship between quality problems and user complaints reinforces the need to focus on user dissatisfaction instead of user complaints. Moreover, the prevention of user complaints is more concerned with decreasing user dissatisfaction than with increasing user satisfaction. More details about this difference between user satisfaction and dissatisfaction are given in Chapter 3.

Considering the results of the previous sections, in order to deal with the increased market uncertainty in consumer electronics, existing quality methods from literature and practice should be complemented with a user-oriented impact assessment of the identified quality problems (and the accompanying quality improvement decisions). The combination of the existing quality methods with an accompanying user-oriented impact assessment would substantially lower the market uncertainty in the design process. Eventually, this approach contributes to a decreasing number of customer complaints in consumer electronics.

This analysis combined with the general aim of this research, results in the formulation of the following research question for this dissertation:

**How to predict the impact of (potential) product quality problems on customer dissatisfaction in consumer electronics industry?**

Like most research questions, this question can only be answered by dividing it into smaller sub-questions. The first sub-question is concerned with the formulation of a concept that encompasses the user-oriented impact assessment. In the context of consumer electronics, an existing or new concept should be defined that expresses the impact of a quality problem on customer dissatisfaction.

* The broad term “dissatisfaction” is deliberately selected for this research question. In consumer behavior literature, customer dissatisfaction is considered a predecessor of customer complaint behavior. In order to limit the number of customer complaints, customer dissatisfaction should be controlled. More details about this customer dissatisfaction concept and the relation between customer dissatisfaction and customer complaint behavior are presented in Chapter 3.
This is summarized in the first sub-question:

**Sub-question 1:** How to express and measure the impact of quality problems on customer dissatisfaction in consumer electronics industry?

In order to predict the impact of potential quality problems, the underlying factors that influence this customer dissatisfaction concept should be investigated and determined. This is summarized in the second sub-question:

**Sub-question 2:** What are the underlying factors of the quality problems that influence this customer dissatisfaction concept?

Lastly, the (degree of) influence of these different factors on this customer dissatisfaction concept should be examined. This is expressed in the third sub-question:

**Sub-question 3:** How do (some of†) these factors influence this customer dissatisfaction concept in consumer electronics industry?

Eventually, the answers to these sub-questions should result in a (partial) model that predicts the influence of a certain quality problem on customer dissatisfaction. Obviously, this (partial) model is the answer to the main research question of this research.

### 2.5 Research Approach

As indicated earlier, theory within the field of quality improvement is well-established. However, in order to answer the earlier mentioned research questions integration between this field of research and other fields of research is necessary. Furthermore, new constructs need to be build that supplement these existing mature theories. Because of that, this research belongs to the category of intermediate theory research; research that employs a partially focused lens to study a particular issue or set of variables in which quantitative as well as qualitative methods can be adopted to answer the research question [EDM04].

For the collection of qualitative and quantitative information, different research strategies can be adopted. Each is a different way of collecting and analyzing empirical evidence following its own logic and has its own advantages and disadvantages [YIN94]. Yin states that three conditions determine the selection of a research strategy, namely:

- The type of research question that is posed
- The extent of control an investigator has over actual behavioral events
- The degree of focus on contemporary events

Table 2.3 demonstrates these conditions and indicates how each is related to five major research strategies: experiments, surveys, archival analysis, histories and case studies.

The research question to be answered in this thesis can be considered *explanatory*. The first research phase is concerned with the development of a concept to express and measure the impact of quality problems on customer dissatisfaction.

---

† Given the fixed time-span of this research, depending on the number of potential influencing factors, probably not for all these factors the influence on customer dissatisfaction will be examined.
This concept is developed based on literature from different academic fields such as marketing, consumer behavior and psychology. Subsequently, the underlying factors that potentially influence this customer dissatisfaction concept are identified based on academic literature and business information.

Lastly, the relevance of these (potential) influencing factors is investigated by answering questions like “how do these factors influence this customer dissatisfaction concept?” and “how do these aspects relate to each other?”.

Table 2.3: Relevant situations for different research strategies [YIN94]

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Form of research question</th>
<th>Requires control over behavioral event?</th>
<th>Focuses on contemporary events?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>how, why</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Survey</td>
<td>who, what, where, how many</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Archival analysis</td>
<td>who, what, where, how many</td>
<td>no</td>
<td>yes/no</td>
</tr>
<tr>
<td>History</td>
<td>how, why</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Case study</td>
<td>how, why</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Complying with the explanatory nature of these questions, three research approaches are considered applicable in this research phase: case studies, histories and experiments. However, since this research focuses on contemporary events, the history research approach should be eliminated from this set of three. The case study approach is preferred in examining contemporary events when relevant behavior cannot be manipulated. On the other hand, experiments can be done when an investigator can manipulate behavior directly and systematically. In this research, the relevant behavior (being the user dissatisfaction caused by quality problems) can be directly and systematically manipulated by varying the values of the different underlying factors. This possibility enables the selection of the experimental research approach for the impact prediction of quality problems on customer dissatisfaction.

Additionally, the case study research approach requires multiple sources of evidence, with data needing to converge in a triangulation fashion [YIN94]. The explorative case studies in this chapter demonstrated that the possibilities of using current reliability feedback information for this impact prediction are very restricted. Since in the field of consumer electronics products, no other sources of case study evidence than these after market processes are available, the experimental research approach is selected for answering the main research question of this thesis.

Darius and Portier define an experimental study or an experiment as “any research study in which the researcher deliberately changes or assigns values (levels) to one factor in order to determine its effect by comparing the groups created by this factor” [ADE99]. According to Van Vliet, the experimental method is used to find out whether a factor influences behavior in a certain way and in that case to discover the level of influence [VLIE95]. However, experimentation is very diverse [GAL03]. On the one hand, there is the field of chemical and biological experimentation which can be characterized as quantitative laboratory research. On the other hand, there is the field of social and behavioral experimentation with generally more qualitative aspects and less constant experimental conditions. The experimental work executed in this research is within the field of social and behavioral experimentation.
In order to give an overview of the approach within this research project, a research model is presented in Figure 2.10. A research model is the schematic representation of the research goal together with the global steps required to achieve this goal [VER04]. The first step of this research is the formulation of the research questions based on academic literature and the explorative case studies. This step eventually resulted in the research questions as presented in Section 2.4. Subsequently, in Section 3.3, the first sub-question is answered based on literature from different academic fields about user (dis)satisfaction, emotions, affect and complaint behavior. This step results in the formulation of a new concept to measure the impact of quality problems on user dissatisfaction. Next, in Section 4.1, a hypothetical model is formulated to predict this impact based on different influencing factors. These hypothetical influencing factors are identified based on literature and consumer electronics business information. Subsequently, parts of this model are tested in Section 4.2 resulting in an improved version of the hypothetical prediction model in Chapter 5. Chapter 6 identifies and validates measurement approaches for the different variables of the experimental model. Subsequently, Chapter 7 and 8 are concerned with the validation of this improved impact prediction model. Eventually, in Section 8.4, this results in a partially validated model to predict the impact of quality problems on user dissatisfaction. This model is the answer to the main research question of this thesis. Lastly, in Chapter 9 overall conclusions and recommendations for further research are presented.
Figure 2.10: Research model based on [VER04] (numbers in brackets indicate related chapters and sections of this thesis)
3  User Perceived Failure Severity (UPFS)

The previous chapter discussed some important implications of the current trends in consumer electronics industry. Literature review illustrated the inability of existing quality improvement methods in dealing with these business trends. Particularly, the unknown impact of identified quality problems on user dissatisfaction is limiting the potential contribution of these methods to the quality improvement decision process in consumer electronics. This indistinctness of impact eventually results in “incorrect” design decisions and consequently in an increasing number of product complaints. In order to reduce this uncertainty in the product development process, companies need a validated prediction model of the impact of product quality problems on customer dissatisfaction in consumer electronics.

The first step in the development of this impact prediction model is the definition of a measure for user dissatisfaction that expresses the impact of quality problems in consumer electronics. A vast amount of literature is available on consumer (dis)satisfaction, affect, emotions and complaint behavior. This literature is presented in the first two sections of this chapter. In Section 3.3, related concepts from literature are combined into a new theoretical model for this thesis. Lastly, in Section 3.4, a new measure for user dissatisfaction that expresses the impact of quality problems in consumer electronics is defined: User Perceived Failure Severity (UPFS).

3.1  Consumer (dis)satisfaction and complaints

Many different definitions of consumer satisfaction can be found in literature. But the conceptualization that has received the most support in literature is the view that satisfaction is a post-choice evaluative judgment concerning a specific purchase selection [WES91]. Consistent with theoretical and empirical evidence, Oliver defines satisfaction as follows:

“Satisfaction is the consumer’s fulfillment response. It is a judgment that a product or service feature, or the product or service itself, provided (or is providing) a pleasurable level of consumption-related fulfillment, including levels of under- or overfulfillment” [OLI96].

Oliver states that pleasurable implies that fulfillment gives or increases pleasure, or reduces pain. So users can be satisfied by going back to normalcy or neutrality, as in the removal of an aversive state [OLI96]. The notion of fulfillment in this definition implies that a certain goal exists. In this context, a goal is something to be achieved. Thus, fulfillment can only be judged with reference to a standard. This standard forms the basis for comparison. Oliver stresses that a fulfillment, and hence a satisfaction judgment, involves at the minimum two stimuli; an outcome and a comparison referent [OLI96], [FOU99]. In this context, dissatisfaction is simply defined by substituting the word unpleasant for pleasurable in this above definition. So, the displeasure of underfulfillment is typically dissatisfying [OLI96].

Different authors confirm the assumed relationship between particularly user dissatisfaction and user complaining behavior (see Figure 2.9). Generally, user dissatisfaction is considered to be a precursor of a complaint; a higher level of dissatisfaction increases the complaint
readiness of the user [GIL82], [BEA83], [SIN91], [OLI96]. However, only few instances of dissatisfaction result in formal complaints [MUL98A]. In other words, user dissatisfaction is a strict prerequisite for complaining but other moderating variables (economical and behavioral) also influence the eventual filing of a formal complaint by the user [SIN91], [OLI96], [MUL98A]. Thus in order to reduce the number of user complaints in consumer electronics, user dissatisfaction is a good starting point for measuring the impact of (potential) quality problems. This positive relationship between dissatisfaction and complain readiness of the user is illustrated in Figure 3.1.

![Figure 3.1: Relationship between user dissatisfaction and user's complain readiness](image)

As mentioned above, dissatisfaction is concerned with product consumption related underfulfillment. This underfulfillment is caused by a difference in expected performance and actual performance of the involved product. Based on the performance of a product, the consumer may experience certain emotions; specific human affects resulting from e.g., blame and gratitude [OLI96]. Examples of emotions resulting from product use are [OLI96]:

- anger directed at the manufacturer for producing a defective product
- guilt, a feeling of self-blame for making a bad decision
- delight over the choice of an exceptionally good product

This thesis is mainly concerned with the prediction of user dissatisfaction caused by quality problems in consumer electronics products to reduce the number of product complaints. Westbrook indicates that emotions affect consumer satisfaction [WES91]. In this context, it is important to mention that especially the “irritation” emotion has a negative impact on customer satisfaction [MAN91], [DOL01]. Although other situational and behavioral factors influence user dissatisfaction as well, user irritation can be considered as major influencer of user dissatisfaction in consumer electronics. Moreover, the irritation emotion is concerned with “blaming others” and not with “blaming yourself”. As a consequence, this external attribution property embedded in the irritation emotion also has a direct positive impact on user's comply readiness in consumer electronics [FRI86], [BAG99]. These relationships are presented in Figure 3.2 on the next page.

Considering the direct influence of irritation on user satisfaction and user's comply readiness, the irritation emotion is selected as a starting point for measuring the impact of quality problems in consumer electronics products.

* Please note that a complete description of the user complaining process is beyond the scope of this thesis. For more details about this process please refer to the references that are mentioned in this section.
The next section gives an overview of the relevant literature with respect to emotions in general and the irritation emotion in particular. This eventually results in a definition of a measure for the user impact of quality problems in consumer electronics. This measure is presented in the last section of this chapter.

![Diagram of User Irritation, Dissatisfaction, and Complain Readiness]

**Figure 3.2:** Relationship between user irritation, dissatisfaction and complain readiness [MAN91]

### 3.2 Emotion, personality and irritation

In order to express the impact of quality problems on customer dissatisfaction/irritation, literature from several academic research areas can be contributive. Therefore, this subsection gives an overview of the relevant theories with respect to irritation in a use context. At this point, it is also important to stress the difference between behavior-, personality- and emotion research.

**Consumer behavior**

Consumer behavior is the study of how and why people consume products and services [WEB05]. To predict the impact of product quality problems on user dissatisfaction, consumer behavior research models can be helpful. According to Chaudhuri, all behavior can be broadly attributed to three classic influences [CHA06]:

- the particular characteristics of the individual
- the environment that surrounds the individual
- the inherited genetics that constitute the biological makeup of the individual

The characteristics of an individual consist of the personality, perceptions, attitudes, needs, and motivations of the individual. The environment consists of the culture, family, friends and institutions that the individual lives in. The genetics of an individual are the biological codes (DNA) that are unique for each individual [CHA06]. The consumer behavior model in Figure 3.3 illustrates that stimuli result in emotional and rational responses in the individual's mind, which may lead to a particular response by the individual.

There is a strong link between this model and the earlier presented models in Section 3.1. This thesis focuses on the prediction of dissatisfaction/irritation (emotional response) caused by product quality problems (stimulus) and resulting in product complaints (behavior). According to this behavioral model, the future behavior prediction model should also consider the influence of individual characteristics, environmental factors and genetics.
The influence of individual differences on (use) behavior is one of the main issues of personality research.

**Figure 3.3:** Consumer behavior model [CHA06]

### Personality and emotion research

Personality research brings together contributions from developmental, social, cognitive and biological psychology. Personality research studies the whole person, the dynamics of human affect and motivation and the identification and empirical measurement of the individual differences among persons [MCA97]. As illustrated in Figure 3.3, emotions and other affective phenomena are important for human behavior [WRA05]. Affect plays a central role in personality theories. Affect refers to the emotional side of consciousness. Davitz made a division for person characteristics related to affect: first, stable traits that describe how people usually or typically are; second, transitory states that describe the affective condition of people during a specific moment in time [DAV69]. Emotion researchers are primarily interested in the affective states, whereas personality researchers are primarily interested in affective traits. However, the boundaries between affect-related traits and affect states are vague [END76], [CHA88]. Both states are part of an emotional response (see Figure 3.3).

In a classical book on personality theory, Allport defines traits as “the building blocks of personality, the guidepost for action, the source of individual's uniqueness” [ALL37]. Traits are inferred predispositions that direct the behavior of an individual in consistent and characteristic ways. Traits produce consistencies in behavior because they are enduring attributes, and they are general or broad in scope [GIB03]. “The fundamental goal of trait theory is to characterize individuals in terms of comprehensive but finite, preferably small sets of stable dispositions that remain invariant across situations and that are distinctive for the individual, determining a wide range of behaviors” [MIS98].

Emotion research is concerned with the transitory states that describe the affective condition of people during a specific moment in time. Therefore, predicting the user impact of product quality problems involves predicting the emotional response (transitory state) to these problems given certain predetermined user characteristics (traits). In academic literature, there is no consensus concerning the definition or exact nature of emotions [WRA05]. However, the cognitive appraisal theories have been among the most influential within this field [HAM88], [LAZ91], [WRA05]. Cognitive appraisal theories of emotion state that the elicitation and differentiation of emotions is based upon a process of cognitive evaluations of
appraisals [SCH84], [FRI86], [HAM88], [LAZ91]. In these theories, it is not an event that
determines an emotional response, but the evaluation and interpretations the individual makes
of the event [WRA05]. In everyday language, the words affect and emotion are often used as
synonyms. However, in the cognitive appraisal theories, a distinction is made between these
two concepts. Affect refers to the feeling side of consciousness. This feeling side includes
pleasure and displeasure, happiness and sadness, liking and disliking, and the psychological
and visceral sensations brought on by the neural-hormonal bodily systems [DAV69],
[OLI97], [RUS99]. Emotion includes arousal, various forms of affect, and cognitive
interpretations of affect. Emotion is more cognitively involved than affect [DAV69],
[OLI97].

Several researchers have demonstrated strong links between specific cognitive evaluations
(so-called appraisal patterns) and emotions [FRI86], [HAM88], [WRA05]. For example, fear
is associated with evaluating a situation as threatening, guilt with self-blame and
anger/irritation with blaming someone else for an undesirable situation. This link explains the
earlier presented relationship between irritation and complain readiness. Scherer defines an
emotion as “the interface between an organism and its environment mediating between
contantly changing situations and events and the individual's behavior responses” [SCH84].
He explains that an emotion state is the result of a process of cognitive evaluation, which can
be characterized as a rapidly and hierarchically structured sequence of stimulus processing
steps.

Irritation

The last half decade, many researchers have contributed to the development of a descriptive
model of the human affective space. In 1985, Watson and Tellegen proposed their so-called
“consensual model of affect”. Although some of the earlier models claim the existence of
three major affect dimensions, affect researchers gradually inclined towards a two-factor
structure [WAT99]. The two major affect dimensions in this model are: Pleasantness versus
Unpleasantness and Activation versus Arousal [RUS79], [RUS80]. In the early 1980's,
Russell made a great step forward by proposing that these two dimensions define a
circumplex, that is, a model in which mood descriptors can be systematically arranged around
the perimeter of a circle [RUS80], [RUS82]. This affect circumplex consisted of four bipolar
dimensions that were spaced 45° apart:

- Pleasantness (pleasure versus misery)
- Excitement (excitement versus depression)
- Activation (arousal versus sleepiness)
- Distress (distress versus contentment)

Based on the results of many self-rating studies, Watson and Tellegen reconfirmed this basis
structure of the circumplex affect model. Moreover, they presented a circular structure that
was designed to resemble Russell’s circumplex as much as possible [WAT85]. This model is
presented in Figure 3.4. Like Russell's model, this structure consists of four bipolar
dimensions that are spaced 45° apart [WAT85]:

- Pleasantness (happy versus sad)
- Positive Affect (excited versus sluggish)
- Engagement (aroused versus still)
- Negative Affect (stressed versus relaxed)
The terms in adjacent octants are moderately positively correlated. Words 90° apart are in essence unrelated to another, whereas those 180° apart are opposite in meaning and negatively correlated. Different from Russell, Watson and Tellegen emphasized the importance of the Positive Affect (PA) and Negative Affect (NA) dimensions that are represented by solid lines in Figure 3.4 [WAT99]. Larsen and Diener, as well as Feldman Barrett and Russell, proposed a model very similar to this circumplex in the 1990's [LAR87], [LAR92], [FEL98]. This confirms the general agreement among researchers about the high level structure of emotions and affect.

The main components of this model are PA and NA. High PA is composed of terms reflecting enthusiasm, energy, mental alertness and determination. In contrast, NA is a general factor of subjective distress and subsumes a broad range of aversive mood states, including distressed, nervous, afraid, angry, guilty and scornful [WAT88]. This affect model can be used to express both mood factors as states as well as traits (emotional responses and individual characteristics, see Figure 3.3) [WAT88A].

The results of many affect related research studies confirm the appropriateness of this two dimensional structure of affect and emotion [DIE85], [DIE86], [WAT91], [WAT92], [BAR98], [WAT99], [CRO03], [HIL05]. However, the independence of PA and NA is not
undisputed. Some findings support the older view that considers PA and NA as two opposite ends of a single bipolar continuum [BAR98], [RUS99], [TER03]. This reasoning implicates that a lower negative affect level corresponds to a higher positive affect level (and the other way around). Nevertheless, most recent studies confirm the (theoretical) independence of the PA and NA dimensions [GRE99], [TEL99], [REM00]. Although this finding is rather counterintuitive, most researchers currently agree that a higher level of positive affect does not necessarily implicate a lower negative affect level (or the other way around) [BAR98], [WAT92A]. The irritation emotion is part of the “anger family” emotions like anger, contempt, frustration and irritation [WRA05]. Irritation can be considered as a weaker form of anger. Therefore, in the circumplex model of affect, irritation is part of the NA dimension.

3.3 Overall theoretical model

Based on the theory presented in this chapter, an overall theoretical model for this thesis is presented in Figure 3.5. In Chapter 2, the results of the case studies confirm the lack of user complaint information in consumer electronics service centres. Moreover, the presence of user dissatisfaction as a mediating variable in the relationship between quality problems and user complaints endorses the decision to focus on user dissatisfaction instead of user complaints.

However, user dissatisfaction is a complex theoretical concept with various influencing situational and behavioral factors. The direct influence and conditionality of irritation on user dissatisfaction and user's complain readiness makes this emotion a good starting point in measuring the impact of quality problems in consumer electronics products in this research. Moreover, current insights into emotion theory confirm the appropriateness of the circumplex model of affect and emotion. Research performed with respect to this circumplex model provides a solid theoretical foundation for the application of irritation (being part of NA) as a response measurement dimension in this research. Furthermore, behavioral research indicates the need to control the environment and the individual characteristics of test participants when investigating the influence of the stimulus (quality problem) on the emotional response (irritation) in this research. Based on this related theoretical information, the remainder of this thesis focuses on the relationship between quality problems (independent variable) and user irritation (dependent variable) as displayed in the dotted square of Figure 3.5.

Figure 3.5: Overall theoretical model (within dotted square: focus of this research, investigating the relation between quality problems and user irritation)
3.4 User Perceived Failure Severity (UPFS)

Based on the insights presented in the previous sections, the following concept is defined that captures the user impact of quality problems in consumer electronics products:

User Perceived Failure Severity (UPFS) is the level of irritation experienced by the user caused by a product failure

The UPFS concept expresses the impact of product quality problems on customer dissatisfaction in consumer electronics. Insight into the expected/predicted UPFS resulting from a certain design decision would reduce uncertainty in the product development decision process and decrease the number of product complaints. UPFS is an emotion measurement and therefore depends on a certain stimulus (product failure) and individual- and environmental characteristics. Especially these last characteristics should not be ignored in the experimental validation of potential UPFS prediction models.

The next chapter describes the development of a first UPFS prediction model in consumer electronics (see sub-question 2 and 3, Section 2.4). This model is tested in an exploratory consumer experiment. Based on the results of this first experiment, the original UPFS prediction model is improved and extended.
4 An UPFS Model based on Best Practices from Industry

The previous chapter illustrated the significant relationship between quality problems (as independent variable) and user irritation (as dependent variable). The UPFS concept was defined to capture the impact of quality problems on user dissatisfaction. This chapter describes the development of a first UPFS prediction model for consumer electronics. Starting point for this prediction model is the approach that is currently applied in consumer electronics industry for user dissatisfaction prediction within the product development process. In case of fast business development it is not uncommon for practical approaches to be ahead of the theoretical knowledge base. Consequently, given the current trends in consumer electronics industry, practical user dissatisfaction approaches can be considered a good starting point for the development of theoretical models. Moreover, experimental testing of the practical models should reveal useful information about the relationship between quality problems and user irritation.

In the first section of this chapter, the approach that is currently applied in industry for user dissatisfaction prediction is presented resulting in a so-called expert-based UPFS prediction model. Subsequently, this hypothetical UPFS model is experimentally tested in a consumer experiment. Although the design of this first experiment is rather elementary, the experimental results are meaningful for the development of the UPFS prediction model. In the last section, based on the experimental results, some conclusions are drawn about the suitability of the expert-based UPFS prediction model to capture the impact of quality problems on user dissatisfaction in consumer electronics industry.

4.1 The expert-based UPFS prediction model

As indicated earlier, the first version of the UPFS prediction model is based on the user dissatisfaction prediction approach as applied in the development process of consumer electronics products. In order to develop this UPFS prediction model from industry, the user dissatisfaction prediction approach for quality problems in a large consumer electronics company was analyzed. This analysis was based on unstructured interviews with three product testers and two (software) product developers in this company.

These interviews resulted in the following conclusions with respect to the user dissatisfaction prediction of quality problems before market introduction in this company:

- A relatively simple user dissatisfaction prediction tool is used by the product testers/developers. This tool requires that testers/developers, sometimes independently but mostly in expert groups, classify the failures according to their estimation of failure impact on customer satisfaction. The different failure classes are indicated with the letters S, and A till D. S indicates a Safety issue which should be solved immediately. An A product failure is a major issue that should be solved before product market introduction and D being a minor issue that could be solved in the next generation of the product.

* For confidentiality reasons, the name of this company cannot be displayed.
Problem solving or so-called “debugging” starts with failures in category S, followed by the failures in category A, and so on. All failures in the first three failure classes (A-C) should be solved before market introduction. The testers/developers indicate that most failures appear in these three failure categories. However, as a result of the time-to-market pressure, some of these failures can only be solved partially before product market introduction. For the same reason, other failures are transferred to lower failure categories. Which failures are transferred depends on the judgment of the product testers. These testers perform another user dissatisfaction prediction of the category A failures to identify the most critical failures that should actually be solved before market introduction.

The results of these interviews indicate that developers and testers of consumer electronics products independently or in groups predict the user dissatisfaction level of an identified product quality problem. In this context, the different failure classes (S, A-D) are considered a prediction of the UPFS level that is caused by a specific failure. Consequently, S and A failures correspond to an expected high UPFS level and D failures correspond to an expected low UPFS level. These developers/testers base their prediction on their personal experience as a product developer/tester at the company and their own practice as consumer product user at home. Moreover, as presented in Section 2.3, developers/testers of consumer electronics products do not receive (valuable) feedback from the field in order to adjust or update their failure impact judgment approach for user dissatisfaction prediction.

Based on the description of this practical approach to user dissatisfaction prediction in consumer electronics, a hypothetical UPFS prediction model is formulated. This hypothetical model is the so-called “expert-based UPFS prediction model” and is presented in Figure 4.1.

![Figure 4.1: Expert-based UPFS prediction model](image-url)

In this hypothetical model, the prediction of the user dissatisfaction level of a certain product failure made by the product developers/testers (category S and A-D) is used to predict the actual level of failure severity that is perceived by the users when experiencing this failure, the so-called UPFS. In other words, this model states that product developers/testers' evaluation of the user dissatisfaction of a certain failure agrees with the actual user dissatisfaction of that failure as experienced by the users. In the next section, the experimental design to test the validity of this hypothetical model is presented.
4.2 Experimental design

A behavioral measure like the expert-based UPFS is only valid when it measures what it has been designed to measure [KER93]. Especially construct validity is relevant in the context of this research. Construct validity concerns two issues at the same time: whether the construct being measured by a particular instrument is a valid construct and whether the particular instrument is adequate for measuring it [GOO05]. Construct validity is never established or destroyed with a single study. Rather, confidence in construct validity accumulates gradually and inductively as research produces supportive results [GOO05].

In order to validate the expert-based UPFS prediction model of Figure 4.1, an explorative consumer experiment was performed. In this experiment, a group of TV experts selected a high dissatisfaction failure scenario for the TV. Subsequently, the impact of this failure on actual users’ dissatisfaction (UPFS) was measured. According to the expert-based UPFS prediction model, this selected high dissatisfaction failure scenario should result in corresponding high values of UPFS among test participants. Consequently, a low average UPFS score among test participants in this experiment would seriously lower the construct validity of the expert-based UPFS prediction model. A detailed description of the experimental design is presented in the following subsections.

4.2.1 Failure scenario

A project team of ten TV experts (developers and testers) came up with a TV failure scenario that would according to their expertise result in a high level of UPFS among TV users. The TV experts unanimously agreed that this selected failure scenario would cause a high level of irritation among test participants experiencing the scenario. In other words, this failure scenario would be a “category A” failure in the product development process. The selection of this failure scenario by TV experts is an essential part of the experimental design: according to the expert-based UPFS model, the prediction of the scenario’s failure severity by the TV experts should predict the UPFS. As mentioned before, a mismatch between this prediction and the actual/measured UPFS would lower the construct validity of the expert-based UPFS model.

The selected failure scenario involves a failure in the teletext function of the TV. This failure scenario involves a deadlock in the information retrieval process of a requested teletext page. When triggered, this deadlock results in a black TV screen except from the teletext page number. The user can resolve this problem by closing the teletext function, by putting the TV on standby or by completely switching off the TV. The visibility of this failure scenario is presented in Figure 4.2a (normal teletext function) versus Figure 4.2b (failing teletext function).

4.2.2 Implementation of the failure scenario into the TV

Three members of the project team implemented this teletext failure scenario into the software of a TV such that the failure scenario was triggered when requesting a certain teletext page (in this case, page 757). Triggering this failure scenario resulted in the above described teletext failure.

---

* Teletext is an information retrieval service that offers text-based information via the TV cable signal.
4.2.3 Formulation of the task list

Before the experiment, test participants are not informed about the actual purpose of the experiment which is the validation of the expert-based UPFS model. Instead, participants are told that they are going to test the usability of the teletext functionality of the TV. This so-called participant deception occurs whenever research participants are not completely and fully informed about the nature of the research project before participating in it [STA98]. The use of deception in consumer research is often defended on the grounds that it is needed to get participants to act naturally and to enable the study of social phenomena [GOO05]. Moreover, telling test participants the purpose of the study and the procedures to be followed removes their unbiased and spontaneous attitude [ADA85]. In order to alleviate the potential impact of the deception on participants, it is important to extensively debrief all test participants. This debriefing should occur immediately after the research has ended and is designed to explain the purposes and procedures of the research and remove harmful after-effects of participation [ADA85], [STA98]. Consequently, participant debriefing is incorporated in the design of this experiment.

The “usability evaluation” of the teletext function is performed by means of a predefined task list. Based on the defined failure scenario, this task list for participants is formulated. The purpose of this task list is to guide the participants to the implemented failure scenario without directly instructing or influencing them. The task list for this experiment consists of three separate tasks. By performing the first two tasks, the test participants are familiarized with the teletext functionality in the TV. Moreover, by performing these tasks, participants can get used to the TV and the test environment. During the execution of the third task, participants are confronted with the actual failure scenario. In this third task, participants are asked to find the arrival time of flight KL 1168 from Helsinki. This arrival time can only be found on teletext page 757. However, when this page is triggered, the implemented failure scenario causes a black TV screen, except from the teletext page number. The participant can only resolve this problem by closing the teletext function, by putting the TV on standby or by completely switching off the TV. Subsequently, the UPFS of this failure scenario should be measured. The complete task list as used in this experiment is presented in Appendix B (in Dutch).
4.2.4 Selection of measurement tools

In the previous chapter, UPFS was defined as the level of irritation experienced by the user caused by a product failure. In order to validate the expert-based UPFS model, the UPFS scores corresponding to the defined failure scenario should be determined. Although the definition of UPFS is based on literature from several scientific fields, no off-the-shelf UPFS measurement tool is available. Nevertheless, irritation as experienced by the user is part of the Negative Affect (NA) dimension of the circumplex model (see Section 3.2). For this circumplex model and especially for NA, a measurement scale is available to determine e.g. the (change in) level of experienced irritation [BAG93], [WAT96], [WAT99A]. However, the application of this rather complex measurement scale was refrained from given the explorative nature of this first experiment.

Instead, a relatively straightforward approach of self-report was followed to determine the UPFS of the identified failure scenario. In self-report approaches, participants are asked straightforward to express their opinion or feelings with respect to a certain object, issue or situation [MYU00]. The main advantages of this self-evaluation approach are the relative ease of use and formulation. In a short survey, the self-report tool directly asks participants to score their level of irritation as caused by the teletext failure. However, self-report approaches are often susceptible to social desirability bias [FIS93], [JO97]. Social desirability bias is the pervasive tendency of individuals to present themselves in the most favorable manner relative to prevailing social norms [KIN00]. This bias is particularly influential in research involving social sensitive questions [RAG96], [HEI00]. Although the social norm concerning the expression of failure related irritation is unclear at this moment, the involved questions are unlikely to be considered “social sensitive”.

The self-report survey consists of five statements to measure the level of irritation experienced by the user as caused by the TV failure scenario:

**Statement 1:** The teletext function works without problems

**Statement 2:** I consider the problem in the teletext function as relatively unimportant

**Statement 3:** I would describe this problem in the teletext function as “very irritating”

**Statement 4:** The teletext function of this TV shows problems in use

**Statement 5:** At home, I would consider this failure as a minor shortcoming of the teletext function

After the appearance of the failure scenario, participants are asked to score the failure scenario according to these five statements. For each statement, participants indicate their level of agreement on a five point scale; from 1 meaning complete agreement till 5 meaning complete disagreement. In order to prevent so-called answering tendency, which is the inclination of participants to answer questions systematically, some of the statements have a reversed formulation [JAN98]. For statements 1, 2, 5 a high survey score corresponds to a high UPFS level. For statements 3 and 4, low survey scores correspond to a high UPFS level.

Another approach to limit this answering tendency of participants is to add other questions/statements than solely the ones directly related to the subject of interest.
Therefore, five other statements are added that do not consider UPFS. These added statements describe the expected reactions of the participant to the experienced failure scenario:

**Statement 6:** If I experienced this problem at home, I would tell my acquaintances about this problem

**Statement 7:** At home, I would consult the helpdesk of the TV manufacturing company

**Statement 8:** I would accept these kinds of problems in the teletext function of my TV

**Statement 9:** If I experienced this problem at home, I would ask a family member/friend for help

**Statement 10:** At home, I would have brought the TV back to the dealer for repair

Again, for each statement participants indicate their level of agreement on a five point scale. By mixing these reaction statements with the UPFS statements, the answering tendency of participants is restricted, because participants are forced to read each item carefully and to make decisions item-by-item. Moreover, these statements give extra insight into the expected response of users to the failure scenario. These reaction statements are not used for the measurement of the UPFS level. However, as indicated in Chapter 3, user reaction (behavior) to a failure also depends on the emotional response (UPFS) to that failure. Therefore, these last five statements will probably correlate with the first five UPFS statements. This correlation is not further investigated in this research. Altogether, the final survey consists of 10 statements measuring UPFS as well as participant reaction to the failure scenario. This survey is presented in Appendix C (in Dutch).

### 4.2.5 Selection of participants

Based on the available time and budget, thirty test participants were selected for the consumer experiment. Given the explorative nature of this experiment, these test participants were all acquaintances of the researchers in this project. In this context, it is important to consider the possible influence of this selected convenience sample on the experimental results. Acquaintances of the researchers are generally aware of the concerning research subject. Moreover, as test subjects they can be more tolerant to product failures because they are selected based on “their goodwill” regarding the involved researchers.

Nevertheless, the selection of participants prevented an unequal distribution of participants with respect to age (see Figure 4.3).

![Age Distribution](image)

*Figure 4.3: Age Distribution of test participants*
All participants were Dutch citizens and received a €15,- gift for their participation. As indicated above, participants were not informed about the actual purpose of the experiment, but were elaborately debriefed afterwards.

4.2.6 Experiment execution

One of the important aspects of a good experiment is that it has experimental control, which occurs to the extent that the experimenter is able to eliminate effects on the dependent variable (UPFS) other than the effects of the independent variable (the failure scenario) [STA98]. In order to limit the influence of these other effects, all the experiments were executed in a laboratory environment. This lab environment consists of two rooms which are separated by a one-way mirror; a living room and an observation room. The living room is the domain of the test participant. This room is decorated as a standard living room except for the three observation camera's and the one-way mirror. These observation tools allow people in the observation room to observe the experiment and to tape the actions of the test participants. Figure 4.4 on the next page gives an overview of the layout of both rooms.

During the experiment, one researcher was present as an observer in the living room. At the start of each experiment, this researcher welcomed the participant in the living room and briefly explained the purpose of the experiment. Subsequently, the researcher handed out the task list of the experiment to the participant (see Appendix B). During the experiments, this researcher only answered questions about operational aspects of the experiment and not about the TV or its functions. Furthermore, the researcher stimulated the participants to think-aloud during the experiment, so that the line of reasoning and operation of the participants became distinct for all researchers. Two minutes after the activation of the failure scenario, the experiment was terminated and the researcher handed out the UPFS survey to the test participant (see Appendix C). After finishing this survey, the test participant was debriefed about the actual purpose of the experiment. During this experiment, the other researchers were present in the observation room. They took care of the camera operation and videotaping of the experiments.

Based on all the information provided in this section, Figure 4.5 gives an overview of the design of this experiment. The experiment started with the selection of thirty test participants in the pre-experimental phase.

Figure 4.4: The experimental room (The arrows indicate the positions and capture directions of the three cameras)
Subsequently, each participant performed two introduction tasks followed by the failure scenario task. Next, the UPFS of this failure scenario was measured for each participant using the UPFS survey. Lastly, each participant was elaborately debriefed.

In order to validate the expert-based UPFS prediction model, the results of this explorative consumer experiment need to be analyzed. Particularly, the UPFS scores of the test participants should be compared with the high dissatisfaction prediction of the experts for this failure scenario. The next section, Section 4.3, discusses the experimental results in detail.

4.3 Experimental results

The actual experiments were performed in week 5 till week 7 of 2006. As mentioned above, thirty participants evaluated the teletext failure scenario using the UPFS survey. The results of these experiments should indicate whether the expert-based UPFS prediction model is valid within the context of consumer electronics products. However, before these experimental results can be discussed, the reliability of the UPFS measurement tool should be evaluated. A good reliability of the measurement tool is a prerequisite for the correct appraisal of the experimental results. Therefore, the first subsection deals with the reliability of the UPFS measurement results. Subsequently, the second subsection deals with the UPFS evaluation of the failure scenario and the validity of the expert-based UPFS prediction model. Lastly, the third subsection deals with some other useful experimental results that provide more insight in the UPFS evaluation process of the participants.
4.3.1 Reliability of the UPFS measurement

As mentioned above, a good reliability of the UPFS measurement scale is a prerequisite for the correct appraisal of the experimental results. Reliability of a measure refers to the extent to which it is free from random error [KER93], [GOO05]. Especially, internal consistency has become the most popular and most accurate way of assessing reliability for both trait and state measures [STA98]. Internal consistency refers to the extent to which the scores on the items (statements) correlate with each other and thus all measure the true score rather than random error [STA98]. The most common, and the best, index of internal consistency is known as Cronbach's coefficient alpha or Cronbach's alpha (symbolized as $\alpha$) [STA98]. The Cronbach's alpha is a measure that assesses the consistency of the entire measurement scale and ranges from $\alpha = 0.00$ (indicating that the measure is entirely error) to $\alpha = 1.00$ (indicating that the measure has no error) [STA98], [HAI05]. A general rule of thumb indicates that $\alpha = 0.70$ is considered the minimum required Cronbach's alpha value for a measurement scale in order to be considered reliable [HAI05]. This rule of thumb is applicable for measurement scales consisting of a limited number of items (number of items < 11) since the Cronbach's alpha value may increase with the number of included items [PET94], [CHU79]. The UPFS measurement scale consists of five items meaning that the minimum value of $\alpha = 0.70$ can be applied in this analysis. Table 4.1 gives an overview of the Cronbach's alpha values of the UPFS measurement scale.

<table>
<thead>
<tr>
<th>Statement excluded</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.71</td>
</tr>
<tr>
<td>Statement 1</td>
<td>0.79</td>
</tr>
<tr>
<td>Statement 2</td>
<td>0.60</td>
</tr>
<tr>
<td>Statement 3</td>
<td>0.59</td>
</tr>
<tr>
<td>Statement 4</td>
<td>0.65</td>
</tr>
<tr>
<td>Statement 5</td>
<td>0.63</td>
</tr>
</tbody>
</table>

The first value (0.71) indicates that the overall reliability of this scale is sufficient. However, for application in follow-up experiments, the reliability of the UPFS measurement scale could be improved by adjusting the first statement of the UPFS survey. Just by omitting this statement (“The teletext function works without problems”) in the reliability evaluation, Cronbach's alpha raises to 0.79. Exclusion of other statements does not further improve the reliability of the measurement scale. This analysis demonstrates that the self-report UPFS measurement scale used in the experiment is reliable.

4.3.2 Validity of the expert-based UPFS model

The results of the analysis in the previous subsection demonstrated the reliability of the UPFS measurement scale in this experiment. In this section, the validity of the expert-based UPFS prediction model is further analyzed.

As presented earlier, the UPFS scale consists of five statements. For each statement participants indicate their level of agreement on a five point scale; from 1 meaning complete agreement till 5 meaning complete disagreement. For every participant, the scores for all statements are added to one aggregated UPFS score.
Consequently, these aggregated UPFS scores can range from 5 (a very low UPFS score) till 25 (a very high UPFS score). All UPFS scores above 20 are considered high. The expert-based UPFS model predicts a high average UPFS score for the failure scenario which can be translated into the following null- and alternative hypothesis:

\[
H_0: \text{Average UPFS} > 20 \quad \quad H_1: \text{Average UPFS} \leq 20
\]

Table 4.2 gives an overview of the UPFS scores of the participants in this experiment.

<table>
<thead>
<tr>
<th>UPFS score of participants (N=30)</th>
<th>Average score</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 16 13 15 19 14 11 15 9 13 7 18 22 11 19</td>
<td>14.60</td>
<td>4.36</td>
</tr>
</tbody>
</table>

Subsequently, in Figure 4.6 on the next page, the UPFS scores of Table 4.2 are summarized in a bar chart. The results in the table and graph indicate a low average level of UPFS among the participants. The average UPFS score among the participants is 14.60. Twenty participants scored the failure scenario lower than or equal to the numerical centre of the UPFS measurement scale, which is 15.

However, in order to (in)validate the expert-based UPFS prediction model, the above formulated null hypothesis should be statistically tested. For samples of size \( n \geq 30 \) regardless of the shape of the population, the normal-based population confidence interval will provide good results [MON99]. A 95\% confidence interval of the UPFS data ranges from 12.97 till 16.23 \((14.6 \pm 1.63)\), not including the expected high UPFS value \(20\). In other words, this confidence interval demonstrates that the null hypothesis should be rejected. In other words, the measured UPFS scores indicate that on average this failure scenario does not result in a high level of irritation experienced by the participants of the experiment.

![Figure 4.6: Overview of UPFS scores (Low: UPFS score \(\leq 10.00\), Below average: \(10.00 < \text{UPFS score} \leq 14.60\), Above average: \(14.60 < \text{UPFS score} \leq 20.00\), High: UPFS score > 20.00)
As mentioned earlier, according to the expert-based UPFS prediction model the selected high dissatisfaction failure scenario should result in corresponding high values of UPFS among test participants. Consequently, a low average UPFS score among test participants in this experiment seriously affects the construct validity of the expert-based UPFS prediction model. But again, construct validity is never established or destroyed with a single study. Rather, confidence in construct validity accumulates gradually and inductively as research produces supportive results [GOO05].

However, in addition to this experiment, several other recent sources in literature dispute the ability of experts to predict users' preferences and behavior in consumer electronics. In his book “The inmates are running the asylum”, Cooper states that “programmers devote so much of their time and energy to learning about software that it is inconceivable to them that a user would not want to take the time to understand his work” [COO04]. Moreover, these product developers use themselves as their frame of reference, so they make products that are easy to use for other engineers, not for average users [COO04]. Nielsen indicates that “designers are not users” [NIE93]. Kujala and Kauppinnen state that developers often have a vague or contradictory sense of their intended users and that these developers tend to base use scenarios on people similar to themselves [KUJ06]. Hsu et al. illustrate the fact that many differences exist between designers and users, and subsequently they explain the resulting challenge of designers to transfer consumer needs into technical and design specifications [HSU00]. Furthermore, a very recent study performed by Karapanos and Martens confirms these prior findings by demonstrating that designers often fall short in accounting for users' views on the product and foreseeing their preferences [KAR07].

Given the restricted time and resources that are available in the product development process of consumer electronics products, the overestimation of the dissatisfaction levels that are caused by unimportant product failures automatically results in the underestimation of these levels for actual important failures. Although the construct validity of a model cannot be destroyed with a single experiment, the combined contribution of the above mentioned prior research results and the results of this experiment highly invalidate the expert-based UPFS prediction model in consumer electronics industry. In other words, the combination of results demonstrates that product developers are often unsuccessful in the prediction of the level of irritation experienced by a user caused by a product failure. Consequently, the theoretical foundation for the application of this UPFS prediction model in practice (e.g. consumer electronics industry) is lacking. And as explained earlier, the limited validity of this practical UPFS model might well contribute to the increasing number of customer complaints about consumer electronics products.

4.3.3 Other results

The previous subsection showed that the expert-based UPFS prediction model is unsuitable for predicting the impact of product failures on user dissatisfaction in the context of consumer electronics. Nevertheless, this first experiment generated some extra results that may be useful for the development or validation of a new hypothetical UPFS model in the following chapters of this thesis. The first group of additional results is an overview of reactions of the test participants to the failure scenario. By responding to the five reaction statements (statement 6 till 10) participants indicated how they would respond to a similar failure scenario at home. The combined results to these statements give some insight into the escalation behavior of a product failure into the service channel of a company. Figure 4.7 gives an overview of the reactions by the participants to the experienced failure scenario.
This figure indicates that on average people would accept these kinds of problems in the teletext function of their TV (statement 8). Hardly anybody would call the helpdesk for this problem (statement 7) or bring the TV to the dealer for repair (statement 10). On the other hand, several persons would tell other people about this problem (statement 6) or would ask a family member for help (statement 9). These results indicate that the escalation of this failure into the service channel of the consumer electronics company is very restricted. This could be interpreted as another indication of the limited impact of the failure scenario on user (dis)satisfaction. In other words, participants do not consider the failure serious enough to consult the helpdesk or product dealer.

![Participant reaction](image)

**Figure 4.7:** Participant reaction (Average agreement levels from 1: complete disagreement till 5: complete agreement)

A second group of additional results relates to the assumed causes of the failure. Just before the debriefing, participants were questioned about the cause of the failure in the teletext function. Participants were given complete freedom to formulate their assumed cause of the failure scenario. Eventually this resulted in six categories of failure causes:

- **TV:** the failure is caused by an internal failure of the television
- **Cable:** the failure is caused by a disruption in the cable signal
- **Broadcast:** the failure is caused by a mistake of the broadcasting company
- **Airport:** the failure is caused by the airport by not providing the flight information to the broadcasting company
- **Self:** the failure is caused by me. I did something wrong
- **Unknown:** the cause of the failure is unknown

Figure 4.8 gives an overview of the number of times a certain failure cause was mentioned by the participants. Surprisingly, the TV was only appointed four times as assumed failure cause. On the other hand, the cable and broadcasting companies were frequently (together fifteen times which corresponds to 50% of all cases) allocated as causes of the concerning failure scenario.
These results indicate that most participants did not consider the TV as the main cause of the failure scenario. Although the failure was caused by an internal failure in the TV, the majority of the participants allocated another external cause to the failure. However, the influence of this (wrong) external failure attribution on the experienced UPFS is unknown. In other words, the attribution of the failure cause outside the TV might have negatively or positively influenced the measured levels of UPFS among participants.

![Assumed causes of failure](image.png)

**Figure 4.8:** Assumed failure causes

### 4.4 Conclusions

The first version of the UPFS prediction model is based on the user dissatisfaction prediction approach as applied in the development process of consumer electronics products. This model states that product developers'/testers' evaluation of user dissatisfaction of a certain failure agrees with the actual user dissatisfaction of that failure expressed by the UPFS. In order to validate this expert-based UPFS prediction model, an explorative consumer experiment was performed.

The results of this experiment do not support the validity of the expert-based UPFS model. With respect to customer dissatisfaction prediction the practical approach is proven to be invalid. Furthermore, several other recent sources in literature dispute the ability of experts to predict users' preferences and behavior with respect to consumer electronics. The combination of these results confirms that product developers are often unsuccessful in the prediction of the level of irritation experienced by a user caused by a product failure. Consequently, the theoretical foundation for the application of this UPFS prediction model in practice (e.g. in consumer electronics industry) is lacking.

Although the design of this first explorative experiment is rather straightforward, the results are quite revealing. Based on the combined results of previous research and this experiment, the validity of the expert-based UPFS model is proven to be very limited. Furthermore, the first attempts into the development of a reliable UPFS measurement scale appear to be quite promising. Although for future experiments some of the statements should be improved, the overall reliability of the scale was already sufficient (Cronbach's alpha: 0.71). Moreover, the exit interviews among participants about the failure cause revealed a possible influence of failure attribution on UPFS in consumer electronics.
The explorative character of the experiment is also associated with some limitations of the results. On the first place, as mentioned before, it is important to consider the possible influence of the selected convenience sample on the experimental results. Acquaintances of the researchers are generally aware of the concerning research subject. Moreover, as test subjects they can be more tolerant to product failures because they are selected based on “their goodwill” regarding the involved researchers. This combination of effects might have caused lower UPFS levels among test participants than would have been measured among a random group of test participants.

On the second place, the convenience sample of participants does not control for any differences between test participants. A more homogeneous group of test participants would most presumably interfere less with the relationship between failure scenario and UPFS.

Thirdly, to further support the validity of the self-report UPFS measurement scale, this scale should be complemented with an existing and more sophisticated measurement scale that is based on the circumplex model of affect. For this circumplex model and especially for NA, a measurement scale is available to determine for example the (change in) level of experienced irritation [BAG93], [WAT96], [WAT99A]. The application of such a measurement scale for measuring UPFS in future experiments could further validate the self-report UPFS measurement scale.

Furthermore, the possible influence of the failure scenario description on the experimental results should not be ignored. Although in this description of the scenario the importance of the teletext content is stressed (“you have to pick up your wife/husband at the airport, find the time of arrival on teletext”, see Appendix B) this importance might not have reached the test participants. In that case, the description of the scenario might have lowered the measured UPFS among test participants.

Lastly, it is important to mention that the laboratory environment in which the experiments took place may have influenced the behavior of the test participants. As mentioned before, all the experiments were executed in a laboratory environment in order to limit the influence of other (external) effects. However, this laboratory environment makes test participants respond differently to the failure scenario than they would have done in their normal home situation. Consequently, this laboratory environment may influence the UPFS measurement results of this experiment. But, in general, the results of laboratory research mirror the results of field research as elaborate research demonstrates a high degree of correspondence between the results found in and outside the lab [AND99], [GOO05].

Based on the lessons learned from this first experiment and complemented with relevant literature, the next chapter describes an adjusted UPFS prediction model. Again, the validity of this adjusted UPFS model is evaluated in a consumer experiment.
5 An UPFS Model with Theoretical Aspects

Chapter 3 illustrated that insight into the UPFS of product failures would be useful in order to reduce the number of product complaints. For this reason, the previous chapter formulated a so-called expert-based UPFS model based on a practical approach that is currently used in consumer electronics industry. However, a first consumer experiment together with related literature indicated that product developers are often unsuccessful in the prediction of the level of irritation experienced by a user, caused by a product failure.

This chapter focuses on developing a theoretical UPFS prediction model in consumer electronics. This adjusted prediction model is based on the results of the first UPFS consumer experiment (see Chapter 4), literature from different scientific fields, and an expert validation session with people from academia.

In Section 5.1, this adjusted UPFS prediction model is presented. The starting point of this model is a high-level UPFS model based on emotion theory. Subsequently, this high-level model is developed into a more detailed UPFS prediction model, resulting in a so-called theory-based UPFS prediction model. In Section 5.2, one hypothesized influential part of this theory-based UPFS model is described in more detail, namely the potential influence of Function Importance on UPFS.

5.1 The theory-based UPFS prediction model

This section describes the development of an adjusted UPFS prediction model in consumer electronics. As indicated earlier, the adjusted UPFS model is based on the analysis of the results of the first UPFS consumer experiment (see Section 4.3), literature from different scientific fields (e.g. quality and reliability, medicine, marketing and safety) and a validation session with people from academia with a background in product reliability, usability, product development and testing.

5.1.1 A high-level UPFS prediction model

The starting point for this adjusted model is the emotional response model from consumer behavior theory (see Section 3.2), as the original definition of UPFS is based on this model.

In Chapter 3, UPFS is defined as the level of irritation experienced by the user as caused by a product failure. In this definition, UPFS is characterized as an emotion measurement and therefore depends on a certain stimulus (product failure) and individual and environmental characteristics [CHA06]. The direct translation of the emotional response model from consumer behavior theory into a high-level UPFS model is illustrated in Figure 5.1. In this high-level UPFS model, product failure is indicated as the stimulus of the emotional response (UPFS). Consequently, the characteristics of this failure directly influence the intensity of this emotional response. Besides the stimulus, several individual and environmental characteristics influence the emotional response as well. In the context of this UPFS model, individual characteristics correspond to the personal traits of the product user. In other words, the characteristics of a user directly influence the level of UPFS.
In this high-level UPFS model, the environmental characteristics correspond to the use conditions in which a user operates the product. These use conditions also directly influence the level of UPFS.

![High-level UPFS model](image)

The scope of this high-level UPFS model is rather broad. Given the strict time and resource constraints of a PhD project (4 years and 1 person) this research mainly focuses on the explanation of the influence of the stimulus (failure characteristics) on the emotional response (UPFS) and treats the individual- and environmental characteristics (user characteristics and use conditions) as extraneous variables. An extraneous variable is a variable that is not of interest in the study, but which might have influence on the relationships being studied [STA98], [GOO05]. In order to restrict the influence of these extraneous variables on UPFS in this research, these variables should be kept as constant as possible. This research focus is also represented in Figure 5.1 by the solid arrow from failure characteristics to UPFS and the dotted arrows from the user characteristics and use conditions to UPFS.

This high-level UPFS model describes the expected influence of the failure characteristics on UPFS in general. However, in order to predict the UPFS resulting from a certain product failure, the relationships between the individual failure characteristics and UPFS should be explicated. The exploration of these relationships is presented in the next subsection.

### 5.1.2 Detailed UPFS prediction model

In order to develop a more detailed UPFS prediction model, the potentially influential failure characteristics should be determined. The identification of these failure characteristics is based on the following information sources:

- Literature from different scientific fields (e.g. quality and reliability engineering, safety engineering and chemical engineering)
- The results of the first UPFS consumer experiment (see Chapter 4)
- Expert validation session with people from academia with a background in product reliability, usability, product development and testing

**Failure Characteristics: literature from different fields**

In Section 2.2, it was concluded that the combination of lacking user-orientation and confined specificity restricts the potential uncertainty reduction of current quality improvement methods.
However, quality improvement methods that do bear a certain degree of specificity suggest some conceivable relevant failure characteristics. The most promising method in this context is the FMEA method. This approach was presented in Section 2.2. FMEA is a traditional quality and safety analysis technique, which has enjoyed extensive application to diverse products over several decades. The FMEA method consists of a systemized series of activities intended to discover failures and to recommend corrective actions for design improvements [LEV03]. In essence, the FMEA provides a systematic method of examining all possible product failure scenarios. The FMEA includes an overview of failure modes, possible causes of each failure, effects of the failure and their seriousness, failure occurrence and its likelihood of detection [BLI00]. There are three components that help to define the priority of failures in a classical FMEA:

- Likelihood of Occurrence (O)
- Severity (S)
- Likelihood of Detection (D)

The impact of these components (occurrence, severity and likelihood of detection) on UPFS is not described in literature. On the other hand, these components could quite easily be translated into failure characteristics that possibly influence UPFS. For example, the failure occurrence clearly relates to failure frequency. In this context, failure frequency (FF) is defined as the number of failures per time unit under standardized use conditions. The severity in a FMEA, which in meaning distinctly differs from User Perceived Failure Severity, relates to the failure characteristic of failure impact. Failure impact (FI) can be defined as the percentage loss of functionality as a result of the failure. Likelihood of detection also relates to failure impact and frequency. Frequent failures with a high impact are easily detected. However, likelihood of detection is also related to failure reproducibility (FR), which is the degree of repeatability of a failure by the user since a product problem that is reproducible by the user is expected to be easier to recognize as being a failure. Altogether, the FMEA components can not be directly used as failure characteristics in the UPFS model, but some relevant failure characteristics can be derived from this approach.

In order to identify more relevant failure characteristics for the UPFS model, literature from other research fields has been consulted. Several scientific fields (e.g. medicine, marketing, safety) use arithmetical models similar to FMEA for failure impact assessment, resulting in the failure characteristics of failure frequency, failure impact and failure reproducibility. On the other hand, the field of chemical engineering makes use of Risk Matrices [STE97], [BAY02]. Although these Risk Matrices show some agreement with the FMEA approach, several analysis factors are added. A typical Risk Matrix contains the following factors [CRO01]:

- **Consequence**: This factor is related to the severity of the hazard (similar to the severity factor in FMEA) but also to the nature of the hazardous condition. The severity of a failure again relates to failure impact. However, the nature of the hazardous condition relates to the failure characteristic of function importance (FUI): the relative importance of the function affected by the failure.

- **Exposure frequency**: Exposure frequency is calculated from the expected personnel exposure in the hazard zone. The factor is directly related to the earlier mentioned failure characteristic of failure frequency.

- **Possibility to escape**: This parameter deals with the opportunity of avoiding injury after the hazard has occurred. This factor is related to the failure
solvability (FS), which indicates the required effort a user should take after failure occurrence to return to normal functioning of the product (excluding the failed part of the function or product).

- **Possibility of avoiding the hazardous event:** As the name indicates, this factor describes the opportunity of preventing the failure/problem. This factor relates to the failure characteristic *failure work around*. Failure work around (FWA) indicates the degree to which the failure occurrence can be prevented by the user by operating the product differently.

**Failure Characteristics: results of the first UPFS consumer experiment**

In the previous chapter, the validity of the expert-based UPFS model was evaluated in an explorative consumer experiment. The results of this first UPFS consumer experiment indicated that most participants did not consider the consumer electronics product (TV) as the main cause of the experienced failure. Although the failure was caused by an internal failure in the TV, the majority of the participants attributed an external cause to the failure. However, the influence of this (incorrect) external failure attribution on the experienced UPFS is unknown. In other words, the attribution of the failure cause outside the TV might have negatively or positively influenced the measured levels of UPFS among participants. Consequently, based on these experimental results, the variable *failure attribution* is added as failure characteristic to the theory-based UPFS model. Failure attribution or causal attribution knows several definitions in contemporary literature [HEI58], [WEI85], [TAY90] and [OLI96]. In the context of this research, failure attribution (FA) is defined as the cause to which users attribute the failure.

**Failure Characteristics: expert validation session with people from academia**

Six people from different scientific areas joined an expert validation session on the hypothetical failure characteristics. All these people have a background in product reliability, usability, product development or testing. The purpose of this session was twofold:

- Identification of relevant failure characteristics other than the characteristics identified in the literature search
- Determination of the hypothetical relations between the identified failure characteristics and UPFS

The participants of the validation session were not informed about the relevant failure characteristics identified from literature. The purpose of this approach was to prevent participants from limiting their contribution to this session to similar or related failure characteristics. Nevertheless, the session resulted in many failure characteristics rather similar to earlier identified characteristics. One failure characteristic was added during the validation session namely, *failure moment in use process*. This failure moment in use process (FMUP) indicates when the failure occurs in the total use-life of a product (e.g. during unpacking, installation, first use regular use or just before the end of the warranty period).

The participants of the expert validation session confirmed that all failure characteristics identified in the literature search are potential influencing factor of UPFS. The combination of the high-level UPFS model (see Figure 5.1) and all these hypothetical failure characteristics results in the theory-based UPFS model. This model is presented in Figure 5.2 on the next page.

The second purpose of the validation session was the determination of the hypothetical relations between the eight identified failure characteristics and UPFS.
The outcome of the session related to this purpose, is presented in Table 5.1. This table gives an overview of the eight hypothetical failure characteristics together with their abbreviation, a description and their expected relation with UPFS. A positive influence is expected of Failure Frequency, Failure Impact, Failure Solvability and Function Importance on UPFS. In other words, a higher level of these failure characteristics is expected to result in a higher level of UPFS. A negative influence is expected of Failure Reproducibility and Failure Workaround on UPFS. This means that a higher level of these failure characteristics is expected to result in a lower level of UPFS. For Failure Moment in Use process, the hypothetical relation implicates that a failure earlier in the use-life of the product results in a higher level of UPFS. As indicated above, the direction of the expected relation between Failure Attribution and UPFS is unknown at this moment.

![User Perceived Failure Severity Diagram](image)

**Figure 5.2:** The theory-based UPFS model

### 5.2 Influence of Function Importance on UPFS

As mentioned before, given the time and resource constraints of this research project, the complete validation of the theory-based UPFS model within this project is unfeasible. Furthermore, the validation of the complete model requires a gradual approach in which the (combined) added values of all failure characteristics to the model are evaluated. However, this research starts small-scale with investigating the influence of one of these failure characteristics on UPFS. In this research, Function Importance is selected as the failure characteristic to be evaluated because the influence of this variable on UPFS is expected to be substantial. In other words, the influence of Function Importance (FUI) as independent variable on UPFS as dependent variable is investigated.

A positive relation is expected between the FUI level and UPFS. In other words, a failure in an important product function is expected to cause a higher level of user irritation than a failure in a less important function of that same product would cause. This hypothetical relation between FUI and UPFS is evaluated in a second consumer experiment.
Table 5.1: Overview of the identified Failure Characteristics

<table>
<thead>
<tr>
<th>Failure Characteristic</th>
<th>Abbrev.</th>
<th>Description</th>
<th>Hypothesized relation with UPFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure Attribution</td>
<td>FA</td>
<td>The cause to which users attribute the failure</td>
<td>Unknown</td>
</tr>
<tr>
<td>Failure Frequency</td>
<td>FF</td>
<td>Number of failures per time unit under standardized use conditions</td>
<td>Positive</td>
</tr>
<tr>
<td>Failure Impact</td>
<td>FI</td>
<td>The percentage loss of functionality as a result of the failure</td>
<td>Positive</td>
</tr>
<tr>
<td>Failure Moment in Use Process</td>
<td>FMUP</td>
<td>The phase of the product use life in which the failure occurs</td>
<td>Negative*</td>
</tr>
<tr>
<td>Failure Reproducibility</td>
<td>FR</td>
<td>The degree of repeatability of a failure by the user</td>
<td>Negative</td>
</tr>
<tr>
<td>Failure Solvability</td>
<td>FS</td>
<td>The required effort a user should take after failure occurrence to return to normal functioning of the product (excluding the failed part of the function)</td>
<td>Positive</td>
</tr>
<tr>
<td>Failure Work Around</td>
<td>FWA</td>
<td>The degree in which the failure occurrence can be prevented by the user by operating the product differently</td>
<td>Negative</td>
</tr>
<tr>
<td>Function Importance</td>
<td>FUI</td>
<td>The relative importance of the function affected by the failure</td>
<td>Positive</td>
</tr>
</tbody>
</table>

* In this case a negative relationship implies: A failure earlier in the use life of a product results in a higher level of UPFS
The experimental model for this first validation step of the theory-based UPFS model is presented in Figure 5.3.

![Experimental model FUI experiment](image)

Before the relation between FUI and UPFS can be evaluated in an actual consumer experiment, the variables in this experimental model need to be described in more detail. An elaborated description of the dependent variable UPFS was already given in Chapter 3. Additionally, the next subsection describes the independent variable FUI. Subsequently, in the second subsection, the extraneous variables of this experimental model are presented.

### 5.2.1 Independent variable: Function Importance

In the previous section, FUI was defined as the relative importance of the function affected by the failure. This FUI definition requires a further interpretation of the term “function”. Therefore, this subsection starts with the definition of a function in the context of this research. Subsequently, the second part of this subsection deals with relevant FUI literature.

#### Functions

An important undefined term in the definition of FUI is the word “function”. A function indicates the task fulfillment of a product. In the context of this research, a function is defined as:

\[ \text{The whole of technical attributes that in the eyes of the user is required to execute one task} \]

An attribute is a property of a product that, together with other attributes, is needed to execute a task of a function. This user-oriented definition differs from more traditional and technical definitions. In these technical definitions, a function is often defined as a technical characteristic of a product; something the product has or is able to do. However, in the context of this UPFS research, a definition that specifies a function from a user perspective is considered more useful. In this thesis, the above mentioned technical characteristics of a product are termed features. Table 5.2 gives an overview of the above mentioned concepts.

### Table 5.2: Overview of concepts

<table>
<thead>
<tr>
<th><strong>Function</strong></th>
<th>The whole of technical attributes that in the eyes of the user is required to execute one task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attribute</strong></td>
<td>A property of a product that, together with other attributes, is needed to execute a task of a function</td>
</tr>
<tr>
<td><strong>Feature</strong></td>
<td>A technical characteristic of a product, something a product has or is able to do</td>
</tr>
</tbody>
</table>

69
Function Importance in literature

The expected relation between FUI and user dissatisfaction in consumer electronics products is not discussed in contemporary literature. Nevertheless, the general concept of FUI and several related concepts are presented in literature. This subsection gives an overview of this FUI related literature.

Consumers' evaluation

Myers and Alpert define two product evaluation levels of consumers [MYE68]:

- Overall attitude towards the product in terms of its suitability or desirability
- Attitudes towards each of the product's component functions

This second type of attitudes can be used to determine the important functions of a product and their influence on the overall evaluation of a product. Myers and Alpert use the terms “feature determinance” or “attribute determinance” to characterize FUI [MYE68]. The attitudes toward the various functions together determine the overall product evaluation or attitude.

There are three keywords that are used in attitude research to determine the preference of users for the different functions of a product [MYE77]:

- Salience
- Importance
- Determinance

Salience has to do with a function being at “the top of users' minds”. Salience functions are the functions that are mentioned first when users are asked to elicit functions that are considered important. Important functions are presumed to have some consequence or significance in making choices among products and brands or in the overall evaluation of the product. Determinant functions are functions which are most closely related to preference or to the actual purchase decisions of the user [MYE68], [MYE77].

Earlier studies into the relationship between importance and salience generally show a positive correlation between the two concepts [KRE48], [HAR00]. Apparently, for the evaluation of products people rely on a limited number of important functions, which are most salient (accessible in their memory). However, there are several situations in which salient functions are unimportant and vice versa [KRE48]. Important product functions are often taken for granted by users (e.g. the dialing function of a mobile phone) and consequently become non-salient among users. This example of a contrast between function importance and function salience also illustrates the distinction between important and determinant functions. A function that is considered important but non-salient because it is taken for granted by the user, will not determine product choice, since that function is supposed to have equal and satisfactory levels among products [MYE77]. A function that is ranked high in importance by consumers, but is not considered to differ widely among products and/or brands, is not a determinant function in the actual purchase [MYE68]. Functions for which the differences are greater among products can cause a true differentiation between product offers. Subsequently, these functions are the ones that sell a product. In this thesis, the earlier defined FUI concept corresponds to the function importance concept of Myers and Alpert. As demonstrated by different sources of literature, function salience and function determinance are related concepts, but represent a different aspect of the product evaluation process.
**Importance to buy versus importance to use**

There are two distinct phases of the product life cycle process in which FUI can be determined. The FUI can be measured before buying the product, in the so-called product selection phase. However, the FUI can also be determined after buying the product, in the product use phase. In the product selection phase, the emphasis is on what a consumer considers important and what drives the decision-making process. This is called “function importance to buy”. In the product use phase, the emphasis is on the actual use of the product and on what the user experiences as being important functions. Consequently, this is called “importance to use”. The functions that are considered most important by users often differ between these two phases. A user might buy a product because it has a specific function. But in the use phase, the same user may consider this function rather unimportant and may never use it.

In this thesis, “the function importance to buy” is selected as measure for FUI. This decision is based on two considerations. In the first place, the large majority of all FUI measurement methods described in literature is developed to measure pre-sales importance from a marketing viewpoint. Therefore, these measures generate reliable results when used in a pre-sale context and results are of unknown reliability when these measures are used in an after-sale context. Secondly, in the context of this research, the after-sales FUI measurement might be influenced by the raised UPFS level in the consumer experiment. In other words, the level of user dissatisfaction that is caused by an implemented product failure in the consumer experiment might affect the results of the FUI measurement. Oliver indicates that failing functions during use naturally become more important because of their immediacy [OLI96]. The FUI should be an independent variable, without any bias from the dependent UPFS variable. This would be harder to realize if the used method would measure FUI in the product use phase.

**5.2.2 Extraneous variables: users and their environment**

The experimental model (Figure 5.3) describes the relation between FUI and UPFS. However, this relation between FUI and UPFS may be biased by the influence of extraneous variables (see Figure 5.2 and 5.3). As mentioned before, an extraneous variable is a variable that is not of interest in the study, but which might have influence on the relationships being studied [STA98], [GOO05]. In order to limit the influence of these variables on the dependent variable (UPFS), extraneous variables should be controlled.

Previous analysis identified two relevant classes of extraneous variables in this research, namely user characteristics and use conditions. User characteristics correspond to the personal traits of the product user and use conditions indicate the circumstances in which a user utilizes the product. Although the influence of these variables on UPFS is not of interest in this research, extraneous variables should be kept as constant as possible to limit interference with the outcomes of the experiment.

**Use conditions**

In the context of this research, this use conditions can be divided into two subgroups:

- **The use environment**: the physical conditions in which the product is used e.g. the space where the product is used, the presence of other products or people in this space and the light conditions in this space
**The use approach:** the way the product is utilized by the user e.g. whether he/she uses the product according to a predefined task list, whether he/she has the possibility to ask questions and whether he/she starts using the product from "out-of-the-box" or already unpacked.

Both subgroups should be controlled in order to limit their interference with the outcomes of the experiment.

**User characteristics**

As mentioned in Section 5.1, investigating the influence of different user characteristics on UPFS is not within the scope of this research project. However, unconsidered disproportionalities in certain user characteristics may seriously bias the experimental results. By controlling for these user characteristics, sounder experimental results can be obtained.

The experimental model of Figure 5.3 describes the relation between FUI and UPFS. A user characteristic that is expected to interfere with the UPFS is so-called irritableness. Some people get irritated by minor product failures, whereas other people only get irritated by major product failures. A disproportionate distribution of irritableness levels over the experimental groups may reduce the likelihood of finding a significant correlation between FUI and UPFS. Therefore, there should be controlled for the irritableness level of participants. Furthermore, the influence of all other user characteristics should be limited by the selection of a homogeneous group of test participants.

**Other failure characteristics**

Lastly, it is important to mention that all other failure characteristics of which the influence is not investigated (see Figure 5.2: Failure Attribution, Failure Work Around, Failure Reproducibility, Failure Moment in Use Process, Failure Solvability, Failure Frequency and Failure Impact) are also extraneous variables in this research that should be controlled. Otherwise, the influences of these variables on UPFS may bias the results of the actual research.

**5.3 Conclusions**

The starting point for the adjusted UPFS model was the emotional response model from consumer behavior theory. Subsequently, this high-level emotional response model was developed into a more detailed UPFS model, resulting in the theory-based UPFS prediction model. This theory-based UPFS model predicts user dissatisfaction that is caused by a product failure based on the characteristics of the failure, characteristics of the user and the product use conditions (see Section 5.2). This research concentrates on examining the impact of failure characteristics on UPFS. More specifically, the influence of Function Importance (FUI) (independent variable) on UPFS (dependent variable) is to be investigated. It is hypothesized that a positive relation exists between FUI and UPFS. Moreover, the use environment, the use approach, user irritableness and other user characteristics are expected to influence this relation between FUI and UPFS.

In order to validate this relation between FUI and UPFS as presented in the experimental model of Figure 5.3, a second consumer experiment is performed. However, before this relation can be evaluated in an actual consumer experiment, measurement approaches for the different variables of the experimental model need to be determined and validated. Therefore, the next chapter deals with the description and validation of the measurement approaches for the independent, dependent and extraneous variables of the model.
Measurement of the Experimental Variables

The previous chapter presented the theory-based UPFS prediction model. This research concentrates on validating part of this model, namely the impact of failure characteristics on UPFS. More specifically, the influence of Function Importance (FUI) on UPFS is to be investigated. This hypothesized relationship is presented in the experimental model of Figure 6.1.

![Experimental model FUI experiment (based on Figure 5.3)](image)

This chapter deals with the description and validation of the measurement approaches for the independent (FUI), dependent (UPFS) and extraneous variables (use environment, use approach, user irritableness and other user and failure characteristics) of Figure 6.1.

This chapter is organized as follows. The first section (6.1) deals with the measurement of the independent variable FUI. A measurement approach is identified to effectively determine the FUI of the different product functions. Subsequently, this approach is applied to assess the importance of the different TV functions. Eventually, this adopted approach results in a FUI ranking of all the TV functions. This FUI ranking can be applied in the design of the proposed consumer experiment.

The second section (6.2) deals with the measurement of the dependent variable UPFS. Chapter 4 already described some first attempts into the development of a reliable UPFS measurement scale (see Section 4.3.1). The overall reliability of this first self-report UPFS measurement scale was already sufficient (Cronbach's alpha: 0.71). An additional review of the previous experimental results has resulted in some further improvements to this self-report UPFS measurement scale in Section 6.2. Moreover, this self-report UPFS measurement scale is complemented with an existing and more sophisticated measurement scale that is based on the circumplex model of affect (see Chapter 3). Before this additional measurement scale can be applied in the proposed consumer experiment, its reliability and validity should be evaluated. This evaluation process is presented in Section 6.2. Subsequently, the third section (6.3) deals with the control of the extraneous variables.

The previous chapter indicated that there are five categories of extraneous variables that should be controlled for in the proposed consumer experiment:
Section 6.3 illustrates how these extraneous variables can be controlled for in this consumer experiment.

Lastly, in the fourth section (6.4) some conclusions are drawn about the measurement approaches for the different variables of the experimental model of Figure 6.1.

### 6.1 Independent variable: FUI measurement

Before the relation between FUI and UPFS can be evaluated in an actual consumer experiment, an FUI measurement method should be identified. Section 6.1.1 discusses the measurement method to be used based on literature review. In Section 6.1.2 this method is applied to determine the importance of TV functions.

#### 6.1.1 Function Importance measurement methods

Several techniques for the identification of important product functions can be found in literature. Table 6.1 gives an overview of five validated methods for FUI measurement.

Different studies in literature compare the properties and performances of the various FUI measurement techniques [HEE79], [JAC86], [LIN95], [BEC99], [BRE03]. These studies indicate that in general, the “Ratings of functions on a Pre-established List” (RPL) method outperforms the other methods with respect to validity of the results and user-friendliness. Furthermore, the predictive ability of all these methods is approximately equal. Based on the results from these studies, the RPL method is selected for determining FUI in the context of this research.

There are several weight elicitation methods available in literature, to determine ratings given to functions in a decision task. Earlier research demonstrates that straightforward methods work well in the context of subjective preferences [SCH82]. Examples of such uncomplicated methods to give weight to functions are [BOT00], [BOT01]:

- **Max100**: Respondents are asked to give the most important function(s) in the set a rating of 100. All other functions are then rated relatively on a scale of 0-99.
- **Min10**: Respondents are asked to give the least important function a rating of 10. All the other functions are then rated relatively on a scale with no upper-bound
- **Direct Rating (DR) method**: Respondents are asked to rate each function on a scale of 1-100 (or 1-10)
- **Point Allocation (PAL) method**: Respondents are asked to divide a total of 100 points between the given functions

---

- Use environment
- Use approach
- User irritableness
- Other user characteristics
- Other failure characteristics
Table 6.1: FUI measurement methods from literature [HEE79], [JAC86], [LIN95], [BEC99], and [BRE03]

<table>
<thead>
<tr>
<th>FUI measurement methods</th>
<th>Abbrev.</th>
<th>Short Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triadic Sorting using Kelly's repertory grid</td>
<td>TS</td>
<td>This procedure starts by the respondent being shown triple combinations of the product concerned. For each triple combination, the respondent is repeatedly asked for an important function on which two products are alike and at the same time different from the third. This technique is developed with the purpose of mapping cognitive structures, and the added value is that the respondent's own terminology describes the product, as opposed to technocratic terminology.</td>
</tr>
<tr>
<td>Direct Elicitation</td>
<td>DE</td>
<td>The respondent is asked to come up with the functions most important to him/her when choosing among the assortment of products presented. The method can be altered by letting the respondents come up with functions they consider as important and less important when a failure occurs in it. This technique comes closest to a &quot;natural speech&quot; interviewing technique.</td>
</tr>
<tr>
<td>Rank Ordering and Elicitation</td>
<td>ROE</td>
<td>First the respondent is asked to rank order alternatives according to some choice related criterion (e.g. likelihood of buying). Then, the respondent is asked to explain why alternative A is preferred over alternative B, and which product functions influence this preference. The respondent is asked not to repeat functions when explaining their choice.</td>
</tr>
<tr>
<td>Ratings of functions on a Pre-established List</td>
<td>RPL</td>
<td>The respondent is asked to indicate on a scale the importance of the functions provided in a list. This RPL method can be performed directly, by asking respondents to rank function importance by allocating points of a constant sum scale to the various functions. But the RPL method can also be performed indirectly, by letting respondents give the opinion of &quot;other&quot; or &quot;most people&quot;. The difference between a direct and indirect method is that in a direct method people reason from their own perspective and in an indirect method people reason from a third person perspective.</td>
</tr>
<tr>
<td>Information Display Board</td>
<td>IDB</td>
<td>In this method, the respondents choose information piece by piece from a function matrix board until they have acquired enough information to make a product selection. If information on a function is chosen to be displayed, it will be displayed for all alternative products. The information can be requested as many times as wanted, but is only shown one piece at a time.</td>
</tr>
</tbody>
</table>
The Max100 method generates more internally consistent (test-retest reliable) results than the DR method, but these differences are very small [BOT01]. The Max100 and DR method are both considerably more reliable than the PAL and Min10 method. Moreover, test participants prefer using both the Max100 and DR method compared with using the PAL and Min10 method [BOT00], [BOT01]. Therefore, only the Max100 and DR method are considered suitable for application in combination with the RPL technique in a consumer experiment. In this research, the DR method is selected as weight elicitation method for function importance. Respondents are asked to rate all the selected functions on importance on a scale from 0-10. A higher function rating corresponds to a higher importance of that function.

6.1.2 FUI measurement for TV functions

The previous subsection described and evaluated different methods to measure FUI. The RPL method with DR for weight elicitation was selected for the FUI measurement in the context of this research. This subsection describes the application of these methods for the development of a FUI ranking of TV functions. Subsequently, this FUI ranking can be applied in the design of the proposed consumer experiment to investigate the influence of FUI on UPFS. The development of the FUI ranking of TV functions requires the following steps:

- Generation of a list of TV functions
- Development of the function importance survey
- Description of the FUI ranking of the TV functions based on the survey results

TV function list

In the previous chapter (subsection 5.2.1), a function was defined as “the whole of technical attributes that in the eyes of the user is required to execute one task”. A list of TV functions can be drawn from several sources like:

- Websites of TV manufacturing companies, where TV specifications and features are listed that can be translated into general TV functions
- Consumers’ magazines, where products (e.g. TVs) are tested according to some specified criteria, to create an overview of the product performance for consumers
- Web forums, where product users post comments and questions

Based on these TV information sources, a comprehensive list of TV functions was developed based on the above mentioned definition of a function. This list is presented in Table 6.2 on the next page.

The function importance survey

Based on the TV function list of Table 6.2, a FUI survey is developed to determine the TV’s FUI ranking. The importance of the TV functions is measured using the selected RPL method. The survey respondents are asked to rate the TV functions on a scale from 0 till 10 according to the FUI in their purchasing decision. The scores do not have to sum to 10. This means that the first function can be given a rating of 8 and the second function can be given a rating of 5. A high score means that the function is important.

To limit the influence of user characteristics on this FUI ranking, it is decided to select a homogeneous group of participants to complete this FUI survey. Students are a relatively homogeneous group which minimizes variability.
The use of students has the disadvantage that it may not be possible to generalize the results of the study to groups other than university students. But in general, students are considered to be a sound, homogeneous respondent group in consumer (survey) research [PET01]. Moreover, Stangor states that although the use of college students poses some limitations, it must be realized that any sample of research participants is limited in some sense [STA98].

Table 6.2: List of TV functions based on a functionality check of TVs on the market

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch desired program</td>
<td>Choose the program you want and watch it on TV without flaws in the image or sound</td>
</tr>
<tr>
<td>Use of teletext</td>
<td>Use of teletext and all features that belong to it</td>
</tr>
<tr>
<td>Automatic channel installation</td>
<td>Automatic installation of channels to presets without consideration of the preferences of the user</td>
</tr>
<tr>
<td>Manual channel installation</td>
<td>Manual installation of channels to presets that are selected by the user</td>
</tr>
<tr>
<td>Set-up of preferences</td>
<td>Set-up of image, audio and feature preferences and storing the adjusted preferences to presets</td>
</tr>
<tr>
<td>Switch on timer</td>
<td>Use the timer to switch on the TV at a pre-specified time period</td>
</tr>
<tr>
<td>Sleep timer</td>
<td>Use the timer to switch off the TV after a pre-specified time period</td>
</tr>
<tr>
<td>Watch two programs at the same time</td>
<td>Watch two programs at the same time with Picture in Picture (PiP): one program is visible in the main window and the other program in a smaller window in the corner of the screen</td>
</tr>
<tr>
<td>Lock channels</td>
<td>Put a child lock on adult related channels: a code has to be entered to unlock the channel and watch it</td>
</tr>
<tr>
<td>Mute</td>
<td>Mute the sound of the TV</td>
</tr>
<tr>
<td>Change picture format</td>
<td>Select the desired picture format - the different formats are: Auto format, Super zoom, 4:3, 14:9, 16:9, Subtitle zoom, Widescreen</td>
</tr>
<tr>
<td>Previous program</td>
<td>To switch between the currently on and the previously on TV channel</td>
</tr>
<tr>
<td>Use of the motorized swivel</td>
<td>Use the motorized swivel to turn the screen between + and - 30 degrees, to optimize the viewing angle</td>
</tr>
</tbody>
</table>

Some additional questions are added to the function importance survey. The first group of added questions investigates whether the respondents are indeed university students. The second group of questions investigates the possession of a TV with the different TV functions. These questions are raised to investigate whether people who own a TV that possesses a certain function do rate this function differently from people who own a TV that does not possess this function. If a relation appears to exist between function possession and function importance, this should be taken into consideration when investigating the relation between FUI and UPFS.
If function possession influences the function importance evaluation of test participants, the FUI of the involved functions in the experiment should be measured after the participants have actually used these functions. Otherwise, the resulting FUI values are incorrect and consequently the evaluation of the relation between FUI and UPFS is biased. The combination of the FUI measurement and the additional questions resulted in the function importance survey as presented in Appendix D (in Dutch).

As mentioned before, a group of university students is asked to complete the survey. Respondents are selected using a convenience sample of students from the Eindhoven University of Technology. No distinction is made between students from different faculties or educational programs. Students are randomly invited to join this survey at the Auditorium Building of this University. Exchange students are not asked to complete the questionnaire since their frame of reference regarding TVs may be different from Dutch students and this might negatively influence the homogeneity of the sample.

**FUI ranking of the TV functions**
The actual data collection (distribution and collection of the FUI surveys) took place in week 46 and 47 of 2006. A total of 127 surveys was completed. After a first preliminary analysis of the returned surveys, the data of six respondents were removed since they appeared to be non-students. The results of the 121 remaining surveys were analyzed in detail.

The data are analyzed as follows. First, a test for normality is done for the rankings of each function given by the 121 respondents. For normal distributed ranking data, a parametric comparison can be performed among rankings of different TV functions. Otherwise, a non-parametric comparison should be performed. Based on the comparison results, TV functions can be ranked statistically. The detailed analysis results are discussed below.

First, the data was checked on normality. The Kolmogorov-Smirnov test was used to determine if the observed distribution corresponds to a normal distribution (see Appendix E). The analysis shows that all but one scores obtained in the FUI measurement have small significance values (< 0.05) and therefore do not correspond to a normal distribution. The only rating that corresponds to a normal distribution is the rating of the function “Watch two programs” (significance value: 0.083), as can be seen in Table E.1.

Since the FUI data do not correspond to a normal distribution, the ranking of functions on importance cannot be produced by using average ratings and standard deviations of the functions. Instead, the Friedman test has to be used to produce a ranking of functions on importance [CON81]. The Friedman test is designed to determine if scores differ significantly across alternatives. In this case, the Friedman test examines whether the FUI ratings differ significantly between functions. The results of this Friedman ranking test are presented in Table 6.3 on the next page.

The low asymptotic significance value (0.000) indicates that at least one of the functions is rated significantly different from other functions. Kendall's W is a measure that indicates the ranking agreement of the different functions among respondents. Values near 1 indicate a high agreement, and values near 0 indicate little agreement among respondents. In this survey, the value for Kendall's W is 0.493 which indicates moderate agreement among respondents. The function “Watch the desired program” has the highest Friedman rank (1) meaning that on average respondents consider this function as the most important.
### Table 6.3: Ranking of functions based on Friedman test

<table>
<thead>
<tr>
<th>Rank based on Friedman test</th>
<th>Function</th>
<th>Normal Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>Watch the desired program</td>
<td>9.3</td>
</tr>
<tr>
<td>2</td>
<td>Mute</td>
<td>7.1</td>
</tr>
<tr>
<td>3</td>
<td>Manual channel installation</td>
<td>7.0</td>
</tr>
<tr>
<td>4</td>
<td>Set-up preferences</td>
<td>6.9</td>
</tr>
<tr>
<td>5</td>
<td>Teletext</td>
<td>6.1</td>
</tr>
<tr>
<td>6</td>
<td>Previous channel</td>
<td>6.2</td>
</tr>
<tr>
<td>7</td>
<td>Change of picture format</td>
<td>6.3</td>
</tr>
<tr>
<td>8</td>
<td>Automatic channel installation</td>
<td>5.6</td>
</tr>
<tr>
<td>9</td>
<td>Watch two programs at the same time</td>
<td>4.0</td>
</tr>
<tr>
<td>10</td>
<td>Motorized swivel</td>
<td>3.9</td>
</tr>
<tr>
<td>11</td>
<td>Sleep timer</td>
<td>3.1</td>
</tr>
<tr>
<td>12</td>
<td>Child lock</td>
<td>3.1</td>
</tr>
<tr>
<td>13</td>
<td>Switch on timer</td>
<td>2.7</td>
</tr>
</tbody>
</table>

**Friedman Test Statistics**

- N: 121
- Kendall's W: 0.493
- Degrees of freedom: 12
- Asymptotic Significance: 0.000

### Table 6.4: FUI scores split between “in possession” and “not in possession”

<table>
<thead>
<tr>
<th>Function</th>
<th>In possession</th>
<th>Not in possession</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>Average</td>
</tr>
<tr>
<td>Watch the desired program</td>
<td>121</td>
<td>9.3</td>
</tr>
<tr>
<td>Teletext</td>
<td>118</td>
<td>6.1</td>
</tr>
<tr>
<td>Automatic channel installation</td>
<td>97</td>
<td>5.7</td>
</tr>
<tr>
<td>Manual channel installation</td>
<td>111</td>
<td>7.1</td>
</tr>
<tr>
<td>Set-up of preferences</td>
<td>110</td>
<td>7.0</td>
</tr>
<tr>
<td>Switch on timer</td>
<td>27</td>
<td>3.3</td>
</tr>
<tr>
<td>Sleep timer</td>
<td>58</td>
<td>3.3</td>
</tr>
<tr>
<td>Watch two programs</td>
<td>22</td>
<td>3.9</td>
</tr>
<tr>
<td>Child lock</td>
<td>40</td>
<td>2.5</td>
</tr>
<tr>
<td>Mute</td>
<td>111</td>
<td>7.1</td>
</tr>
<tr>
<td>Change picture format</td>
<td>82</td>
<td>6.5</td>
</tr>
<tr>
<td>Previous channel</td>
<td>65</td>
<td>6.9</td>
</tr>
<tr>
<td>Motorized swivel</td>
<td>1</td>
<td>1.0</td>
</tr>
</tbody>
</table>
The TV FUI list in Table 6.3 can be used for the selection of functions for the FUI consumer experiment.

Subsequently, it should be investigated whether people who own a TV that possesses a certain function do rate this function different from people who own a TV that does not possess this function. The “in possession” column of Table 6.4 presents the number of respondents that possess a certain function, the sample mean rating given by these respondents to that function and the standard deviation of that rating. The “not in possession” column presents the number of respondents that do not possess a certain function, the sample mean rating given by these respondents to that function and the standard deviation of that rating. Since these data do not follow a normal distribution, Table 6.4 cannot be used to make a judgment about the influence of function possession on the FUI rating. Instead, a nonparametric test should be selected to check on differences in mean FUI scores for people who own a TV that possesses certain functions and people who own a TV that does not possess those functions. The Mann-Whitney U test is considered most suitable for this purpose [MON99]. Appendix E gives an overview of the details and results of this non-parametric test performed with the function importance survey data.

The results of this test indicate that two functions are rated significantly different by respondents that possess these functions than by respondents that do not possess these functions. The functions “Set-up preferences” and “Previous channel” have small significance values (<0.05), which indicates a significant difference in mean average scores between respondents who own the considered function and respondents who do not own it. For both functions, the mean score that is given by respondents that possess the function is higher than the mean score that is given by respondents that do not possess the function (see Table 6.4). This indicates that, for some functions, people might not appreciate their advantage unless they have tried the function for themselves. The implication of this observation is that the FUI measurement of the functions in this research should be performed after test participants have actually used the involved function. Otherwise, the FUI rating might change significantly during the experiment and thereby bias the measured relation between FUI and UPFS.

Investigating the influence of the independent FUI variable on the dependent UPFS variable in a consumer experiment requires the ability to control the variation of the FUI level and the ability to measure its effect on UPFS. The FUI ranking of TV functions enables the variation of the independent FUI variable by selecting TV functions with different FUI levels. The next section deals with the measurement of the dependent UPFS variable.

6.2 Dependent variable: UPFS measurement

The results of the explorative consumer experiment indicate that the reliability of the self-report UPFS measurement scale was sufficient, but can be improved by adjusting some statements. On the other hand, this self-report UPFS measurement scale can be complemented with an existing and more detailed measurement scale that is based on the circumplex model of affect. Application of this added UPFS measurement could further validate the results of the self-report UPFS measurement. Moreover, this added measurement scale measures the emotional state of participants instead of directly asking participants about their irritation level. This indirect and detailed approach is expected to be more specific on the emotional consequences of the implemented failures for the test participants.

Currently, the material of this subsection is being prepared for publication in an academic journal.
6.2.1 Improvement of the existing UPFS measurement scale

The self-report survey consists of five statements to measure the level of irritation experienced by the user as caused by the TV failure scenario (UPFS) and five added statements to describe the expected reactions of the participant to the experienced failure scenario (see Section 4.2.4). The five irritation statements together form the self-report UPFS measurement scale.

The results of the previous consumer experiment indicate that the overall reliability of this scale is sufficient (Cronbach's alpha 0.71). However, the reliability of the UPFS measurement scale can be improved by adjusting one statement of the UPFS survey. Just by omitting this statement (“The teletext function works without problems”) in the reliability evaluation raises the Cronbach's alpha to 0.79. In other words, the formulation of this statement should better fit the description of the other UPFS statements.

Based on this analysis, the self-report survey is adjusted for the application in this research. The statement “the function works without problems” is replaced by the statement “I would consider the failure in this function very irritating”. In addition to this adjustment, some other survey statements were slightly rephrased in order to achieve an improvement to the overall reliability of the measurement scale. The resulting survey is presented in Appendix F (in Dutch).

6.2.2 PANAS-X for measuring UPFS

UPFS is the level of irritation experienced by the user that is caused by a product failure. Chapter 3 explained the theoretical context of the irritation emotion. It was expounded that the irritation emotion is part of the “anger family” emotions like anger, contempt, frustration and irritation [WRA05]. Irritation can be considered as a weaker form of anger. In the circumplex model of affect, irritation is part of the NA dimension (see Section 3.2).

Moreover, emotion research is concerned with the transitory states that describe the affective condition of people during a specific moment in time. Therefore, predicting the user impact of product quality problems involves predicting the UPFS (transitory state) to these problems given certain predetermined user characteristics (traits). Consequently, a more sophisticated UPFS measurement scale should be able to measure the difference in the (transitory) UPFS state of a user that is caused by a product failure. This measurement scale can be based on the NA dimension in the circumplex model.

PANAS-X

Based on the circumplex model of affect, Watson et al. developed the so-called Positive And Negative Affect Schedule (PANAS). However, this 20 item emotion measurement schedule failed to provide a close fit with the emotion data obtained from the field [WAT99]. Watson and Clark developed PANAS further, trying to resolve this problem. The resulting PANAS eXpanded form (PANAS-X), was developed in 1994 and is a 60-item questionnaire. These 60 items are one-word items which represent an emotional state like happy, sad, angry, etcetera. For every item, participants have to indicate with a score from 1 to 5 to what extent the particular item is applicable to him/her. PANAS-X measures the general dimensions NA and PA, but also measures 11 separate affects. An overview of the item composition of PANAS-X is presented in Table 6.5 on the next page. Several scientific studies demonstrate that the PANAS-X scales are stable over time, show significant convergent and discriminant validity, are correlated with corresponding measures of state affect (short term and long term) and are strongly and systematically related to measures of personality and emotionality [WAT99],...
However, one exception holds for the Surprise-scale when measuring long term state affect. The validity of this scale was not as good as the other scales. Based on these results, it is not recommended to use the Surprise-scale.

According to Watson and Clark, the PANAS-X scales are sensitive to dynamic circumstances in a small time period [WAT99]. In other words, PANAS-X is capable of measuring transitory states that describe the affective condition of people during a specific moment in time. Moreover, several emotion studies demonstrate this capability in practice [MCI90], [CLA89], [HOF06].

**Table 6.5: PANAS-X item composition [WAT99A]**

<table>
<thead>
<tr>
<th>General Dimension Scales</th>
<th>Negative Affect</th>
<th>Positive Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>afraid, scared, nervous, jittery, irritable, hostile, guilty, ashamed, upset, distressed</td>
<td>active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud, strong</td>
</tr>
<tr>
<td>Basic Negative Emotion Scales</td>
<td>Fear</td>
<td>Hostility</td>
</tr>
<tr>
<td></td>
<td>afraid, scared, frightened, nervous, jittery, shaky</td>
<td>angry, hostile, irritable, scornful, disgusted, loathing</td>
</tr>
<tr>
<td>Sadness</td>
<td>sad, blue, downhearted, alone, lonely</td>
<td></td>
</tr>
<tr>
<td>Basic Positive Emotion Scales</td>
<td>Joviality</td>
<td>Self-Assurance</td>
</tr>
<tr>
<td></td>
<td>happy, joyful, delighted, cheerful, excited, enthusiastic, lively, energetic</td>
<td>proud, strong, confident, bold, daring, fearless</td>
</tr>
<tr>
<td>Other Affective States</td>
<td>Shyness</td>
<td>Fatigue</td>
</tr>
<tr>
<td></td>
<td>shy, bashful, sheepish, timid</td>
<td>sleepy, tired, sluggish, drowsy</td>
</tr>
</tbody>
</table>

Furthermore, the PANAS-X scale is simple and easy to use. Most subjects complete the entire 60-item schedule in 10 minutes or less. In case of strict time constraints, it is also possible to select those scales that are most relevant within the context of the research [WAT99A].

In general, a failure is perceived negatively by the user. Persons who are externally oriented look for the causality of the failure externally and therefore experience more anger, irritation and frustration. Internal oriented persons search for failure causality internally and experience more embarrassment, shame and guilt [WEI85], [LAZ91], [SCH01]. Anger, irritation, shame and guilt are all items of the General Dimension Scale of Negative Affect. Therefore, this General Dimension Scale should be used for measuring UPFS.

The PANAS-X scale also has four negative basic emotion scales: Fear, Hostility, Guilt and Sadness. As mentioned before, the irritation emotion is part of the “anger family” consisting of anger, irritation, contempt and frustration [WRA05].
The typical instigation to anger is a value judgment. More than anything else, anger is an attribution of blame, directed to others [AVE83]. Anger involves an attribution of responsibility, an accusation, that the target has done something wrong. It follows that the target of anger must be a person or object to whom responsibility can be assigned [AVE82]. In the context of this research, the anger of the user is probably directed to the television. The television “is responsible for” the failure. Anger occurs in negative, unexpected and important failure situations when causality is directed externally [WEI85]. The basic negative emotion scale Hostility of PANAS-X reflects the anger emotion very well. Firstly, since anger and irritation are both items of the Hostility scale. Secondly, the items hostile, scornful, disgusted and loathing are emotions which point to external factors. These items emphasize that anger is directed to externals. Altogether, the basic negative emotion scale hostility should be used as well for measuring UPFS in this research.

The basic negative emotion scale Guilt is related to internal causal attribution. Guilt is associated with blaming oneself and is a negatively experienced emotion that typically arises in response to personal failure [SMI93], [WEI00], [ROS01], [SCH01]. Guilt can be experienced by users in this research. There are probably participants who blame themselves for the incorrect functioning of the TV. Therefore, the subscale Guilt is also used in the UPFS measurement in this research. Fear can be associated with evaluating a situation as threatening [SMI93]. Weiner et al. demonstrated that in failure situations, the most reported emotions were anger, depression, fear and frustration [WEI79]. So, the user perceived failure severity may also be expressed in fear. Consequently, the basic Negative Emotion Scale Fear is also included in the UPFS measurement.

Sadness is associated with helplessness in an undesirable situation where there is little or no hope of improvement [SMI93]. Persons who are internally oriented look for causality of a failure internally and experience more embarrassment, shame and guilt [WEI85], [LAZ91], [SCH01]. People who search for internal causes of negative events will do more self-blaming and experience more negative emotions in general or more depression and sadness [WRA05]. In other words, also the subscale Sadness should be used in measuring UPFS. Since the basic emotion scales Fear, Hostility, Guilt and Sadness form the Basic Negative Emotion scale, this complete scale is included in the UPFS measurement. Only using the scales related to Negative Affect for the UPFS measurement would lead to response acquiescence – a tendency to agree with statements [GOO05]. Therefore, for the measurement of UPFS, the Positive Affect Scales and Other Affective Scales are incorporated too.

**UPFS measurement using PANAS-X**

The above findings indicate that all the PANAS-X measurement scales should be used to measure the UPFS of a product failure. As mentioned above, predicting the user impact of product quality problems involves predicting the UPFS (transitory state) that is caused by these problems given certain predetermined user characteristics (traits). In other words, a product failure causes an emotional state transition within the user which should be measured with the PANAS-X Negative Affect scales. The transition from the user's emotional NA state just before the product failure into the user's emotional NA state directly after the product failure can be considered as a measure for UPFS. This UPFS measurement approach is presented in Figure 6.2.
A practical limitation to the application of PANAS-X in this research is the fact that the PANAS-X scheme is not (completely) available in Dutch. Although the general dimension scales have already been translated and validated in earlier research [HIL05], the basic emotion scales and other affective states have not been translated yet. Therefore, these scales and states were translated for this purpose by a group of five researchers. This translation process was done on consensual agreement and resulted in the PANAS-X measurement scale as presented in Appendix G.

Obviously, this new translation of PANAS-X requires a validation of the complete measurement scale before application in the proposed consumer experiment. This validation of the translated PANAS-X scheme is presented next.

### 6.2.3 Reliability and validity of PANAS-X

In this subsection, the reliability and validity of the new PANAS-X translation is evaluated. First, the data collection approach is explained. Subsequently, with the collected data an examination is done with regard to missing values. A considerable lot of missing values can affect the generalizability of the results [HAI05]. Further, a comparison is made between the PANAS-X results of this research and the PANAS-X results of an earlier study performed with the existing incomplete Dutch translation of PANAS-X, the so-called Maastricht Aging Study [HIL05]. With this analysis, a part of the translated PANAS-X can be validated. Subsequently, this newly translated part is validated with a factor analysis. Lastly, the reliability and validity of the PANAS-X subscales are checked.

#### Data collection

The actual data collection (distribution and collection of the PANAS-X surveys) took place in week 46 and 47 of 2006. Potential participants were approached and asked to fill in the PANAS-X survey (see Appendix G). This approach resulted in 207 returned PANAS-X surveys. Seven of these surveys appeared to be incomplete (more than 10 missing values) and were therefore excluded from the analysis. This means that in total 200 students successfully completed the PANAS-X survey.

#### Data examination

In this subsection, the data (N=200) are examined with regard to missing values. Too much missing data can affect the generalizability of the results [HAI05].
The sample size available for analysis is reduced and statistical results based on a nonrandom missing data process could be biased [HAI05]. A descriptive analysis of the data showed that from the 200 cases, seven cases had one missing item and one case had two missing items. So, in total nine values of the 12 000 values (200 * 60 values) were missing (< 1 %). It is remarkable that eight of these missing values are missed in the affective state Shyness. Four times a value was missed at Bashful. This could be an indication that the affective state Shyness is not translated very well. Apparently, Dutch students do not understand the Dutch translation of Sheepish, Timid and Bashful. These missing values are replaced. When the missing data are under 10%, any imputation method can be applied [HAI05]. The Mean Value Regression method is used as imputation method which predicts the missing values of a variable based on its relationship with other variables in the data set [HAI05]. In other words, the missing values are “replaced” with the predicted values and a complete dataset of 200 cases is used for the reliability and validity analysis.

**Comparison with Maastricht aging study**

PANAS-X consists of two higher order scales, the so-called General Dimension Scales: Positive and Negative Affect (see Table 6.5). These two scales consist each of ten items, which together form PANAS. PANAS can be seen as the shorter version of PANAS-X. PANAS consists only of these two scales and has in total 20 items. PANAS-X consists of higher order scales and 11 affective states. The Dutch translation of the two scales Positive and Negative Affect of PANAS is validated by Hill et al. [HIL05]. The results of this analysis are compared with the data collected in this research. The analysis conducted by Hill et al. is referred to as the “Maastricht Aging Study”. In this section, the analysis of the data of this research is called the “Failure Severity Study”.

The analysis of the Failure Severity Study is done exactly the same as in the Maastricht Aging Study. Only the 20 items which are relevant for the general dimensions PA and NA are kept in the database of the Failure Severity Study for this analysis. Two factors are extracted out of this dataset in order to get the NA “factor” and the PA “factor”. The analysis method and rotation method used is the same as used in the Maastricht Aging Study: principal component analysis and varimax rotation. The appropriateness of these analysis and rotation methods is discussed further on. Now the focus is on comparing the results of the Failure Severity Study and the Maastricht Aging Study. This comparison can demonstrate that the 20 items of PANAS (more specifically 20 items of the 60 items of PANAS-X) are used in a reliable and valid way. The results of these factor analyses for both studies are presented in Table 6.6.

Both analyses produced an underlying factor structure which is in line with the factor structure described by Watson et al., namely a positive affect factor and a negative affect factor that together encompassed all 20 items [WAT88]. A visual inspection of Table 6.6 shows that there is a large difference (compared with other items) in the mean score of the items Guilty, Proud and Excited. It has to be noted that the mean age of the participants of the Maastricht Aging Study is 62.67 years where in the Failure Severity Study the mean age is around 22\(^\circ\). Apparently, these younger people (students) feel more guilt than older people. Students are also more proud and excited than older people. The two factors explain in both studies nearly the same percentage of variance (around 38%). The internal consistencies of the factors in both studies are around 0.80 (Cronbach’s alpha).

\(^\circ\)Not all participants filled in their year of birth (N=112)
However, it has to be noted that Cronbach’s alpha of the PA factor is lower in the Failure Severity Study but is still above the minimum recommended level of 0.70. Besides these differences, the results are nearly the same for both studies. These results confirm that the 20 items forming PANAS are reliable and can be used for further analysis of irritableness and UPFS.

**Factor analysis**

In addition to the two higher order scales, there are also 11 specific affects within PANAS-X (not within PANAS). These 11 specific affects can be categorized in three groups: Basic Negative Emotions, Basic Positive Emotions and Other Affective States (see Table 6.5). The scales Fear, Hostility, Guilt and Sadness together form the Basic Negative Emotion Scales. These scales are also related to the General Dimension scale Negative Affect. The same reasoning can be applied to the Basic Positive Emotion Scales. However, it is not clear whether the states Shyness, Fatigue, Serenity and Surprise are related to Negative or Positive Affect (or a combination of Negative and Positive Affect).

**Table 6.6: Principal components analysis of PANAS items with varimax rotation**

<table>
<thead>
<tr>
<th>Item</th>
<th>Failure Severity Study</th>
<th>Maastricht Aging Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean St. dev. Factor 1 Factor 2 Mean St. dev. Factor 1 Factor 2</td>
<td>Mean St. dev. Factor 1 Factor 2</td>
</tr>
<tr>
<td>Afraid</td>
<td>1.50 0.716 0.72 1.73 1.07 0.62</td>
<td>Ashamed 1.80 0.835 0.64 1.47 0.82 0.51</td>
</tr>
<tr>
<td>Scared</td>
<td>1.51 0.736 0.71 1.63 0.95 0.67</td>
<td>Irritable 2.25 0.975 0.60 2.27 1.08 0.62</td>
</tr>
<tr>
<td>Distressed</td>
<td>1.43 0.661 0.71 1.76 1.14 0.73</td>
<td>Hostile 1.60 0.897 0.59 1.34 0.74 0.36</td>
</tr>
<tr>
<td>Jittery</td>
<td>2.03 0.847 0.70 2.36 1.30 0.68</td>
<td>Guilty 2.79 1.120 0.40 1.51 0.90 0.55</td>
</tr>
<tr>
<td>Nervous</td>
<td>2.04 0.841 0.69 2.44 1.23 0.75</td>
<td>Enthusiastic 3.57 0.767 0.69 3.24 1.05 0.71</td>
</tr>
<tr>
<td>Upset</td>
<td>1.63 0.904 0.65 1.86 1.11 0.78</td>
<td>Active 3.54 0.929 0.67 3.73 1.08 0.66</td>
</tr>
<tr>
<td>Ashamed</td>
<td>1.80 0.835 0.64 1.47 0.82 0.51</td>
<td>Alert 3.30 0.833 0.65 3.32 1.11 0.54</td>
</tr>
<tr>
<td>Irritable</td>
<td>2.25 0.975 0.60 2.27 1.08 0.62</td>
<td>Determined 3.46 0.861 0.61 3.40 1.08 0.61</td>
</tr>
<tr>
<td>Hostile</td>
<td>1.60 0.897 0.59 1.34 0.74 0.36</td>
<td>Interested 3.68 0.728 0.60 3.90 0.91 0.55</td>
</tr>
<tr>
<td>Guilty</td>
<td>2.79 1.120 0.40 1.51 0.90 0.55</td>
<td>Inspired 3.35 0.811 0.60 3.06 1.10 0.70</td>
</tr>
<tr>
<td>Strong</td>
<td>3.22 0.902 0.39 3.11 1.12 0.64</td>
<td>Strong 3.22 0.902 0.39 3.11 1.12 0.64</td>
</tr>
</tbody>
</table>

**Eigenvalues**

- Failure Severity Study: 4.37, 3.12, 4.12, 3.58
- Maastricht Aging Study: 4.12, 3.58

**% of variance explained**

- Failure Severity Study: 21.87, 15.62, 20.58, 17.89
- Maastricht Aging Study: 21.87, 37.49, 20.58, 38.46

**Cronbach’s alpha**

- Failure Severity Study: 0.83, 0.75, 0.80, 0.84
- Maastricht Aging Study: 0.83, 0.75, 0.80, 0.84

* Loadings < [0.30] are not shown
All these 11 specific affect states are composed of a number of items, see Table 6.5. For example, the specific affect state Fear consists of the following items: afraid, scared, frightened, nervous, jittery and shaky. All together, PANAS-X is made up of 60 items. These 60 items together form 11 specific affects and the two general Positive and Negative Affects. The score on the General Dimension Scale Negative Affect (or Positive Affect) is calculated by summing the scores of the 10 accompanying items. The 11 specific affect states are also calculated by adding the scores of the accompanying items. For the Basic Negative Emotion scale, the scores of the specific affect scales Fear, Hostility, Guilt and Sadness are summed up and consequently divided by four. For the Basic Positive Emotion scale, the scores of the specific affect scales Joviality, Self-assurance and Attentiveness are added and consequently divided by three. PANAS-X can be described as “a hierarchical taxonomic scheme in which two broad, higher order dimensions are each composed of several correlated, yet ultimately distinguishable states” [WAT99]. With factor analysis, it is assessed whether the proposed structure also exists within the Dutch version of PANAS-X. With this factor analysis, the two higher order dimensions PA and NA are assessed. Also, the 11 specific scales are examined. First however, the assumptions underlying factor analysis are checked.

**Assumptions underlying factor analysis**

The critical assumptions underlying factor analysis are more conceptual than statistical [HAI05]. A basic conceptual assumption underlying factor analysis is that some underlying structure exists in the set of selected variables [HAI05]. The existence of this underlying structure is widely proven by Watson, Clark en Tellegen [WAT99]. A statistical issue is to ensure that the variables are sufficiently correlated to produce representative factors [HAI05]. When all of the correlations are low or equal, it is questionable whether factor analysis can be applied. For testing the applicability of factor analysis, two methods are widely used: Bartlett test of sphericity and the Measure of Sampling Adequacy (MSA) [HAI05]. A Bartlett test of sphericity tests whether there is sufficient correlation among the variables. The MSA quantifies the degree of correlation among the variables and the appropriateness of factor analysis. This index ranges from 0 till 1, reaching 1 when each variable is perfectly predicted without error by other variables. The MSA values must exceed 0.5 for both the overall test and each individual variable in order to show a sufficient degree of correlation among the variables [HAI05].

Table 6.7 indicates that the overall MSA test shows a value of 0.812, which is meritorious and largely exceeds 0.50 [HAI05]. Also the Bartlett test of sphericity is highly significant (see Table 6.7). Based on this test, it can be concluded that there are significant correlations among the variables. The MSA values of the individual items all pass the criteria of 0.5, see Table 6.8. Each item seems to fit within the structure of the other items. Overall, there is a sufficient degree of correlation between variables, which indicates that a factor analysis may be useful for this data.

**Table 6.7: MSA and Bartlett test**

<table>
<thead>
<tr>
<th>Measure of Sampling Adequacy (MSA)</th>
<th>0.812</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett's test of sphericity</td>
<td></td>
</tr>
<tr>
<td>Chi-Square</td>
<td>5480.413</td>
</tr>
<tr>
<td>df</td>
<td>1770</td>
</tr>
<tr>
<td>Significance</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Performing factor analysis

The results of the previous subsection show that it is useful to do a factor analysis for identification of the underlying structures in the dataset. PANAS-X consists of two broad higher dimensions: Negative Affect and Positive Affect. With this factor analysis, these two higher order dimensions are examined. This analysis is done for assessing the general structure of PANAS-X. The goal is to extract two factors: Negative Affect and Positive Affect. This means that all items which are related to Negative Affect should load on one factor. These are the items of the General Dimension Scale Negative Affect and the items of the Basic Negative Emotion Scale. All items which are related to Positive Affect should load on the other factor. For the items related to other affective states, no clear pattern is expected. These items can load on any factor.

There are two methods for doing factor analysis: common factor analysis and component analysis. Component analysis is especially suitable for summarizing most of the original information in a minimum number of factors for prediction purposes [HAI05]. Common factor analysis is used primarily to identify underlying factors or dimensions that reflect what the variables share in common [HAI05]. Common factor analysis has two problems. With common factor analysis, it is possible that several different factor scores can be calculated from a single factor model. No single unique solution is found, as in component analysis. The second issue is that sometimes communalities are not estimable or may be invalid, requiring

A factor score is a composite measure created for each observation on each factor extracted in the factor analysis. The factor weights are used in conjunction with the original variable value to calculate each observation’s score. The factor score can be used to represent the factor(s) in subsequent analysis. Factor scores are standardized to have a mean of 0 and a standard deviation of 1 [HAI05].

---

**Table 6.8: MSA of individual items**

<table>
<thead>
<tr>
<th>Item</th>
<th>MSA</th>
<th>MSA</th>
<th>MSA</th>
<th>MSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bang</td>
<td>0.840</td>
<td>Enthousiast</td>
<td>0.825</td>
<td>Boos op jezelf</td>
</tr>
<tr>
<td>Angstig</td>
<td>0.853</td>
<td>Levdend</td>
<td>0.848</td>
<td>Walgend van jezelf</td>
</tr>
<tr>
<td>Bevreeds</td>
<td>0.843</td>
<td>Energiëk</td>
<td>0.840</td>
<td>Ontevreden met jezelf</td>
</tr>
<tr>
<td>Nerveus</td>
<td>0.882</td>
<td>Trots</td>
<td>0.729</td>
<td>Verdrietig</td>
</tr>
<tr>
<td>Zenuwachtig</td>
<td>0.900</td>
<td>Sterk</td>
<td>0.669</td>
<td>Zwaarmoedig</td>
</tr>
<tr>
<td>Rusteloos</td>
<td>0.790</td>
<td>Zelfverzeker</td>
<td>0.677</td>
<td>Teneergeslagen</td>
</tr>
<tr>
<td>Boos</td>
<td>0.909</td>
<td>Dapper</td>
<td>0.717</td>
<td>Alleen</td>
</tr>
<tr>
<td>Vijandig</td>
<td>0.811</td>
<td>Moedig</td>
<td>0.756</td>
<td>Eenzaam</td>
</tr>
<tr>
<td>Prikkelbaar</td>
<td>0.840</td>
<td>Onbevreesd</td>
<td>0.691</td>
<td>Overstuur</td>
</tr>
<tr>
<td>Minachtend</td>
<td>0.729</td>
<td>Alert</td>
<td>0.684</td>
<td>Van streek</td>
</tr>
<tr>
<td>Walgen van</td>
<td>0.790</td>
<td>Opletterend</td>
<td>0.589</td>
<td>Gelukkig</td>
</tr>
<tr>
<td>Afkeer hebben van</td>
<td>0.817</td>
<td>Geconcentreerd</td>
<td>0.768</td>
<td>Opgewekt</td>
</tr>
<tr>
<td>Schuldewust</td>
<td>0.737</td>
<td>Aandachtig</td>
<td>0.591</td>
<td>Opgetogen</td>
</tr>
<tr>
<td>Beschaamd</td>
<td>0.897</td>
<td>Actief</td>
<td>0.862</td>
<td>Vrolijk</td>
</tr>
<tr>
<td>Schuldig</td>
<td>0.830</td>
<td>Geënspireerd</td>
<td>0.733</td>
<td>Uitgelaten</td>
</tr>
</tbody>
</table>

* The results presented in this section are language dependent. The purpose of this section is to validate the Dutch translation of PANAS-X. Therefore, the tables in this chapter only present the Dutch PANAS-X items. Direct translation between the Dutch and English PANAS-X items is not possible. However, for the non-Dutch reader, an indicative English translation of the Dutch PANAS-X items is given in the textual explanations of the tables.
the deletion of variables from the analysis [HAI05]. Communality is the total amount of variance an original variable shares with all other variables included. There is some debate over which factor model is more appropriate. However, empirical research shows that similar results are obtained in many instances when using component or common factor analysis especially if the number of variables exceeds 30 [HAI05]. Subsequently, the factor solution is rotated for improving the interpretation of the results. The goal of the rotation is to derive theoretically meaningful factors and, if possible, the simplest factor structure [HAI05]. The objective of all rotation methods is to simplify the rows and columns of the factor matrix to increase ease of interpretation. In a factor matrix, columns represent factors with each row corresponding to a variable's loadings across the factors. A variable's loading on a single factor is maximized. By simplifying the columns, as many variables in each column are made as close to zero as possible, making the number of high loadings as few as possible.

There are two rotation methods: orthogonal and oblique. In orthogonal factor rotation, the axes are maintained at 90 degrees. Within oblique rotation, the axes do not have to be retained at 90 degrees. There are no specific rules for preferring one factor method over another [HAI05]. Orthogonal rotation methods are the most widely used rotation methods among researchers. The three most known orthogonal rotation methods are Quartimax, Varimax and Equimax. The most popular orthogonal rotation method in literature is Varimax [CON03]. Therefore, the Varimax rotation method is used. Varimax tries to derive high loadings of -1 or +1 in each column of the matrix, thus indicating a clear positive or negative association between variable and factor. A principal component factor analysis with Varimax rotation was performed. The results are presented in Table 6.9. The mean and the standard deviation are given for each item. Also, the factor pattern matrix is given. The factor pattern matrix contains the factor loading of each variable on each component.

The results show that the items which are related to the General Dimension Scale Negative Affect and Basic Negative Emotion Scale load on one factor. Only the item “eenzaam” (lonely) loads also negatively on the second factor. The items which are related to the General Dimension Scale Positive Affect and the Basic Positive Emotion Scale load on the other factor. One exception is the item “oplettend” (alert). This item does not load on the second factor (and also not on factor one). For the items related to the other affective states, no clear pattern is visible, as expected. These results show that there are two general dimensions. The items “lui”, “kalm” and “ontspannen” (sluggish, calm and relaxed respectively) do not load on either component 1 or component 2. These three items are part of the affective states Fatigue and Serenity. So based on these results, the items which should measure NA (or PA) also do measure NA (or PA).

The overall structure of the two broad higher dimensions is confirmed. In this analysis, it also became visible that most items load on the high pole ends (the activated ends). This is in line with Watson et al. who claim that the high pole ends are much more important than the low pole ends (the inactivated ends) [WAT99A]. All important instruments that assess mood contain many more high-activation terms (located at high pole ends) than low activation terms [WAT99A]. Also, the high end pole ends define the dimensions. The low poles of these dimensions present the absence of a particular activation [WAT99A]. The next step is the evaluation of the lower dimension scales of PANAS-X. For assessing the lower dimension scales, a range of factor solutions was examined, until a solution was reached that contained an interpretable and clear factor structure. This means that each solution contained a different number of extracted factors. A solution with six extracted factors gives the best interpretable result. In this context, interpretable means a clear factor solution: an item loads on one factor.
Table 6.9: Principal components analysis of PANAS-X items with varimax rotation

<table>
<thead>
<tr>
<th>Mean</th>
<th>St. dev.</th>
<th>Component 1</th>
<th>Component 2</th>
<th>Mean</th>
<th>St. dev.</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bang</td>
<td>1.50</td>
<td>0.716</td>
<td>0.623</td>
<td>Enthousiast</td>
<td>3.57</td>
<td>0.767</td>
<td>0.678</td>
</tr>
<tr>
<td>Angstig</td>
<td>1.51</td>
<td>0.736</td>
<td>0.650</td>
<td>Levendig</td>
<td>3.59</td>
<td>0.778</td>
<td>0.654</td>
</tr>
<tr>
<td>Bevreed</td>
<td>1.64</td>
<td>0.840</td>
<td>0.660</td>
<td>Energiek</td>
<td>3.56</td>
<td>0.819</td>
<td>0.747</td>
</tr>
<tr>
<td>Nerveus</td>
<td>2.04</td>
<td>0.841</td>
<td>0.631</td>
<td>Trots</td>
<td>3.22</td>
<td>0.856</td>
<td>0.416</td>
</tr>
<tr>
<td>Zenuwachtig</td>
<td>2.03</td>
<td>0.847</td>
<td>0.647</td>
<td>Sterk</td>
<td>3.22</td>
<td>0.902</td>
<td>0.413</td>
</tr>
<tr>
<td>Rusteloos</td>
<td>1.81</td>
<td>1.016</td>
<td>0.480</td>
<td>Zelfverzeker</td>
<td>3.43</td>
<td>0.818</td>
<td>0.456</td>
</tr>
<tr>
<td>Boos</td>
<td>1.69</td>
<td>0.841</td>
<td>0.613</td>
<td>Dapper</td>
<td>2.85</td>
<td>0.950</td>
<td>0.490</td>
</tr>
<tr>
<td>Vaijandig</td>
<td>1.60</td>
<td>0.897</td>
<td>0.549</td>
<td>Moedig</td>
<td>3.12</td>
<td>0.903</td>
<td>0.541</td>
</tr>
<tr>
<td>Prikkelaar</td>
<td>2.25</td>
<td>0.975</td>
<td>0.577</td>
<td>Onbevreeds</td>
<td>2.95</td>
<td>0.955</td>
<td>0.421</td>
</tr>
<tr>
<td>Minachtend</td>
<td>1.73</td>
<td>0.966</td>
<td>0.376</td>
<td>Alert</td>
<td>3.30</td>
<td>0.833</td>
<td>0.533</td>
</tr>
<tr>
<td>Walgen van</td>
<td>1.50</td>
<td>0.783</td>
<td>0.495</td>
<td>Oplettend</td>
<td>3.42</td>
<td>0.810</td>
<td></td>
</tr>
<tr>
<td>Alkeer hebben van</td>
<td>1.84</td>
<td>0.934</td>
<td>0.545</td>
<td>Geconcentreerd</td>
<td>3.24</td>
<td>0.810</td>
<td>0.489</td>
</tr>
<tr>
<td>Schuldbewust</td>
<td>2.79</td>
<td>1.120</td>
<td>0.344</td>
<td>Aandachtig</td>
<td>3.46</td>
<td>0.861</td>
<td>0.374</td>
</tr>
<tr>
<td>Beschaamd</td>
<td>1.80</td>
<td>0.835</td>
<td>0.599</td>
<td>Actief</td>
<td>3.54</td>
<td>0.929</td>
<td>0.631</td>
</tr>
<tr>
<td>Schuldig</td>
<td>1.63</td>
<td>0.798</td>
<td>0.613</td>
<td>Geïnspireerd</td>
<td>3.25</td>
<td>0.811</td>
<td>0.506</td>
</tr>
<tr>
<td>Boos op jezelf</td>
<td>1.59</td>
<td>0.870</td>
<td>0.615</td>
<td>Geïnteresseerd</td>
<td>3.68</td>
<td>0.728</td>
<td>0.408</td>
</tr>
<tr>
<td>Walgend van jezelf</td>
<td>1.32</td>
<td>0.656</td>
<td>0.626</td>
<td>Verlegen</td>
<td>2.36</td>
<td>0.941</td>
<td>0.352</td>
</tr>
<tr>
<td>Ontevreden met jezelf</td>
<td>1.64</td>
<td>0.846</td>
<td>0.526</td>
<td>Beduusd</td>
<td>1.86</td>
<td>0.807</td>
<td>0.376</td>
</tr>
<tr>
<td>Verdrigig</td>
<td>1.58</td>
<td>0.810</td>
<td>0.601</td>
<td>Bedeesd</td>
<td>1.87</td>
<td>0.814</td>
<td>0.412</td>
</tr>
<tr>
<td>Zwaarmoeheid</td>
<td>1.74</td>
<td>0.816</td>
<td>0.566</td>
<td>Timide</td>
<td>2.28</td>
<td>1.042</td>
<td>0.356</td>
</tr>
<tr>
<td>Teneergeslagen</td>
<td>1.52</td>
<td>0.694</td>
<td>0.593</td>
<td>Slaperig</td>
<td>2.48</td>
<td>1.080</td>
<td>0.505</td>
</tr>
<tr>
<td>Alleen</td>
<td>1.78</td>
<td>0.869</td>
<td>0.560</td>
<td>Moe</td>
<td>2.67</td>
<td>1.080</td>
<td>0.506</td>
</tr>
<tr>
<td>Eenzaam</td>
<td>1.59</td>
<td>0.816</td>
<td>0.590 -0.343</td>
<td>Lui</td>
<td>2.53</td>
<td>1.093</td>
<td></td>
</tr>
<tr>
<td>Overstuur</td>
<td>1.63</td>
<td>0.904</td>
<td>0.599</td>
<td>Suf</td>
<td>1.93</td>
<td>0.961</td>
<td>0.474</td>
</tr>
<tr>
<td>Van streek</td>
<td>1.43</td>
<td>0.661</td>
<td>0.715</td>
<td>Kalm</td>
<td>3.47</td>
<td>0.924</td>
<td></td>
</tr>
<tr>
<td>Geluik</td>
<td>3.92</td>
<td>0.675</td>
<td>0.510</td>
<td>Ontspannen</td>
<td>3.51</td>
<td>0.845</td>
<td></td>
</tr>
<tr>
<td>Opgewekt</td>
<td>3.68</td>
<td>0.678</td>
<td>0.575</td>
<td>Op je gemak</td>
<td>3.73</td>
<td>0.692</td>
<td>0.345</td>
</tr>
<tr>
<td>Opgetogen</td>
<td>3.36</td>
<td>0.789</td>
<td>0.451</td>
<td>Verwonderd</td>
<td>2.25</td>
<td>0.971</td>
<td>0.467</td>
</tr>
<tr>
<td>Vrolijk</td>
<td>3.82</td>
<td>0.658</td>
<td>0.533</td>
<td>Verbaasd</td>
<td>2.15</td>
<td>0.893</td>
<td>0.417</td>
</tr>
<tr>
<td>Uitgelaten</td>
<td>3.92</td>
<td>0.893</td>
<td>0.437</td>
<td>Versteld staan</td>
<td>2.20</td>
<td>0.939</td>
<td>0.529</td>
</tr>
</tbody>
</table>

This solution is presented in Table 6.10. Also for this factor analysis, a component factor analysis and a varimax rotation were used.

The discussion starts with the results of the second factor. The first factor is addressed later on. The items which positively load on the second factor are “energiek”, “opgewekt”, “levendig”, “enthousiast”, “geluik”, “uitgelaten” and “actief” (energetic, joyful, cheerful, lively, enthusiastic, happy, excited, and active). These items are all related to PA. With the exception of the item “actief”, all these items are part of the basic emotion scale Joviality. Therefore, the second factor could be seen as the basic emotion Joviality.

* Loadings < |0.30| are not shown
Table 6.10: Principal components analysis of PANAS-X items with varimax rotation

<table>
<thead>
<tr>
<th>Items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van streek</td>
<td>0.705</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walgend van jezelf</td>
<td>0.686</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boos op jezelf</td>
<td>0.650</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bevreest</td>
<td>0.648</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schuldig</td>
<td>0.626</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beschamad</td>
<td>0.604</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zenuwachtig</td>
<td>0.591</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boos op jezelf</td>
<td>0.585</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teneergeslagen</td>
<td>0.576</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overstuur</td>
<td>0.571</td>
<td>-0.455</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertrouwd</td>
<td>0.567</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eenzamen</td>
<td>0.566</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nerveus</td>
<td>0.563</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>angstig</td>
<td>0.561</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ontevreden met jezelf</td>
<td>0.557</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prikkelbaar</td>
<td>0.554</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Versteld staan</td>
<td>0.549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afkeer hebben van</td>
<td>0.543</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zwaarmoedig</td>
<td>0.540</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walgen van</td>
<td>0.537</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bang</td>
<td>0.527</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alleen</td>
<td>0.524</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vijandig</td>
<td>0.519</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rusteloos</td>
<td>0.458</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbaasd</td>
<td>0.401</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verwonderd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedeezd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schuldbewust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minachtend</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beduusd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energiek</td>
<td>0.685</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opgehekt</td>
<td>0.670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vrolijk</td>
<td>0.631</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levendig</td>
<td>0.620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enthousiast</td>
<td>0.611</td>
<td>0.441</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gelukkig</td>
<td>0.501</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uitgeladen</td>
<td>0.440</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actief</td>
<td>0.405</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opgetogen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Op je gemak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aandachtig</td>
<td></td>
<td>0.710</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opletend</td>
<td></td>
<td>0.609</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alert</td>
<td></td>
<td>0.560</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geintereseeerd</td>
<td></td>
<td>0.560</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geconcentreerd</td>
<td></td>
<td>0.520</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geinspireerd</td>
<td></td>
<td>0.427</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trot</td>
<td></td>
<td></td>
<td>0.722</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moedig</td>
<td></td>
<td></td>
<td>0.645</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dapper</td>
<td></td>
<td></td>
<td>0.625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sterk</td>
<td></td>
<td></td>
<td>0.624</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onbevreest</td>
<td></td>
<td></td>
<td></td>
<td>0.624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zelfverzekerd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.762</td>
<td></td>
</tr>
<tr>
<td>Moe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.726</td>
</tr>
<tr>
<td>Slaperig</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.638</td>
</tr>
<tr>
<td>Lui</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.598</td>
</tr>
<tr>
<td>Suf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.538</td>
</tr>
<tr>
<td>Kalm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.533</td>
</tr>
<tr>
<td>Timide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.520</td>
</tr>
<tr>
<td>Verlegen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.410</td>
</tr>
<tr>
<td>Ontspannen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Loadings < |0.40| are not shown
The items “eenzaam” and “alleen” (lonely and alone) load negatively on this second factor. This negative loading could be explained by the fact that these two items are related to NA, the “opposite” of PA where Joviality belongs to. The third factor contains the items “enthousiast”, “aandachtig”, “oplettend”, “alert”, “geïnteresseerd”, “geconcentreerd” and “geïnspireerd” (enthusiastic, determined, attentive, alert, interested, concentrating and inspired). The items alert, attentive, concentrating and determined belong to the basic positive emotion Attentiveness. The three other items enthusiastic, interested and inspired belong to the general dimension scale PA. Therefore, the third factor could be seen as the basic emotion Attentiveness.

The fourth factor contains the items “moedig”, “dapper”, “sterk” and “onbevreesd” (daring, bold, strong and fearless). These four items are part of the basic emotion Self-Assurance. Therefore, the fourth factor can be seen as the basic emotion Self-assurance.

The fifth factor consists of the items “moe”, “slaperig”, “lui” and “suf” (tired, sleepy, sluggish and drowsy). These items are exactly the items of the affective state Fatigue. For that reason, the fifth factor can be seen as Fatigue.

The sixth factor contains the items “kalm”, “timide”, “verlegen” and “ontspannen” (calm, timid, shy and relaxed). These items are part of the other affective state Shyness and Serenity. This sixth factor shows no clear pattern.

For the first factor it is visible that nearly all items, except for scornful and guilty, of the basic negative emotion scales Fear, Hostility, Guilt and Sadness load on this factor. In other words, this factor contains all items related to the basic negative emotion scales. An attempt was made for subdividing this factor further. This was done by selecting all the items of factor 1 plus the items scornful and guilty, and deleting the items of the other affective states which loaded on factor 1. Several solutions were examined.

The best interpretable solution is a five factor solution, but this solution does not contain the structure as proposed by Watson and Clark [WAT99A]. It is not possible to extract the specific basic negative emotions. This finding is in line with the results presented by Bagozzi [BAG93]. Bagozzi reanalyzed the data of Watson and Clark [WAT92], [WAT92A]. Bagozzi found that when someone wants to examine the overall negative affect, the composite of the basic negative affect scales may be useful to employ. This composite is the Basic Negative Emotion Scale which consists of Fear, Hostility, Guilt and Sadness. However, if one wants to examine the unique contribution of these specific basic negative affect scales, the individual basic scales are only useful when well-defined conditions can be identified that lead to differentiating emotional responses [BAG03]. This conclusion is based on the mixed findings on discriminant validity of the basic negative emotion scales.

Summarizing, the current dataset supports the two general affective states of PA and NA. The lower order scales of PA can be confirmed with the dataset. The underlying structure of PA in the form of the basic emotion scales Joviality, Self-assurance and Attentiveness is clearly visible. The underlying structure of NA can not be confirmed as proposed by Watson and Clark [WAT99A]. Nevertheless, the dataset clearly confirms the existence of the “other” affective state Fatigue. Cultural differences could be a reason for the correlation between the four affective states Fear, Hostility, Guilt and Sadness in the Dutch version of PANAS-X. Cultural variations could influence the emotion process [MES92]. There are strong cultural differences in perceiving emotions.
Probably, Dutch people do not express their negative feelings in a specific way like Fear, Hostility, Guilt and Sadness. They do not distinguish between these specific basic negative emotions and perceive all these basic negative emotions the same, as one NA.

**Reliability of PANAS-X**

The reliabilities (Cronbach's alphas) of the subscales of the PANAS-X are presented in Table 6.11. Reliability is an assessment of the degree of consistency between multiple measurements of a variable. In other words, reliability indicates whether the individual items of a particular scale all measure the same construct.

Cronbach's alpha for the General Dimension Scale Negative Affect is 0.830. The generally agreed lower limit for Cronbach's alpha is 0.70 [HAI05]. The Basic Negative Emotion Scale has a high Cronbach's alpha (0.906). Also the scales Fear, Hostility, Guilt and Sadness have sufficient Cronbach's alphas (0.782, 0.771, 0.755 and 0.779 respectively). The Cronbach's alphas for the General Dimension Scale Positive Affect is 0.747, indicating that the reliability of this scale is acceptable. The Basic Positive Emotion Scale has a high Cronbach's alpha (0.833). The Cronbach's alphas of the Basic Positive Emotion subscales range from 0.666 till 0.802 which can be classified as acceptable. The Cronbach's alpha of Attentiveness is 0.666. This value is acceptable considering that this scale only consists of four items.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Cronbach's alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Dimension Scales</strong></td>
<td></td>
</tr>
<tr>
<td>General Dimension Scale Negative Affect</td>
<td>0.830</td>
</tr>
<tr>
<td>General Dimension Scale Positive Affect</td>
<td>0.747</td>
</tr>
<tr>
<td><strong>Basic Negative Emotion Scales</strong></td>
<td>0.906</td>
</tr>
<tr>
<td>Fear</td>
<td>0.782</td>
</tr>
<tr>
<td>Hostility</td>
<td>0.771</td>
</tr>
<tr>
<td>Guilt</td>
<td>0.755</td>
</tr>
<tr>
<td>Sadness</td>
<td>0.779</td>
</tr>
<tr>
<td><strong>Basic Positive Emotion Scales</strong></td>
<td>0.833</td>
</tr>
<tr>
<td>Joviality</td>
<td>0.802</td>
</tr>
<tr>
<td>Self-Assurance</td>
<td>0.714</td>
</tr>
<tr>
<td>Attentiveness</td>
<td>0.666</td>
</tr>
<tr>
<td><strong>Other Emotion Scales</strong></td>
<td></td>
</tr>
<tr>
<td>Shyness</td>
<td>0.582</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.797</td>
</tr>
<tr>
<td>Serenity</td>
<td>0.350</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.747</td>
</tr>
</tbody>
</table>

The Cronbach's alpha for the Shyness scale is 0.582. This value is rather low and it is questionable whether the Shyness scale is reliable enough. The scale Fatigue and Surprise have a sufficient Cronbach's alpha, which indicates that these scales are reliable. The scale Serenity has an unacceptably low value of Cronbach's alpha: 0.350.
Therefore, the scale Serenity is not reliable. Conclusively, it can be stated that the scales of PANAS-X are reliable, except for the Serenity and Shyness scales.

**Validity of PANAS-X**

Besides reliability, also the validity of PANAS-X has to be tested. Validity is the extent to which a scale or set of measures represents the concept of interest [HAI05]. There are several forms of validity. Two widely used measures of validity are convergent and discriminant validity. Convergent validity assesses the degree to which two measures of the same concept are correlated [HAI05]. Discriminant validity is the degree to which a scale is sufficiently different from other related scales [HAI05]. For assessing convergent and discriminant validity, a correlation matrix is presented in Table 6.12. For determining convergent validity, the correlations have to be high between related scales. For proving discriminant validity, the correlations between unrelated scales have to be low.

**Positive and Negative Affect**

The convergent and discriminant validity of the basic negative emotion scale (Fear, Hostility, Guilt and Sadness) and basic positive emotion scale (Joviality, Self-Assurance and Attentiveness) are examined. First, the correlations in Block A of Table 6.12 are examined. The emotion scales Fear, Hostility, Guilt and Sadness are correlated very significantly. These three scales are also correlated very significantly and highly with the general negative affect and basic negative emotion scale (Block C of Table 6.12). These three scales (Block A) are not significantly correlated with the basic positive emotion scales Joviality, Self-Assurance and Attentiveness, except for the Sadness scale. The Sadness scale is significantly correlated with Joviality, but the size of the correlation is not that high (0.229) in comparison with the correlation of Sadness with Fear, Hostility and Guilt.

Sadness is also significantly correlated with the basic positive emotion scale. It is important to note that the direction of this correlation is negative. The correlation between the basic positive emotion scales Joviality, Self-assurance and Attentiveness is significant (Block B), but the size is (much) lower than the correlation among the basic negative scales. This is also a reason why factor analysis could extract three factors for the three basic positive emotion scales Joviality, Self-Assurance and Attentiveness, but could only extract one factor for the four negative emotion scales Fear, Guilt, Hostility and Sadness. Joviality, Self-assurance and Attentiveness (Block B) load high on general positive affect and basic positive emotion scale (Block D). These results show that sufficient discriminant validity exists between the positive and negative emotion scales, because there is (nearly) no correlation between these emotion scales. The high correlations of these scales with their corresponding higher order factor show that there is sufficient convergent validity. The scales indeed measure positive or negative affect. The discriminant validity of the scales Fear, Hostility, Guilt and Sadness is questionable, because the correlation of these emotion scales is high (Block A). The scales Fear, Hostility, Guilt and Sadness have a high communality. It is questionable whether Fear, Hostility, Guilt and Sadness really measure something different or just stick together measuring one emotion. The correlation between Joviality, Self-assurance and Attentiveness (Block B) is not as high as in the case of Fear, Hostility, Guilt and Sadness (Block A). This lower correlation implies that Joviality, Self-assurance and Attentiveness measure a different concept and that discriminant validity of these scales is sufficient.
Table 6.12: Correlation matrix of the PANAS-X scales

<table>
<thead>
<tr>
<th></th>
<th>Fear</th>
<th>Hostility</th>
<th>Guilt</th>
<th>Sadness</th>
<th>Joviality</th>
<th>Self-Assurance</th>
<th>Attentiveness</th>
<th>Shyness</th>
<th>Fatigue</th>
<th>Serenity</th>
<th>Surprise</th>
<th>General Pos Aff</th>
<th>General Neg Aff</th>
<th>Bas Pos Emo</th>
<th>Bas Neg Emo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hostility</td>
<td>0.614*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guilt</td>
<td>0.654*</td>
<td>0.524**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>0.589*</td>
<td>0.502**</td>
<td>0.580**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joviality</td>
<td>0.229**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.414**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Assurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.352**</td>
<td>0.281**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attentiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.352**</td>
<td>0.281**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shyness</td>
<td>0.450**</td>
<td>0.260**</td>
<td>0.368**</td>
<td>0.510**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.481**</td>
<td>0.373**</td>
<td>0.313**</td>
<td>0.420**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serenity</td>
<td>-0.187**</td>
<td>-0.262**</td>
<td>-0.212**</td>
<td>-0.220**</td>
<td>0.156*</td>
<td>0.168*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td>0.535**</td>
<td>0.413**</td>
<td>0.430**</td>
<td>0.302**</td>
<td>0.214**</td>
<td>0.154*</td>
<td>0.350**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Pos Aff</td>
<td>0.655**</td>
<td>0.492**</td>
<td>0.747**</td>
<td>0.430**</td>
<td>0.450**</td>
<td>0.471**</td>
<td></td>
<td>0.160*</td>
<td>0.251**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Neg Aff</td>
<td>0.870**</td>
<td>0.741**</td>
<td>0.768**</td>
<td>0.635**</td>
<td>0.430**</td>
<td>0.430**</td>
<td></td>
<td>0.450**</td>
<td>-0.237*</td>
<td>0.509**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bas Pos Emo</td>
<td>0.867**</td>
<td>0.812**</td>
<td>0.836**</td>
<td>0.787**</td>
<td>0.800**</td>
<td>0.820**</td>
<td></td>
<td>0.609**</td>
<td>-0.204**</td>
<td>0.187**</td>
<td>0.793**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bas Neg Emo</td>
<td>0.867**</td>
<td>0.812**</td>
<td>0.836**</td>
<td>0.787**</td>
<td>0.800**</td>
<td>0.820**</td>
<td></td>
<td>0.609**</td>
<td>-0.204**</td>
<td>0.187**</td>
<td>0.793**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** correlation is significant at the 0.01 level (2-tailed)
* correlation is significant at the 0.05 level (2-tailed)
Other affective states
Subsequently, the convergent and discriminant validity of the other affective scales (Shyness, Fatigue and Surprise) are examined. Table 6.13 is used as a guide for exploring the relevant correlations with regard to the other affective states.

Table 6.13: Correlations of the other affective states

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fear</strong></td>
<td>0.450**</td>
<td>0.481**</td>
<td>-0.187**</td>
</tr>
<tr>
<td><strong>Hostility</strong></td>
<td>0.260**</td>
<td>0.373**</td>
<td>-0.262**</td>
</tr>
<tr>
<td><strong>Guilt</strong></td>
<td>0.368**</td>
<td>0.313**</td>
<td>-0.212**</td>
</tr>
<tr>
<td><strong>Sadness</strong></td>
<td>0.510**</td>
<td>0.420**</td>
<td>-0.220**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Joviality</th>
<th>Serenity</th>
<th>Shyness</th>
<th>Fatigue</th>
<th>0.156*</th>
<th>0.214**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Assurance</strong></td>
<td>0.168*</td>
<td>0.293**</td>
<td>0.350**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attentiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shyness</td>
<td>0.293**</td>
<td>0.293**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatigue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serenity</td>
<td>0.350**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General Pos Aff</th>
<th>General Neg Aff</th>
<th>Bas Pos Emo</th>
<th>Bas Neg Emo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.430**</td>
<td>0.450**</td>
<td>-0.160*</td>
<td>0.471**</td>
</tr>
<tr>
<td>0.160*</td>
<td>-0.237**</td>
<td>0.204**</td>
<td>-0.266**</td>
</tr>
<tr>
<td>0.160*</td>
<td>0.251**</td>
<td>0.509**</td>
<td>0.513**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** correlation is significant at the 0.01 level (2-tailed)
* correlation is significant at the 0.05 level (2-tailed)

Shyness
Shyness correlates very significantly with the negative basic emotions Fear, Hostility, Guilt and Sadness (especially with Fear and Sadness, see column 1-row A). Shyness correlates very significantly with general negative affect and the basic negative emotion scale (column 1-row D). Shyness correlates not only with the negative affect scales, but also with Fatigue and Surprise (column 1-row C). These correlations are significant (0.293 and 0.350). In other words, the claim that Shyness does possess sufficient discriminant validity can not be proven.

Fatigue
Fatigue correlates very significantly with Fear, Hostility, Guilt and Sadness (see column 2-row A). Fatigue correlates very significantly with general negative affect and the basic negative emotion scale (column 2-row D). Fatigue correlates substantially positively with Shyness (column 2-row C) and negatively with the basic emotion scale (column 2-row D). Also, the substantial correlation between the Fatigue and Shyness scales is a reason to reject the idea that the Fatigue scale possesses sufficient discriminant validity.

Serenity
Serenity correlates negatively with Fear, Hostility, Guilt and Sadness but not as high as the other affective states (Shyness, Fatigue and Surprise), see Row A. Serenity also correlates positively with Joviality and Self-Assurance (column 3-row B).
Serenity correlates positively with general positive affect and the basic positive emotion scale and negatively with general negative affect and the basic negative emotion scale (column 3-row D). However, the Serenity scale is not valid because it is not even reliable.

**Surprise**

Surprise correlates positively with a lot of scales (Fear, Hostility, Guilt, Sadness, Joviality, Attentiveness, Shyness, general positive affect, general negative affect, basic positive emotion scale and basic negative emotion scale), see column 4. Clearly, it is very questionable whether Surprise is sufficiently different from other scales. It is not clear what the Surprise scale measures, so discriminant validity is not sufficient.

In summary, this analysis shows that there is reason to claim that the convergent validity of the basic emotions scales Fear, Hostility, Guilt and Sadness is sufficient. Also, the convergent validity of Joviality, Self-Assurance and Attentiveness is sufficient. Since the correlation between the scales Fear, Hostility, Guilt and Sadness is high (0.502 – 0.654), there is no ground for claiming that these scales possess sufficient discriminant validity. The correlation between Joviality, Self-Assurance and Attentiveness is small enough (0.281 – 0.414) to state that the discriminant validity is sufficient. With regard to the other affective states Shyness, Fatigue, Serenity and Surprise, no clear conclusion can be drawn about convergent validity, because it is unclear what these other affective states should measure. Discriminant validity is not sufficient, due to the many correlations of these scales with other scales. In general, it is appropriate to use the General Dimensions scales Negative Affect and Positive Affect, the Basic Negative and Positive Emotion Scales, and the scales Joviality, Self-Assurance and Attentiveness for the UPFS measurement. Care has to be taken when using Fear, Hostility, Guilt and Sadness, because the discriminant validity is questionable. Using the other affective states Shyness, Fatigue, Serenity and Surprise is not recommended, because of their very limited discriminant validity. Further research related to user perceived failure severity should not use the scales related to the other affective states, because no conclusion concerning the reliability of these scales can be drawn from the data obtained by these scales.

### 6.3 Extraneous variables: control for users and their environments

This research investigates the relation between FUI and UPFS. However, the relation between FUI and UPFS in this experiment may be biased by the influence of extraneous variables (see Figure 6.1). As mentioned before, an extraneous variable is a variable that is not of interest in the study, but which might have influence on the relationships being studied [GOO05], [STA98]. In order to limit the influence of these variables on the dependent variable (UPFS), the extraneous variables should be kept as constant as possible.

In the previous chapter, the relevant extraneous variables in the context of this research are determined to be (see Section 5.2.2):

- The use environment: the physical conditions in which the product is used
- The use approach: the way the product is utilized by the user
- Irritableness: quick-temperredness
- Other user characteristics
- Other failure characteristics
Although the influence of these variables on UPFS is not of interest in this research, these variables should be kept as constant as possible to limit their interference with the outcomes of the experiment. This section explains how, in this research, these variables can be controlled.

### 6.3.1 Use conditions

**Control for the use environment**

In order to limit the influence of the use environment on UPFS, all the experiments should be executed in a laboratory environment similar to the environment of the first experiment (see Figure 4.4). Using a laboratory setting improves the experimental control, which is the extent to which the experimenter is able to eliminate environmental effects on the dependent variable [STA98]. In general the results of laboratory studies mirror the results of field studies as elaborate research demonstrates a high degree of correspondence between the results found in and outside the lab [GOO05], [AND99]. Nevertheless, laboratory research has several limitations. As a general rule can be stated that any unrealistic aspect of a study renders it externally invalid [BER82]. This makes field observation more preferable to laboratory experimentation since laboratory research always involves some unrealistic elements [LYN82]. Moreover, not all consumer research can be performed in a laboratory setting. Especially long-term consumer research requires field observation. Notwithstanding these limitations to laboratory research, the proposed UPFS experiment will be executed in a laboratory environment to limit the influence of the use environment on the experimental results.

**Control for the use approach**

In order to control for the use approach in this research, the experimental protocols should be standardized. Standardization of the experimental protocols results in the equal treatment of all participants and consequently causes minimal differences between experimental use conditions of participants. Optimal standardization of the experimental protocol is obtained when all experimental tasks are executed by the participant without the interference of an observer. Consequently, in order to restrict the influence of the use approach on the results of this research, no observer should be present during the activation and evaluation of the failure scenario by the test participant.

### 6.3.2 User characteristics

**Control for irritableness**

This research explores the relation between FUI and UPFS. A user characteristic that is expected to interfere with the UPFS is irritableness (see Section 5.2.2). This specific extraneous variable of irritableness should be kept constant in this research by matching. In matching, participants are grouped together on some trait (e.g. irritableness) and distributed randomly to the different groups in the experiment [GOO05].

In order to control for irritableness in this research, participants should be selected that reveal an average level of irritableness. Such a selection approach requires measurement of the irritableness level of each potential participant. Subsequently, this irritableness level should be compared with the average irritableness level of all other potential participants. Based on these measurements, the most “extreme” (high and low irritableness level) people can be removed from the group of participants.
In contrast with irritation, irritableness is a stable trait that describes how irritable people usually or typically are. Fortunately, PANAS-X is suitable for measuring changes in states but also for measuring traits [WAT99], [IHR06]. By making some small adjustments to the PANAS-X measurement scale that measures UPFS (see Appendix G), this scale can be used for measuring irritableness as well. The most important adjustment is that instead of asking people to rate how they feel “at this moment” (transitory state) people are questioned about how they feel “in general”. Again, the general negative affect scale is most appropriate for measuring the irritableness of people [IHR06], [WAT99A]. The Basic Negative Emotion scales Fear, Hostility, Guilt and Sadness of PANAS-X (see Table 6.5) are consistently and substantially correlated (loadings ranging from 0.69 to 0.83) with the General Dimension Scale Negative Affect [WAT99], [WAT99A]. In other words, a high score on the subscales of the Basic Negative Emotion Scales probably leads to a high score on the General Dimension Scale Negative Affect. Therefore, also these Basic Negative Emotion scales are used for measuring irritableness.

The other affective scales Shyness, Fatigue, Surprise and Serenity do not load strongly or consistently on one of these two scales [WAT99]. However, only using the scales related to Negative Affect leads to response acquiescence – a tendency to agree with statements [GOO05]. This means that not only the Negative Affect Scales should be incorporated, but also Positive Affect Scales and other affective scales should be incorporated. This should force participants to read each item carefully and to make item-by-item decisions.

Section 6.2.3 demonstrated the validity and reliability of the Dutch translation of the General dimensions scales Negative Affect and Positive Affect, the Basic Negative and Positive emotion scale, and the scales Joviality, Self-Assurance, Attentiveness. The use of the other low level scales is not recommended. Therefore, irritableness is controlled using the General Dimension Scale Negative Affect and the Basic Negative Emotion scale.

Control for other user characteristics
Matching is difficult to accomplish with more than one matching variable [GOO05]. Besides that, it is unclear which other user characteristics may influence UPFS. Therefore, the other (unknown) extraneous user characteristics should be kept constant by creating equivalent experimental groups using random assignment.

The equivalent experimental groups can be created by using only university students as test subjects. Moreover, with random assignment, every test participant for the experiment has an equal chance of being placed in any of the groups to be formed. The goal of random assignment is to take individual difference factors (user characteristics) that could bias the study and spread them evenly throughout the different groups [GOO05].

The use of university students as test subjects in this research is not obvious considering the involved product: high-end TV. In daily life, university students probably do not possess such an expensive TV. It is even conceivable that some students might have never used such a TV before. This unrealistic aspect of the experimental design limits the external validity of the experimental results. In spite of this limitation, university students are still selected as test subjects to control for the other user characteristics by random assignment.
6.3.3 Failure characteristics

The previous chapter indicated that all other failure characteristics of which the influence is not investigated in this research are also extraneous variables that should be controlled. The following list presents simple approaches for controlling these other failure characteristics.

- **Failure Work Around**: all implemented failures cannot be prevented by the test participants
- **Failure Frequency**: all implemented failures only happen once
- **Failure Moment in Use Process**: all implemented failures happen at a predefined moment in the experimental protocol
- **Failure Reproducibility**: all the implemented failures cannot be repeated by the test participants
- **Failure Solvability**: all the implemented failures cannot be solved by the test participants
- **Failure Attribution**: all the implemented failures are caused by the TV itself
- **Failure Impact**: the percentage loss of functionality that is caused by the implemented failures is equal for all failures

The control of the first five failure characteristics is rather straightforward; the required failure preconditions are unambiguous and easy to implement. However, the control of Failure Attribution and Failure Impact is less simple. People can blame other factors than the actual (implemented) cause of the failure. Therefore, the Failure Attribution of the test participants should be evaluated afterwards to check whether it has been successfully controlled. This can be done by asking test participants during the debriefing of the experiment to mention the assumed cause of the experienced failure.

The control of Failure Impact is also more troublesome, because of the absence of a universal FI measurement scale. In other words, it is difficult to compare and tune the impact of failures in different product functions. Therefore, it is important to measure the participant perceived Impact of the implemented failures to check whether Failure Impact has been successfully controlled. For this purpose, the Function Failure List is developed (see Appendix H). This Function Failure List is an overview of TV functions of which the participants should indicate whether they worked flawlessly during the experiment. When the participant indicates that a function contains a failure, he/she should give a score between 0 and 5. A zero means that the function could be used very well despite the failure and a five means that the function could almost not be used. In order to control for this failure characteristic, the Failure Impact score should be equal for all experimental scenarios.

6.4 Conclusions

Investigating the influence of the independent FUI variable on the dependent UPFS variable requires the variation of the FUI level and the measurement of its effect on UPFS. Therefore, this chapter dealt with the description and validation of the measurement approaches for the independent (FUI), dependent (UPFS) and extraneous variables (use environment, use approach, user irritableness and other user characteristics).
The first section of this chapter described the importance analysis of the different TV functions. The resulting FUI ranking of TV functions enables the variation of the independent FUI variable by selecting TV functions with different FUI levels.

The second section was concerned with the measurement of UPFS. On the one hand, the self-report UPFS survey was adjusted to improve its overall reliability. On the other hand, a second UPFS measurement approach was introduced using the PANAS-X scheme. Application of this added UPFS measurement could further validate the results of the self-report UPFS measurement and this approach is expected to be more specific on the emotional consequences of the implemented failures for the test participants. Reliability and validity analyses of this new translation of PANAS-X confirm the appropriateness of using the General Dimensions scales Negative Affect and Positive Affect, the Basic Negative and Positive Emotion Scale, and the scales Joviality, Self-Assurance and Attentiveness for the UPFS measurement.

The third section explained how, in this research, the extraneous variables can be controlled. The use environment and use approach can be controlled by using respectively a laboratory environment and standardized experimental protocols. Further, irritableness should be controlled by matching participants on their irritableness level using PANAS-X. The other user characteristics can be controlled by creating equivalent experimental groups of students using random assignment.

In summary, this chapter resulted in different approaches for the variation, measurement and control of the experimental variables of the experimental model in Figure 6.1. In the next chapter, based on these selected approaches, an experimental design for investigating the relationship between FUI and UPFS is formulated and executed. Subsequently, the experimental results will establish the actual influence of FUI on UPFS.
7 Function Importance Experiment

As mentioned before, this research concentrates on investigating the influence of Function Importance (FUI) on UPFS. The hypothesized relationship between these two variables is presented in the experimental model of Figure 7.1 (the same as Figure 6.1).

![Diagram of Experimental Model FUI Experiment]

The previous chapter described different approaches for the variation, measurement and control of the experimental variables of Figure 7.1. Thereupon, this chapter deals with the actual design and execution of the second consumer experiment in order to investigate the hypothesized relation between FUI and UPFS.

The first section gives an overview of the proposed design for this second consumer experiment. In Section 7.2, some hypotheses are formulated based on the experimental model of Figure 7.1. Subsequently, Section 7.3 gives an overview of the experimental variables of this second consumer experiment. Next, Section 7.4 deals with the design of the experimental tasks: the failure scenarios. The fifth section describes the selection process of the participants for the proposed consumer experiment. Subsequently, a complete description of the experimental protocol is presented in Section 7.6. Before performing the actual consumer experiments, a pilot experiment is executed to test this experimental protocol. The results of this pilot experiment are given in Section 7.7. Then, Section 7.8 presents the results of the actual consumer experiments. Lastly, based on these experimental results, in Section 7.9 some conclusions are drawn about the relationship between FUI and UPFS in consumer electronics.

7.1 Overview of the FUI experiment

In order to investigate the relation between FUI and UPFS, a second consumer experiment is performed. The experimental design forms the basis of this second experiment. In general, two types of experimental design can be distinguished [GOO05]:

- **Between-subject design**: those who participate in the study might be placed in level A or level B of the independent variable

- **Within-subject design**: those who participate in the study receive both level A and level B of the independent variable
In a within-subject design, each participant is exposed to each level of the independent variable. Conversely, the between-subject design requires different test participants for each level of the independent variable. A practical advantage of the within-subject approach is that fewer people need to be recruited for the experiment. Next to that, the initial difference between participant groups at each level of the independent variable is logically negligible. This minimal difference is not self-evident in a between-subject design. However, the major disadvantage of within-subject designs is the presence of sequence effects; for a participant who has completed the first part of a study, the knowledge, experience or altered circumstances could influence performance in later parts of the study [GOO05]. In a between-subject design test participants only deal with one level of the independent variable. Consequently, this design does not suffer from the negative influence of sequence effects. Sequence effects are known to be very strong in consumer experiments. Therefore the between-subject design is selected for this second consumer experiment.

Moreover, a single factor two level design is selected for investigating the influence of FUI on UPFS. In other words, in this between-subject design, participants are placed in an experimental group with a low level or a high level of the independent variable FUI. This means that some participants are confronted with a failure in a low importance function whereas other participants are confronted with a failure in a high importance function. Subsequently, the UPFS levels of the two participant groups are statistically compared in order to investigate the influence of FUI on UPFS.

7.2 Hypotheses

The starting point for investigating this influence of FUI on UPFS is Figure 7.1. This figure illustrates the anticipated positive influence of FUI on UPFS. In other words, a higher importance of the failed function is expected to result in a higher level of user dissatisfaction. These expectations can be translated into the following null hypothesis and alternative hypothesis:

\[ H_0 : \text{There is no significant difference in UPFS caused by failures with different levels of function importance} \]

\[ H_1 : \text{The UPFS that is caused by a failure in an important function is significantly higher than by a failure in an unimportant failure} \]

Consequently, by statistically comparing the UPFS levels of the two experimental groups (high and low FUI levels), the validity of these hypotheses can be tested.

7.3 Experimental variables

The proposed single factor two level experimental design consists of one independent variable (FUI), one dependent variable (UPFS) and several extraneous variables. The previous chapter described different approaches for the variation, measurement and control of these experimental variables. This section gives a short overview of the application of these variables in the proposed consumer experiment.

7.3.1 Independent variable: FUI

FUI is the relative importance of the function affected by the failure (see Table 5.1). In the experimental model of Figure 7.1, FUI is considered the independent variable that influences
UPFS. In order to test the hypotheses of Section 7.2, a single factor two level experimental design is performed. This implies that in the proposed experiment the FUI level of the experimental scenarios should be set at two levels: a high FUI level and a low FUI level. This difference in FUI level among the experimental scenarios can be accomplished by selecting two TV functions with different FUI levels.

A structured FUI evaluation in the previous chapter resulted in a FUI ranking of TV functions (see Section 6.1.2). For the variation of FUI among the experimental scenarios, an important and unimportant TV function can be selected from this ranking of TV functions. The decision for these two functions is based on the FUI Friedman ranking as presented in Table 6.3. Firstly, the two functions need to have a significant difference in perceived importance by the user. Furthermore, the controlled disruption (failure) of the selected functions should be easily applicable in an experiment. It must be possible to let the participant actively use the function during the experiment and to trigger a failure (remotely) in the function while the participant is using that function.

The most important function, according to Table 6.3, is to “Watch the desired program”. This function can be easily affected by degrading the quality of the image or the sound. The deterioration of the image or sound could even be controlled remotely when the participant is using the TV in the laboratory. In other words, the function “Watch the desired program” is very suitable for the failure scenario design of the proposed experiment. Therefore, this function is selected as the high FUI function in the proposed consumer experiment.

For the three least important TV functions “Sleep timer”, “Child lock” and “Switch on timer”, it is very difficult to let the participants actively use these functions during the experiment. However, the “Motorized swivel” function, the fourth least important function, can easily be embedded in the experimental task list. Moreover, it should be rather uncomplicated to cause a failure (remotely) in the “Motorized swivel” function while the participant uses this function.

Nevertheless, the “Watch the desired program” and the “Motorized swivel” function also need to have a significant difference in perceived importance by the user. In order to investigate this second condition, the Wilcoxon Signed Ranks test is used to test if the scores given to the two functions differ significantly [MON99]. The results of this test are presented in Table 7.1.

<table>
<thead>
<tr>
<th>Rankings</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Ranks: rating &quot;Motorized swivel&quot; &lt; rating &quot;Watch the desired program&quot;</td>
<td>116</td>
</tr>
<tr>
<td>Positive Ranks: rating &quot;Motorized swivel&quot; &gt; rating &quot;Watch the desired program&quot;</td>
<td>1</td>
</tr>
<tr>
<td>Ties: rating &quot;Motorized swivel&quot; = rating &quot;Watch the desired program&quot;</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
</tr>
<tr>
<td>Asymptotic Significance (2-tailed)</td>
</tr>
</tbody>
</table>

The asymptotic significance of the Wilcoxon Signed Ranks test is smaller than 0.05, which indicates that there is a significant difference in average ratings between the two functions.
Table 7.1 illustrates that the rating of the “Motorized swivel” function is lower than the “Watch the desired program” function in 116 out of the 121 cases. Based on these results, it can be concluded that the “Watch the desired program” function is considered significantly more important than the “Motorized swivel” function.

Summarizing, for the single factor two level design of the second consumer experiment, two levels of the independent variable are defined. In other words, two TV functions (one important and one unimportant) are selected for the scenario design of this experiment, namely:

- **Important function**: “Watch the desired program”
- **Unimportant function**: “Motorized swivel”

### 7.3.2 Dependent variable: UPFS

UPFS is the level of irritation experienced by the user caused by a product failure (see Section 3.4). In the experimental model of Figure 7.1, UPFS is considered the dependent variable that is influenced by FUI. In the previous subsection, two TV functions were selected for the scenario design of the second experiment. The different FUI levels of these two TV functions enable the variation of the independent FUI variable in this experiment. However, statistical testing of the hypotheses requires the measurement of the influence of these different FUI levels on the UPFS. In other words, the UPFS of the two different failure scenarios (high and low FUI level) should be measured.

The previous chapter proposed two approaches for UPFS measurement in this research:

- **The UPFS survey**: a self-report survey consisting of five statements to measure the level of irritation experienced by the user as caused by the TV failure and five additional statements to describe the expected reactions of the participant to the experienced failure scenario (see Appendix F). The first five statements (the irritation statements) together form the self-report UPFS measurement scale.

- **PANAS-X**: 60-item emotion measurement schedule (see Appendix G). The difference between the user's emotional Negative Affect state just before product failure and just after product failure can be considered as a measure for UPFS. The Dutch translation of the General Dimension scale Negative Affect and the Basic Negative Emotion scale possesses good validity and reliability and is therefore considered appropriate for UPFS measurement (see Section 6.2.3).

Consequently, both these UPFS measurement approaches are applied in the second consumer experiment to measure the influence of FUI (independent variable) on UPFS (dependent variable).

### 7.3.3 Extraneous variables

The relation between FUI and UPFS in this experiment may be biased by the influence of extraneous variables (see Figure 7.1). In order to limit the influence of these variables on the dependent variable (UPFS), the extraneous variables should be controlled.

The previous chapter described several approaches for controlling the relevant extraneous variables in this research (see Section 6.3):

- **The use environment**: all the experiments should be executed in a laboratory environment to improve the experimental control (see Figure 4.4)
- **The use approach:** the experimental protocols should be standardized to minimize the difference between experimental use conditions of participants
- **Irritableness:** participants should be selected that reveal an average level of irritableness using PANAS-X
- **Other user characteristics:** equivalent experimental groups should be created by using only university students and random assignment
- **Other failure characteristics:** in the definition of the failure scenarios, the other characteristics of the failures should be equal for both scenarios

In order to limit the influence of these extraneous variables on the dependent variable (UPFS), all control approaches are applied in the second consumer experiment.

### 7.4 Experimental tasks: failure scenario design

The previous section gave an overview of the application of the experimental variables in the second consumer experiment. Based on these results, this section describes the development of the failure scenarios for the proposed consumer experiment.

#### 7.4.1 Scenario 1: Failure in important function

Participants consider “Watch the desired program” as the most important function of a TV. The description of this function is “to choose the program you want and watch it without flaws in image and sound” (see Table 6.2). In order to test how participants perceive the failure of this high importance function, a suitable failure scenario needs to be developed. A rather convenient way of degrading this function is by disrupting the analog television signal. Therefore, a (remotely controlled) signal disrupter is installed to cause noise in the analog television signal. When activated, this disrupter causes snow in the image of some TV channels and total image failure (black screen) of the other channels. This degradation of the analog television signal using the signal disrupter is embedded in a predefined set of tasks that test participants have to go through. Logically, the test participant is unaware of the actual purpose of the experiment and the implemented failure in this TV function. The task list for this failure scenario consists of the following activities:

- **Introduction tasks:** the participant tries some of the TV functions based on an acquaintance task list. This task list lets the participant become familiar with the TV and the laboratory environment. Moreover, the execution of these introduction tasks creates an equal point of departure for all participants at the start of the failure scenario task list. Lastly, after the introduction tasks, a supplemental FUI measurement is performed with a restricted number of TV functions. This FUI measurement can be used to verify whether a significant difference exist between the FUI ratings of the two selected TV functions among the test participants. The acquaintance task list is presented in Appendix I (in Dutch).

- **Scenario 1 task:** the participant is requested to make a list of the broadcasted channels and programs on TV at that moment. This scenario 1 task list is presented in Appendix J (in Dutch). During the execution of this task by the participant, the signal disrupter is triggered, causing a temporary disruption of the TV image. This disruption continues for one minute. After this minute, the original image quality is restored.
7.4.2 Scenario 2: Failure in unimportant function

Participants consider the “Motorized swivel” function as a significantly less important function than the “Watch the desired program” function. The motorized swivel function can be used to turn the screen between -30 and +30 degrees to optimize the viewing angle (see Table 6.2). Figure 7.2 illustrates the functionality of the motorized swivel.

![The motorized swivel function](image)

If a user changes his position in the room, he might want to adjust the viewing angle of the screen. The “Motorized swivel function” can be used for this, but the viewing angle can also be changed by turning the screen manually.

In order to test how participants perceive the failure of this low importance function, a suitable failure scenario needs to be developed. A failure can be caused in this function by blocking the electrical current to the swivel. Therefore, a (remotely controlled) adjustable socket is installed to cause a power cut-off to the motorized swivel. When activated, this power disruption causes the break down of the motorized swivel function. Also, this disruption of the motorized swivel function using the adjustable socket is embedded in a predefined set of tasks that test participants have to go through. Of course, the test participant is unaware of the actual purpose of the experiment and the implemented failure in this TV function. The task list for this failure scenario consists of the following activities:

- **Introduction tasks:** the purpose and design of the introduction tasks is exactly the same for this scenario as for the first failure scenario (see Appendix I).

- **Scenario 2 task:** the participant is requested to change position in the laboratory room. The participant is asked to take place behind a laptop at the table at the other side of the laboratory room (see Figure 4.4). Subsequently, the test participant is asked to make a list of broadcasted programs using the laptop. The participant cannot watch the television from this angle and so the viewing angle of the TV should be adjusted using the motorized swivel. After the completion of this task, he/she is asked to go back to the couch and watch TV. In order to accomplish the next task, the participant should turn back the TV with the motorized swivel. The task list of this second scenario is presented in Appendix K (in Dutch). When the participant tries to use the motorized swivel for the second time, the adjustable socket is triggered causing a break down of the motorized swivel function. This disruption continues for one minute. After this minute, original function quality is restored.
In the definition of these scenarios, the other failure characteristics are equal in both scenarios. The control of the other failure characteristics is implemented as follows:

- **Failure Work Around**: both implemented failures cannot be prevented by the test participants
- **Failure Frequency**: both implemented failures only happen once
- **Failure Moment in Use Process**: both implemented failures happen at the same predefined moment in the experimental protocol
- **Failure Reproducibility**: both implemented failures cannot be repeated by the test participants
- **Failure Solvability**: both implemented failures cannot be solved by the test participants
- **Failure Attribution**: both implemented failures are caused by the TV itself
- **Failure Impact**: the percentage loss of functionality that is caused by the implemented failures is equal for all failures

Lastly, it is important to mention that the TV content (the program broadcasted on TV) should matter to the test participants. If the content is considered unimportant by the test participants, an implemented failure scenario would probably cause a limited or no emotional response among these participants. They simply would not care about the failure. In order to assure the content importance in this experiment, both scenarios prescribe tasks that require the observation of broadcasted TV programs. Interruption of the broadcasted content, automatically results in the interruption of the task completion by the test participants. This makes the TV content to a certain extent important to the test participants.

### 7.5 Experimental protocol

The previous sections described parts of the approach for the second consumer experiment. This section combines these parts into a final experimental protocol to investigate the relation between FUI and UPFS. Figure 7.3 gives an overview of this final experimental protocol.

The pre-experimental phase consists of the selection of test participants for the experimental phase. In this selection process, the irritableness of potential test participants should be controlled (see subsection 6.3.2). This pre-experimental phase is described in Section 7.6.

The experimental phase starts with the performance of the acquaintance tasks by the test participants. These acquaintance tasks are followed by the completion of a supplemental FUI measurement with a restricted number of TV functions. This FUI measurement is used to verify whether a significant difference exist between the FUI ratings of the two selected TV functions among the test participants after the acquaintance tasks. Because participants can rate the “Motorized swivel” function (unimportant) as more important after using it during the acquaintance tasks (refer to Section 6.1.2), the results of this supplemental FUI measurement are used to divide the participants over the scenarios. In other words, in order to maintain the significant difference in FUI ratings of the two scenario functions, participants of scenario 1 must consider the “Watch the desired program” function important and participants of scenario 2 must consider the “Motorized swivel” function unimportant.
Therefore, participants with a high supplemental FUI rating (3 or higher) for the “Motorized swivel” function always continue the experiment with scenario 1. Other participants are randomly distributed among the two scenarios.

After this supplemental FUI rating, the first PANAS-X measurement is performed to measure the current emotional state of the test participant (see subsection 6.2.2). This PANAS-X-1 measurement is used in the UPFS measurement and to compare the irritableness level of the test participants of both scenarios. Subsequently, the test participants perform the task of scenario 1 or scenario 2. During the execution of this task, the participant is confronted with a failure in the related TV function (scenario 1: “Watch the desired program function”, scenario 2: “Motorized swivel” function). This failure continues for one minute. One minute after the recovery of the TV function, another PANAS-X measurement is performed to measure the current emotional state of the test participant. The difference between this and the previous PANAS-X measurement is considered as a measure for UPFS. In addition, the adjusted UPFS survey also measures the UPFS that is caused by one of the two failure scenarios (see subsection 6.2.1).

![Figure 7.3: Final experimental protocol](image)

Subsequently, after the completion of this UPFS survey, the participant receives a list of TV functions for which he/she should indicate whether they worked flawlessly during the experiment (see Appendix H).
This function failure list is used to check whether the impact of the failures on the functions is equivalent (see Section 6.3.3). In order to control for this failure characteristic, the failure impact scores should be equal for both experimental scenarios.

Lastly, each participant is elaborately debriefed and questioned about the assumed cause(s) of the experienced failure (see Section 6.3.3). This information about the perceived failure cause is used to check whether the Failure Attribution has been successfully controlled.

### 7.6 Participant selection

As explained earlier, only university students are selected as test subjects. Students are a relatively homogeneous group, which minimizes variability within the conditions of the experiment. Participants are selected using a convenience sample of students from the Eindhoven University of Technology. No distinction is made between students from different faculties or educational programs. Students are chosen randomly at the Auditorium Building of this University.

The actual recruitment of test participants took place in week 46 and 47 of 2006. Potential participants were approached and asked for their participation in the FUI consumer experiment. These potential participants were requested to fill in the PANAS-X survey (see Appendix G). This approach resulted in 207 returned PANAS-X surveys. Seven of these surveys appeared to be incomplete (more than 10 missing values) and were therefore excluded from the analysis. This means that in total 200 students successfully completed the PANAS-X survey.

Section 6.3 explained how this PANAS-X survey can be used to control for irritableness in the selection of participants. The results of the PANAS-X validity check (Section 6.2) justify the application of the General Dimension Scale Negative Affect and the Basic Negative Emotion scale for controlling irritableness in this research.

Based on the PANAS-X dataset (N=200), “average irritable” participants are selected for the second consumer experiment. There are no guidelines in literature on the average level of irritableness among people. Therefore, the average irritableness level of the 200 participants is considered the “average irritableness” level of all people.

As mentioned before, the General Negative Affect scale and the Basic Negative Affect scale are used for the selection of the test participants. The “average irritable” participants are considered to be those participants that have General Negative Affect- and Basic Negative Affect- scores within the range of one standard deviation around the means of these scales. In other words, suitable test participants are the ones who have a score on the General Negative Affect Scale of 18.52 (mean) +/- 5.44 (one standard deviation) and who also have a score on the Basic Negative Affect Scale of 9.99 (mean) +/- 2.81 (one standard deviation).

Frequency plots for General Negative Affect and Basic Negative affect are presented in Figure 7.4.
Figure 7.4: Histogram of General Negative Affect and Basic Negative Emotion

The dotted lines in the figure are the lower and upper limits of the acceptance ranges. As mentioned before, the determination of these cut-off values is not discussed in literature. However, by setting the cut-off values of the scales at one standard deviation, it is certain that participants with extreme low or high NA are excluded from this experiment. This is necessary because a disproportionate distribution of “irritableness” levels over the experimental groups biases the experimental results. Moreover, this unequal “irritableness” distribution reduces the likelihood of finding a significant correlation between the independent variable FUI and the dependent variable UPFS. Applying these cut-off values resulted in the selection of 71 participants (out of 200). However, due to resource constraints, only 25 people could be invited for the FUI experiment. The selection of these 25 actual test participants from this bigger group of 71 suitable potential test participants was done completely random.

7.7 Pilot experiment

For evaluating the experimental protocol of Figure 7.3, five pilot experiments were performed. Five additional university students were selected from the group of 71 potential test participants (see Section 7.6) to participate in these pilot experiments. Based on the observations of these pilot experiments and the comments of the test participants, some small adjustments were made to the final experimental protocol:

- The presence of the researcher in the laboratory room should be minimized. In the actual experiment, the researcher is only present in the room during the execution of the acquaintance tasks to help the participant. After that, he/she leaves the room and comes back after the scenario task is finished. This points the attention of the participant to the function failure and not to the researcher. The participant has to deal with the failure him/herself.

- The failure duration should be extended. Two of the five pilot participants did not notice the implemented failure in the TV function. One other participant indicated that the failure was “too short” to be irritating. Therefore, the failure duration was doubled. Instead of one minute, the implemented failures continue for two minutes in the actual experiment.

The other experimental settings worked out well during the pilot experiments and were therefore left unchanged.

\footnote{The 25 participants of the experiment and 5 participants of the pilot experiment received an electronic gadget worth € 25,- as token of appreciation.}

112
7.8 Experimental results

The actual experiments were performed in week 6 till week 8 of 2007. As mentioned above, 25 university students participated in the FUI experiment. Thirteen of these participant underwent scenario 1 (failure in the “Watch the desired program” function) and twelve participants underwent scenario 2 (failure in the “Motorized swivel” function). This section deals with the results of this FUI experiment. In the first subsection, it is investigated whether the approach to control for irritableness by the application of PANAS-X (Section 7.6) has been successful. In other words, subsection 7.8.1 deals with the question: “are the test participants of scenario 1 equally irritable as the test participants of scenario 2?” Subsequently, the second subsection investigates the influence of FUI on UPFS by statistically testing the hypotheses formulated in Section 7.2. Lastly, in subsection 7.8.3 some additional outcomes of the experiment are discussed.

7.8.1 Control for irritableness by applying PANAS-X

This subsection investigates whether the approach to control for irritableness by the application of PANAS-X (Section 7.6) has been successful. Test participants filled in a PANAS-X scheme (Appendix G) just before the start of the failure scenario, see Figure 7.3. This PANAS-X measured the mood of the test participant just before the start of the failure scenario. This mood can be different for each test participant. However, given the applied participant selection approach, the irritableness levels of “scenario 1 participants” should be equal to the irritableness levels of “scenario 2 participants”. In other words, no significant difference should exist between the participant irritableness levels of both scenarios. In order to investigate this assumption, the average values of the scales General Negative Affect, Basic Negative Emotion, Fear, Hostility, Guilt and Sadness of both scenarios are compared, as presented in Figure 7.5.

![Figure 7.5: Examining differences between relevant PANAS-X-1 scales](image)

Care should be taken when interpreting the results with regard to Fear, Hostility, Guilt and Sadness individually, since the discriminant validity of these scales is questionable (see Section 6.2.3). Since the PANAS-X-1 data is not normally distributed, a Mann-Whitney test was used for examining the differences in means of these scales [MON99]. The results of this test are presented in Table 7.2.

<table>
<thead>
<tr>
<th></th>
<th>Fear</th>
<th>Hostility</th>
<th>Guilt</th>
<th>Sadness</th>
<th>General NA</th>
<th>Bas Neg Emo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney</td>
<td>78.00</td>
<td>75.00</td>
<td>54.00</td>
<td>74.00</td>
<td>70.50</td>
<td>77.00</td>
</tr>
<tr>
<td>Significance</td>
<td>1.000</td>
<td>0.894</td>
<td>0.205</td>
<td>0.852</td>
<td>0.689</td>
<td>0.979</td>
</tr>
</tbody>
</table>

These results confirm the assumption that no significant (α = 0.1) differences exist between the moods of “scenario 1 participants” and “scenario 2 participants”.

113
Based on these results, it can be concluded that PANAS-X can be successfully used for the selection of average irritable participants.

In Section 6.2, it is concluded that it is impossible to distinguish between the basic negative emotions Fear, Hostility, Guilt and Sadness. These basic negative emotions correlate to one emotion: Basic Negative Emotion. Table 6.12 showed that the Basic Negative Emotion scale is highly correlated with General Negative Affect (0.916). This correlation indicates that the use of the Basic Negative Emotion scale in addition to General Negative Affect does not add much extra information. Both scales measure nearly the same construct. Therefore, for future applications of PANAS-X in this research, using the General Negative Affect scale alone is recommended.

7.8.2 Influence of FUI on UPFS

As Figure 7.3 indicates, the UPFS of the failure scenarios is measured by two methods:

- **PANAS-X-1 versus PANAS-X-2**: the difference between the user's emotional NA state just before product failure and just after product failure can be considered as a measure for UPFS
- **UPFS survey**: a self-report survey that consists of five statements to measure the level of irritation experienced by the user as caused by the TV failure scenario (see Appendix F)

This subsection investigates the relation between FUI and UPFS by statistically testing the earlier formulated hypotheses:

- \( H_0 \): There is no significant difference in UPFS caused by failures with different levels of function importance
- \( H_1 \): The UPFS that is caused by a failure in an important function is significantly higher than by a failure in an unimportant failure

These hypotheses are tested by using both types of UPFS measurement results (PANAS-X and UPFS survey). Consequently, the results of the statistical tests raise or lower the confidence in the construct validity of (parts of) the theory-based UPFS model presented in Figure 5.2.

PANAS-X to measure UPFS

As mentioned before, a good reliability of the PANAS-X measurement scale is a prerequisite for the correct appraisal of the experimental results. Subsection 6.2.3 presented a comprehensive evaluation of the reliability and validity of the new (Dutch) translation of PANAS-X. This evaluation indicated that it is appropriate to use the General Dimensions Scales Negative Affect and positive Affect, the Basic Negative and Positive Emotion scale, and the scales Joyviality, Self-Assurance, Attentiveness. Subsection 6.2.2 identified the relevant scales for measuring UPFS as: Fear, Hostility, Guilt, Sadness, General Negative Affect and Basic Negative Emotion. However, based on the reliability analysis of PANAS-X, the results of the scales Fear, Hostility, Guilt and sadness should be interpreted with care, because their discriminant validity is questionable.

Before comparing the UPFS scores (\( \Delta(PANAS-X-1 \text{ and } PANAS-X-2) \)) of both scenarios, the ability of PANAS-X to measure UPFS should be investigated. For measuring UPFS, there should be a difference between the scores of PANAS-X-1 and PANAS-X-2 (see Figure 6.2).
In other words, a change in the mood of the test participants should be visible after experiencing the failure scenario. This change in mood can be measured by statistically testing the difference between PANAS-X-1 and PANAS-X-2 for both scenarios together (N=25). A Wilcoxon test should be used to compare the overall means of these two measures [MON99]. This test indicates that there are significant differences ($\alpha = 0.1$) between PANAS-X-1 and PANAS-X-2 in the relevant scales Fear, Guilt and General Negative Affect (see Figure 7.6, $\Delta 2$).

The ability of PANAS-X to measure UPFS should also be examined for the two scenarios separately (N=13) and (N=12). In other words, do both failure scenarios cause a significant change in the mood of the test participants? Again, a Wilcoxon test should be used to compare the means of PANAS-X-1 and PANAS-X-2 for both scenarios separately [MON99]. For scenario 1, only the scale Guilt differs significantly ($\alpha = 0.1$) before and after the TV failure (see Figure 7.6, $\Delta 4$).

In scenario 2, significant ($\alpha = 0.1$) different scores are measured on the scales Fear, Hostility, General Negative Affect and Basic Negative Emotion before and after the TV failure (see Figure 7.6, $\Delta 3$). As mentioned before, UPFS is defined as the differences between PANAS-X-1 and PANAS-X-2 on the scales General Negative Affect and Basic Negative Emotion. The differences between PANAS-X-1 and PANAS-X-2 on these scales are significant for scenario 2, but not for scenario 1.

In other words, the proposed approach to measure UPFS using PANAS-X is unable to measure user dissatisfaction that is caused by the failure in the important “Watch the desired program” function. The inability of PANAS-X to measure UPFS in the first failure scenario makes it impossible to investigate the influence of FUI on UPFS. Therefore, the collected PANAS-X-1 and PANAS-X-2 data cannot be used to statistically test the hypotheses of Section 7.2.
The UPFS survey
In the experiment, a second approach to (try to) measure UPFS was adopted: the UPFS survey (see subsection 6.2.1). This self-report survey consists of five statements to measure the level of irritation experienced by the user as caused by the TV failure scenario (UPFS) and five additional statements to describe the expected reactions of the participant to the experienced failure scenario (see Appendix F). The five irritation statements together form the self-report UPFS measurement scale. Before analyzing the UPFS results of this survey, the reliability of this adjusted UPFS survey should be assessed. Table 7.3 examines the correlations between the items of the UPFS measurement scale. The items of the measurement scale correlate very high with each other. Moreover, the Cronbach's alpha value of the measurement scale is 0.83 indicating that the reliability of this measurement scale is good. This value is well above the threshold value of 0.70 [HAI05]. These analyses confirm the good reliability of the UPFS measurement scale.

Table 7.3: Correlation matrix: irritation items of the UPFS measurement scale

<table>
<thead>
<tr>
<th>Item 1: very irritating</th>
<th>Item 3: get enraged</th>
<th>Item 5: not angry</th>
<th>Item 8: small failure</th>
<th>Item 10: does not bother</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1: very irritating</td>
<td>0.61**</td>
<td>0.51*</td>
<td>0.57**</td>
<td>0.66**</td>
</tr>
<tr>
<td>Item 3: get enraged</td>
<td>0.61**</td>
<td>0.40*</td>
<td>0.53**</td>
<td>0.69**</td>
</tr>
<tr>
<td>Item 5: not angry</td>
<td>0.51*</td>
<td>0.40*</td>
<td>0.53**</td>
<td>0.69**</td>
</tr>
<tr>
<td>Item 8: small failure</td>
<td>0.57**</td>
<td>0.53**</td>
<td>0.42*</td>
<td></td>
</tr>
<tr>
<td>Item 10: does not bother</td>
<td>0.66**</td>
<td>0.69**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Correlation significant at the 0.01 level
* Correlation significant at the 0.05 level

Subsequently, these UPFS measurement results of the UPFS survey can be used to investigate the relation between FUI and UPFS. Table 7.4 presents the UPFS scores of the test participants per failure scenario.

Table 7.4: Individual UPFS scores of test participants per scenario

<table>
<thead>
<tr>
<th>Scenario 1: Watch the desired program</th>
<th>Scenario 2: Motorized swivel</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPFS scores of participants</td>
<td>Average score</td>
</tr>
<tr>
<td>22 17 22 18 15 14 23 18 21 18 22 24</td>
<td>19.50</td>
</tr>
</tbody>
</table>

In scenario 1, there is one participant with a very low UPFS score (9) in comparison with the other participants of this scenario. This outlier can be classified as an extraordinary observation. Extraordinary observations are a class of outliers for which the researchers has no explanation [HAI05]. In the case of this outlier, there are no particular circumstances which are different for this experiment in comparison to other experiments. Nevertheless, this particular experiment resulted in a very low UPFS score. The overall average UPFS score and standard deviation are heavily influenced by this specific UPFS score. Removing this outlier from the data set lowers the UPFS standard deviation of scenario 1 from 4.66 to 3.06. The average UPFS score of scenario 1 increases from 21.08 to 22.08. These results are presented in Table 7.5. Based on the strong influence on the average UPFS score and standard deviation, this outlier is removed from the dataset.
<table>
<thead>
<tr>
<th>Scenario 1: Watch the desired program</th>
<th>UPFS scores of participants</th>
<th>Average score</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 25 22 22 25 23 24 23 15 20 25 23</td>
<td>22.08</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>Scenario 2: Motorized swivel</td>
<td>22 17 22 18 15 14 23 18 21 18 22 24</td>
<td>19.50</td>
<td>3.26</td>
</tr>
</tbody>
</table>

For this dataset (N=24), a non parametric test is performed to test the hypotheses of Section 7.2. The Mann-Whitney U test is considered most suitable for this purpose [MON99]. Table 7.6 presents the test results of this Mann-Whitney U test.

<table>
<thead>
<tr>
<th>UPFS score</th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
<th>Asymptotic Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>36.00</td>
<td>114.00</td>
<td>-2.10</td>
<td>0.036</td>
<td></td>
</tr>
</tbody>
</table>

The test shows a significance value of 0.036 which is lower than 0.05. This indicates that there is a significant difference between UPFS of scenario 1 and scenario 2. Consequently, based on these results, the null hypothesis (“There is no significant difference in UPFS between failures with difference levels of function importance”) should be rejected. Thus, the positive relationship between FUI and UPFS as presented in Figure 7.1 is demonstrated. However, it is important to mention that the difference in UPFS level is not very significant (>0.01). Although the FUI of a failure influences the UPFS, the significance of this effect is rather low.

Moreover, the comparison of these UPFS survey results with the PANAS-X measurement results demonstrates the inability of PANAS-X to measure UPFS in this and similar experimental designs. A plausible explanation for this inability may be the subtlety of the mood changes in participants that are caused by a product failure in a laboratory environment. In real life, similar failures in by the user owned and paid products may cause far more extreme mood changes than in this controlled setting. Consequently, from now on, the application of PANAS-X in the context of this research is restricted to the selection of average irritable test participants.

### 7.8.3 Other results

Besides the confirmed relation between FUI and UPFS, there are also some other experimental results worth mentioning. These results are discussed in this subsection.

**Failure Impact of the scenarios**

Figure 5.2 indicated that Failure Impact (FI) could influence UPFS. FI is defined as the percentage loss of functionality as a result of the failure (see Table 5.1). The influence of FI on UPFS was not examined in this second consumer experiment. Therefore, in order to limit the influence of FI on UPFS, the FI was controlled (see Section 6.3.3). In the design of the failure scenarios, it was tried to keep the FI of both function failures equal. Whether this control for the FI characteristic was successful can be checked with the function failure list (see Figure 7.3 and subsection 7.5). The average overall score of the FI should be equal for both scenarios. Unsuccessful control of this extraneous variable would bias the experimental results and could invalidate the identified significant relation between FUI and UPFS.
Table 7.7 presents the FI scores for both scenarios. The FI score of scenario 2 is high (3.67) compared with the FI score of scenario 1 (2.36). This difference in FI between the two scenarios is significant (p < 0.05). The effect on UPFS could be that in this experiment the measured UPFS in scenario 1 is too low in comparison with scenario 2. It can be expected that when the FI of scenario 1 increases (up to the FI level of scenario 2), the UPFS level increases as well.

A possible explanation for the unequal FI scores of the failure scenarios is the limited picture quality of the TV, unrelated to the failure scenarios. The picture quality of the television in the “normal situation” (outside the failure scenarios) was considered inadequate by several test participants. This was caused by the signal disrupter that was developed for creating the picture disruption. In the normal situation, this device partly disturbed the signal. Because of that, the picture quality was less than normally could be expected. However, another problem is the size of the TV used in this experiment. This television had a very large picture diagonal (107 cm). For optimal picture quality, the size of this TV requires a minimum viewing distance of three meters. However, the maximum viewing distance in the laboratory living room was just above two meters. This limited distance also caused a limited perceived picture quality. The influence of this limited picture quality on the FI can be detected in Table 7.7. In scenario 2, participants did not only mention the failure in the motorized swivel, but also saw the perceived picture quality as a problem. Although these participants are not confronted with the “Watch the desired program” function failure, they score the FI of this function with 2.33. In the actual “Watch the desired program” function failure scenario (scenario 1), participants perceive the failure with a severity of 2.36 (see Table 7.7). This indicates that the failure in the “Watch the desired program” function did (nearly) not worsen the failure severity.

<table>
<thead>
<tr>
<th>Function</th>
<th>Times mentioned</th>
<th>Average score**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motorized swivel</td>
<td>4</td>
<td>1.25</td>
</tr>
<tr>
<td>Watching desired program*</td>
<td>11</td>
<td>2.36</td>
</tr>
<tr>
<td>Teletext</td>
<td>2</td>
<td>2.50</td>
</tr>
<tr>
<td>Changing picture format</td>
<td>2</td>
<td>1.50</td>
</tr>
<tr>
<td><strong>Scenario 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motorized swivel*</td>
<td>12</td>
<td>3.67</td>
</tr>
<tr>
<td>Watching desired program</td>
<td>6</td>
<td>2.33</td>
</tr>
<tr>
<td>Watching two program</td>
<td>3</td>
<td>2.00</td>
</tr>
<tr>
<td>Teletext</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>Changing picture format</td>
<td>1</td>
<td>3.00</td>
</tr>
</tbody>
</table>

* The function that failed in this scenario
** Average score on a scale from 0 till 5

This inequality in FI could explain the relative small effect size of FUI on UPFS. Therefore, this inability to control for FI does not invalidate the presented relation between FUI and UPFS. Consequently, in future consumer experiments, the relation between FUI and UPFS is expected to be even more significant, when the FI is successfully controlled.
The influence of causal attribution

A second group of additional results again relates to the assumed causes of the TV failures by the test participants. The influence of failure attribution on UPFS was not examined in this second consumer experiment. Therefore, in order to limit the influence of failure attribution on UPFS, the failure attribution was controlled. In the design of the failure scenarios, both failure scenarios were designed to be internally attributed. Unsuccessful control of this extraneous variable would bias the experimental results and could invalidate the identified significant relation between FUI and UPFS.

Whether this control for the failure attribution characteristic was successful is checked in the debriefing phase. The debriefing phase of the experiment started with asking the participants to describe the cause of the TV failure. Analysis of these causal attribution results indicates that scenario 1 participants, who experienced a failure in the “Watch the desired program” function, did not blame the television for the degraded picture quality. Table 7.8 presents some of the responses of these scenario 1 participants to the question “What caused the failure in the TV function?” Scenario 1 participants attributed the cause of the TV failure in most cases (10 out of 13) to factors other than the TV itself (e.g. the cable company or the weather). Conversely, all scenario 2 participants (12 out of 12), who experienced a failure in the “Motorized swivel” function, attributed the cause to the TV itself.

Table 7.8: Some scenario 1 causal attribution results

<table>
<thead>
<tr>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I don't think that the television is the problem, probably the signal is bad”</td>
</tr>
<tr>
<td>“Probably it isn't the television that causes the bad picture”</td>
</tr>
<tr>
<td>“There was a short disruption; the problem has to do with the weather I guess”</td>
</tr>
<tr>
<td>“The signal is bad”</td>
</tr>
<tr>
<td>“The fine-tuning is not good enough”</td>
</tr>
<tr>
<td>“The antenna is not properly connected”</td>
</tr>
<tr>
<td>“Maybe the cable is the problem”</td>
</tr>
</tbody>
</table>

Circumstances related to a frustrating incident and the situation of the person itself can lead to a difference in level of frustration experienced by that person [LAZ06]. Users will have a high commitment to a goal when the goal is important to them and it is believed that the goal can be reached [LOC96]. The level of frustration that people experience is influenced by how important the goal is to them, as well as how confident they are in their abilities (self-efficacy) [LAZ06A]. Self-efficacy can be considered as how well a task can be performed when it involves setbacks, obstacles and failures [BAN86].

In scenario 1, the goal “make an overview of the broadcasted channels and programs on TV” could be partly completed, since some channels were still partly visible despite the disturbed TV image. It was still possible to recognize several programs. Whereas in scenario 2, the goal “turn the TV back to the couch position” could not be completed.

There are also other factors that may influence the level of frustration. Not all obstructions are equally frustrating [LAZ06]. People can perceive a failure as being justified by general social rules; in this case, the frustration response may be minimized [BAR07]. Extra information available to the individual may reduce the expectations. Most people have experienced a disruption in the TV cable signal before.
It is generally accepted that these disruptions can happen every now and then. This means that the failure of scenario 1 can be associated with this socially accepted phenomenon of a TV cable disruption. According to Baron, this association leads to a lower level of frustration [BAR77]. In case of the “Motorized swivel” function failure, it is not generally accepted that the motorized swivel fails once in a while. Therefore, the influence of this causal attribution on the expected UPFS of this scenario is higher in comparison with the “Watch the desired program” function failure. This is also supported in the paper by Westbrook and Oliver [WES91]. They state that when an evaluation of a consumer experience (or its associated product) is performed, past experiences and memories such as prior expectancies and disconfirmation beliefs play a role in this evaluation [WES91]. In this second consumer experiment, these prior expectancies could have influenced the causal attribution by the test participants. The traditional expectancy (past usage and buying experience) of a television is that it at least has a “good picture” quality. Therefore, test participants could not believe that, based on their past experience, the bad picture quality was caused by the TV.

Moreover, test participants do not expect that a product such as this expensive TV can fail on such an important function. Individuals are cognitive misers: they are liable to accept the easiest explanation [TAY80], [HAN80]. The easiest explanation for the scenario 1 TV failure is a problem in the TV cable signal and not a failure in the television itself. This difference in failure attribution among the failure scenarios could also have contributed to the relative small effect size of FUI on UPFS.

The possible influence of failure attribution on the experimental results invalids the significant relation between FUI and UPFS. In other words, the measured effect of FUI on UPFS may in fact be caused by the difference in failure attribution between the two scenarios. Therefore, the testing of the research hypotheses requires another consumer experiment in which the influence of failure attribution is distinguished and investigated.

Although the significant positive relation between FUI and UPFS is invalidated by the bias of the failure attribution variable, the results of this experiment still support the expected positive relation between FUI and UPFS. However, the results of the first and second consumer experiment both raise the notion of the probable influence of failure attribution on UPFS (see also subsection 4.3.3). Based on the results of these two experiments, a hypothetical UPFS matrix can be presented that combines the influence of Failure Attribution (FA) and FUI on UPFS. This hypothetical UPFS matrix is presented in Figure 7.7.

![Hypothetical UPFS matrix](image)

*Figure 7.7: Hypothetical UPFS matrix*
In the context of this research, internal failure attribution implies that the TV failure is attributed to the TV itself or the user. External failure attribution means that the failure is considered to be caused by a factor other than the TV or the user. The UPFS of a failure in a low importance function that is internally attributed is expected to be perceived higher than a failure in a high importance function that is externally attributed. Subsequently, a failure in a high importance function that is internally attributed is expected to cause the highest level of UPFS and a failure in a low importance function that is externally attributed is expected to cause the lowest level of UPFS.

This hypothetical UPFS matrix again describes the expected positive relation between FUI and UPFS. However, the influence of FA is added as an extra experimental variable. Subsequently, a logical next step in this UPFS research is the experimental validation of the hypothetical UPFS matrix that combines the influence of FUI and FA on UPFS. Moreover, in case the interaction effect between FA and FUI appears to be absent in the third consumer experiment, the significant relation between FUI and UPFS that was established in the second consumer experiment would also be confirmed. That is, the possible absence of this interaction effect would imply that the lack of FA control in this second consumer experiment did not bias the experimental outcomes.

7.9 Conclusions

The influence of Function Importance (FUI) level on UPFS in the theory-based UPFS prediction model was investigated. A positive relation was expected between FUI level and UPFS. In order to validate this relation, a second consumer experiment was performed.

The results of this second consumer experiment support the positive relation between FUI and UPFS. Statistical tests confirm that a failure in an important function causes higher levels of user dissatisfaction than a failure in an unimportant function. However, the unsuccessful control of the extraneous FA variable invalidates this experimental result. In a proposed third consumer experiment, this relation between FUI and UPFS is expected to be present together with the predicted relation between FA and UPFS. Moreover, in case the interaction effect between FA and FUI appears to be absent in this third consumer experiment, the significant relation between FUI and UPFS established in the second consumer experiment would also be confirmed.

Furthermore, the size of the FUI-UPFS effect is rather small. These results might indicate that the importance of a failing function is less determinant for user dissatisfaction than was expected beforehand. In other words, a user is most dissatisfied by the fact that the product fails in general. Which specific function fails, is only of limited concern to the user.

This limited effect size might also be (partly) explained by the unequal failure impact of the experimental scenarios. The failure impact of the unimportant failure scenario was considerably higher than the failure impact of the important failure scenario. Furthermore, the cause of the unimportant function was attributed to the TV itself, but the cause of the important function was mostly attributed to external causes. This difference in attribution of both failure scenarios might also have caused the limited influence of the FUI on UPFS in this experiment. A follow-up consumer experiment should investigate the origin(s) of this limited FUI effect size. In this proposed experiment, the possible influence of failure impact and failure attribution on UPFS should be considered explicitly.
Furthermore, the adjustments to the UPFS measurement scale appeared to be quite successful (see Section 7.8.2). The reliability of the scale has been improved substantially (Cronbach's alpha value of 0.83 in this consumer experiment, instead of 0.71 in the previous consumer experiment). Based on these results, the UPFS measurement scale can be indicated as reliable.

In this second consumer experiment, the PANAS-X measurement scale was adopted for the selection of average irritable test participants and as additional UPFS measurement. Statistical analysis confirmed the validity and reliability of parts of the Dutch version of PANAS-X (see Section 6.2.3). The results of this analysis indicate that it is appropriate to use the General Dimension scales Negative Affect and Positive Affect, the basic Negative and Positive Emotion scales and the scales Joviality, Self-Assurance and Attentiveness. However, care should be taken when using Fear, Hostility, Guilt and Sadness, because the discriminant validity is questionable. The first application of PANAS-X for the selection of average irritable test participants generates good results. Comparison of the scenarios indicates that no significant (α = 0.1) difference exists between the mood of “scenario 1 participants” and the mood of “scenario 2 participants”. Based on these results, it can be concluded that PANAS-X can be successfully used for the selection of average irritable participants.

The application of PANAS-X for the additional measurement of UPFS appeared less successful. The proposed approach to measure UPFS using PANAS-X is unable to measure user dissatisfaction that is caused by the failure in the important “Watch the desired program” function. This inability of PANAS-X to measure UPFS in this failure scenario makes it impossible to use this data to investigate the influence of FUI on UPFS. Consequently, from now on, the application of PANAS-X in the context of this research is restricted to the selection of average irritable test participants.

Another lesson learned from this second consumer experiment is that the above mentioned conclusions are based on a relatively small sample size (N=25). Moreover, in order to control for user characteristics in this second consumer experiment, the experimental sample consisted of a homogeneous group of university students. Therefore, generalization of the results should be done with care. Consequently, the presented results are an indication of the UPFS that is caused by a TV failure.

Lastly, it is important to mention that the laboratory environment in which the experiments took place may have influenced the behavior of the test participants. Especially the unrealistic aspects of the laboratory environment limit the external validity of the experimental results. Every person or group being studied in a laboratory experiment will behave different in the presence of researchers [CHR04].

Based on the results of this second consumer experiment and complemented with relevant literature, the next chapter investigates the combined influence of FUI and Failure Attribution on UPFS. This influence is evaluated in a third consumer experiment in which the impacts of the failure scenarios are kept as constant as possible. The basis for this third consumer experiment is the hypothetical UPFS matrix of Figure 7.7. This third experiment contributes to the further validation of the theory-based UPFS prediction model.
The previous chapter described the development of an UPFS prediction model in consumer electronics. The results of the second consumer experiment confirmed the positive relation between Function Importance (FUI) and UPFS. However, this significant relation was biased by the influence of Failure Attribution (FA). Moreover, the effect size of the relation appeared to be rather small. The results of this second consumer experiment may therefore indicate that the importance of a failing function is not determinant. This limited effect size may also be (partly) explained by the unequal Failure Impact of the experimental scenarios and the difference in Failure Attribution of both failure scenarios.

Based on these insights, this chapter aims to further validate the theory-based UPFS prediction model. Consequently, this chapter evaluates the combined influence of FUI and Failure Attribution on UPFS based on a new experimental model.

In the first section of this chapter, the variables of this new experimental model are described in more detail. Subsequently, the actual influence of the Failure Attribution variable and the FUI variable on UPFS is investigated in a third consumer experiment. The second and third sections of this chapter describe respectively the design and results of this experiment. In the last section, based on the experimental results, some conclusions are drawn about the influence of Failure Attribution and Function Importance on UPFS and the general validity of this part of the theory-based UPFS model.

### 8.1 Influence of Failure Attribution on UPFS

Because of the possible influence of Failure Attribution, the results of the previous chapter could not irrefutably validate the positive relation between Function Importance (FUI) and UPFS. Nevertheless, the results of the second consumer experiment indicated that the limited significance of this relation may be (partly) explained by the difference in Failure Attribution (FA) of both failure scenarios.

Subsequently, the expected influence of FA on UPFS was presented in the hypothetical UPFS matrix of Figure 7.7. The implication of this UPFS matrix can also be represented in an updated experimental model: the experimental model of the FA experiment (see Figure 8.1 on the next page). This model illustrates the expected positive relation between FUI and UPFS complemented with the hypothesized relation between FA and UPFS. The symbols in this figure indicate that internally attributed failures are expected to cause higher UPFS levels than externally attributed failures. Based on this experimental model, the relations between FUI and UPFS and between FA and UPFS are evaluated in a third consumer experiment. However, before these relations can be investigated in an actual consumer experiment, the experimental variables need to be described in more detail. Therefore, the next subsections present the description of these variables.

* Currently, the material of this chapter is being prepared for publication in an academic journal.
8.1.1 Independent variable: Failure Attribution

In Chapter 5 (see Table 5.1), Failure Attribution (FA) was defined as the cause to which users attribute the failure. This definition indicates that FA is related to the concept of causal search or explanation [BER07]. Anything unusual which stimulates individuals' attention to the outcome will lead to causal search [OLI96]. Disconfirmation of expectations will cause attribution processing.

Heider was one of the first to explain individuals' reactions to other individuals or the environment, such as the weather or economic conditions [HEI58]. As already mentioned in the previous chapter, individuals are cognitive misers: individuals are prone to accept the easiest explanation for events [HAN80], [TAY80]. In their judgment, individuals are biased in the causal attribution of the assessed outcome. This phenomenon is called “the fundamental attribution error” [HEW89], [OLI96]. This fundamental attribution error implies that there is pervasive tendency for attributers to ignore the impact of situational factors and to attribute behavior exclusively to the behavior of people [JON72], [HEW89]. Attributions to situational events are called situational attributions. Attributions to the behavior of people are called dispositional attributions [HEW89]. Positive outcomes of actions are perceived with the perception of egocentric bias and lead to self-serving attributions [OLI96]. This implies that, in general, individuals attribute their success to their own characteristics. On the other hand, the successes of other people are mainly attributed to situational factors. This tendency reverses when considering the attribution of negative events. To protect their ego, individuals ascribe their failure to situational reasons and others' failure to the personal characteristics of the involved people [OLI96].

Weiner has elaborated on Heider's original work. Weiner developed an attribution framework consisting of three dimensions [WEI85]:

- **Locus of causality**: the result of an action is felt to depend on two sets of conditions, namely, factors within the person and factors within the environment (dispositional versus situational attributions) [HEI58]. Weiner rather speaks about the difference between “internal” and “external” attribution.

- **Stability**: some causal factors are perceived as a constant influence (for example, ability or geographical conditions), but other causal factors are perceived more variable (for example, personal effort or the weather).

- **Controllability**: some causal factors are perceived to be under volitional or optional control (for example, laziness), while other factors are not (for example, gift for math).
The stability dimension of the attribution framework by Weiner shows some overlapping with the theory-based UPFS model (see Figure 5.2). This dimension relates to the failure characteristics Failure Reproducibility and Failure Frequency. A stable (constant) failure cause is completely predictable and consequently results in reproducibility of a product failure. Moreover, a stable failure cause is constant (always present) and therefore causes a high Failure Frequency under standardized use conditions. These examples indicate that the stability dimension of a failure is interrelated with some of the failure characteristics of the UPFS model. Accordingly, in the proposed FA experiment the stability of the failure cause should be controlled for (kept constant).

The controllability of the failure cause indicates the degree to which the cause of the product failure can be affected. Some product failures are caused by deliberate product abuse or misuse by the user. In this case, the controllability of the failure cause by the user is self-evident. However, in almost all other cases of product failure, the determination of the controllability of the failure causes is inclined to result in a rather philosophical discussion about this cause of the failure. The thrust of this discussion can be summarized by the statement that any product failure could have been prevented by either "the development of a better product" or "the instruction of the ignorant user". Even if the controllability of a failure is considered from the user perspective, the contribution of this third attribution dimension to the overall explanation of failure attribution is rather unclear. Therefore, in the context of this research, all product failures are considered uncontrollable by the user. In other words, in the proposed FA experiment the controllability of the failure cause should be kept constant.

The meaning of the locus of the causality dimension however, largely corresponds to the description of the FA variable in the theory-based UPFS prediction model. The results of the first two consumer experiments indicate that test participants tend to attribute the cause of product failures externally. This agrees with the earlier described tendency of individuals to ascribe failures to situational reasons and not to their own characteristics [OLI96]. However, both scenarios also raise the notion of the probable influence of FA on UPFS (see Sections 4.3.3 and 7.8.3). Subsequently, the locus of control dimension can be used to define the two levels of this FA variable in this research. The two levels of FA are defined as:

- **Intern**; the cause of the failure is attributed to the product itself or the user. In the product use process, the product and user are considered to be in constant interaction with each other. Therefore, the integration of the product and user as potential failure cause is considered the intern FA level.

- **Extern**; the failure is considered to be caused by a factor other than the product or the user.

The results of the first two consumer experiments indicate that externally attributed failures are expected to cause lower UPFS levels than internally attributed failures. Validating this influence of the independent FA variable on the dependent UPFS variable requires the (controlled) variation of the FA level and the measurement of its effect on UPFS. Based on the above information, this controlled variation of the independent FA variable can be accomplished by selecting TV functions with different failure attributions (internal or external).

* However, in future research this intern FA level could be subdivided into an intern-user level and intern-product level to investigate their influence on the UPFS.
In this experiment, test participants will be guided to make the specific failure attribution of the involved failure scenario (internal or external attribution). This is done by providing them with textual information about the cause of the presented failure. However, this guided failure attribution is different from independent failure attribution. The experimental results of this guided attribution experiment are therefore likely to be different from experimental results with independent failure attribution by test participants. Nevertheless, the increasing discrepancy between user’s mental models of current consumer electronics products and the complex technical architectures of these products results in guided attribution supported by the product itself. Nowadays, many products give feedback on the possible cause of their disfunctioning to the user (intern: e.g. software update and error numbers or extern: e.g. bad TV signal). The research approach of this FA experiment therefore relates to this trend of guided attribution in practice.

8.1.2 Independent variable: Function Importance

Function Importance (FUI) is the relative importance of the function affected by the failure. The results of the FUI consumer experiment already confirmed the positive relation between FUI and UPFS. However, the difference in FA of these two scenarios invalidated the experimental result. Moreover, the effect size of this relation between FUI and UPFS appeared to be rather small. Therefore, this third consumer experiment tries to validate the relation between FUI and UPFS and to explain the limited effect size by evaluating the combined influence of FUI and FA on UPFS. An elaborated description of the independent variable FUI can be found in Chapter 5 (Section 5.2.1). Moreover, Chapter 6 (Section 6.1.2) presented a validated FUI ranking of the TV functions. Investigating the influence of the independent FUI variable on the dependent UPFS variable requires the (controlled) variation of the FUI level and the measurement of its effect on UPFS. Also, in this third consumer experiment, the FUI ranking of TV functions enables the variation of the independent FUI variable by selecting TV functions with different FUI levels. The next section deals with the measurement of the dependent UPFS variable.

8.1.3 Dependent variable: UPFS

An elaborate description of the dependent variable UPFS was already presented in Chapter 3 (Section 3.4). In the design of the FUI consumer experiment, two different UPFS measurement approaches were applied:

- The PANAS-X scheme
- The self-report UPFS survey

The results of this experiment demonstrated the inability of PANAS-X to measure UPFS. Consequently, the application of PANAS-X in the context of this research is restricted to the selection of average irritable test participants (see Section 7.8.2).

Nevertheless, the results of the FUI consumer experiment indicated that the items of the self-report UPFS measurement scale correlate very high with each other and that the reliability of this measurement scale is good (Cronbach’s alpha: 0.83). Moreover, the results of this self-report UPFS survey confirmed the positive relation between FUI and UPFS. In other words, the experimental results demonstrated the ability of this survey to measure UPFS. Based on these results, only the self-report UPFS survey is selected for the measurement of the dependent UPFS variable in the proposed FA experiment.
8.1.4 Extraneous variables

The proposed FA consumer experiment investigates the influence of FA and FUI on UPFS. However, in an experimental context, this influence may be biased by the interference of extraneous variables. In order to limit the influence of these variables on the dependent variable (UPFS), the extraneous variables should be controlled.

The results of the FUI consumer experiment demonstrated the successful control of the following extraneous variables:

- **The use environment**: all the experiments were executed in a laboratory environment to improve the experimental control
- **The use approach**: the experimental protocols were completely standardized and consequently minimized the difference between the experimental use conditions of participants
- **Irritableness**: participants were selected that revealed an average level of irritableness using PANAS-X. The experimental results demonstrated that PANAS-X can be successfully used for the selection of average irritable participants (see Section 7.8.1)
- **Other use characteristics**: equivalent experimental groups were created by using only university students and random assignment
- **Other failure characteristics**: in the definition of the failure scenarios, the other characteristics of the failures should be equal for both scenarios

The successful control of these extraneous variables in the previous experiment support the repeated application of these methods in the proposed FA experiment. Nevertheless, the analysis of the experimental results also indicated one extraneous variable (failure characteristic) of which the experimental control was less successful: Failure Impact (FI). In the FUI experiment, the difference between the FI of the two experimental scenarios was significant and might have caused the relative small effect size of the FUI-UPFS relation. Therefore, in the proposed FA experiment the FI should be better controlled.

A significant difficulty in the control of FI is the absence of a universal FI measurement scale. In other words, it is difficult to compare and tune the impact of failures in different product functions. In this context, only two universal FI levels can be defined:

- **No failure FI-level**: the product function can be used without problems (no loss of functionality)
- **Full failure FI-level**: the product function cannot be used at all (complete loss of functionality)

Since the proposed FA experiment investigates the influence of different failure scenarios on UPFS, the no failure FI-level is not applicable in this context. Therefore, the successful control of the FI variable in the FA experiment requires the complete disruption of the selected functions in the failure scenarios (full failure FI-level).

Lastly, it is important to mention that the stability and controllability of the different failure scenarios should be kept constant by selecting failure scenarios that are outside the control of the test participant and are equally stable (see Section 8.1.1).
8.2 Experimental design

In order to investigate the influence of FA and FUI on UPFS, a third consumer experiment is performed. The design of this so-called FA consumer experiment is based on the experimental model of Figure 8.1 and is described in this section.

8.2.1 Overview of the FA experiment

The experimental design forms the basis of the third FA consumer experiment. Based on the expected influence of sequence effects, the between-subject design is again selected for this FA experiment (see Section 7.1). Moreover, a $2 \times 2$ full factorial design is selected to investigate the influence of FA and FUI on UPFS. In other words, in this between-subject design, participants are placed in one of the four experimental groups:

- **Group 1**: Test participants are confronted with a failure in a high importance function that is internally attributed
- **Group 2**: Test participants are confronted with a failure in a high importance function that is externally attributed
- **Group 3**: Test participants are confronted with a failure in a low importance function that is internally attributed
- **Group 4**: Test participants are confronted with a failure in a low importance function that is externally attributed

Subsequently, the UPFS levels of these four participant groups are statistically compared in order to investigate the influence of FA and FUI on UPFS.

8.2.2 Hypotheses

The starting point for investigating the influence of FA and FUI on UPFS is presented in Figure 8.1. This figure illustrates the anticipated positive influences of FUI and FA on UPFS. In other words, a higher importance of the failed function and internal attribution of the failure cause are expected to result in a higher level of user dissatisfaction. These expectations can be translated into two groups of null and alternative hypotheses; one for the influence of FUI on UPFS and one for the influence of FA on UPFS.

The null hypothesis and alternative hypothesis for the relation between FUI and UPFS are again:

| $	ext{H}_0$: | There is no significant difference in UPFS caused by failures with different levels of function importance |
| $	ext{H}_1$: | The UPFS that is caused by a failure in an important function is significantly higher than by a failure in an unimportant failure |

The null hypothesis and alternative hypothesis for the relation between FA and UPFS are:

| $	ext{H}_0$: | There is no significant difference in UPFS caused by internally or externally attributed failures |
| $	ext{H}_1$: | The UPFS that is caused by an internally attributed failure is significantly higher than by an externally attributed failure |
Moreover, the influence of FA on UPFS is expected to be stronger than the influence of FUI on UPFS. The expected UPFS levels for the different experimental scenarios are again presented in the hypothetical UPFS matrix of Figure 8.2.

![Failure Attribution Matrix](image)

### Figure 8.2: Hypothetical UPFS matrix

Before this consumer experiment can be performed, the actual design of the experiment should be worked out. One of the limitations of the previous consumer experiment is the limited sample size (N=25) on which the conclusions are based. Therefore, the next subsection deals with the minimum required sample size for this third consumer experiment.

#### 8.2.3 Lesson learned: sample size

One important limitation of the previous consumer experiment is the limited sample size (N=25) on which the conclusions are based. For testing the hypotheses of Section 8.2.2, a 2×2 full factorial experimental design is performed. Section 8.2.1 described the four combinations of the independent variables in this experimental design (FUI: high and low importance function, FA: internal and external attribution of the failure cause). The minimum recommended sample size for this third consumer experiment depends on: the uncertainty level alpha (\(\alpha\)), the power (1-\(\beta\)), the effect size (\(f\)) and the number of groups (\(k\)). The general recommended level for alpha is 0.05 and for the power is 0.80. The number of groups in this third consumer experiment is 4 (see Figure 8.2). The data of the second consumer experiment can be used to calculate the effect size (see Table 7.5). However, this data only related to two of the four experimental groups (see Section 5.4.3):

- **Group 1**: High function importance and internal failure attribution
- **Group 2**: High function importance and external failure attribution

The data of this second consumer experiment are used for calculating the effect size. This effect size is determined using one-way ANOVA: one independent variable (FUI) and one dependent variable (UPFS). One of the assumptions underlying one-way ANOVA is that the data of the dependent variable are normally distributed. This is not the case for the UPFS data of the second consumer experiment. However, even when the normality assumption is violated, the one-way ANOVA yields reasonable results when applied to moderate or larger sample size data [GRE05].

Therefore, the indication of the expected effect size is calculated with the data from Table 7.5. These data indicate that the mean difference between the UPFS levels for the two scenarios is 2.58 (22.08-19.50). The standard deviations of scenario 1 and 2 are respectively 3.06 and 3.26. Subsequently, the expected effect size can be calculated using Cohen's \(f\). The calculation of this effect size is presented in Figure 8.3.
Based on these data, the estimate of the (FUI-UPFS) effect size is 0.38. According to Cohen this effect size of 0.38 can be considered large (0.40 = large, 0.25 = medium, 0.10 = small) [GRE05]. Subsequently, using this effect size of 0.38, alpha of 0.05 and a power of 0.80, a simple statistical program called “Power and Precision” is applied to calculate the minimum recommended sample size for the proposed 2×2 factorial experimental design. This minimum recommended sample size is N = 56 (28 cases per level of the independent variables). Altogether, based on the results of the previous consumer experiment, the minimum required sample size for the third proposed consumer experiment is N = 56. However, if the effect size of the FUI-UPFS relation or the FA-UPFS relation is lower than 0.38, a higher experimental power is required to establish these relations. This higher experimental power can be attained by increasing the sample size per experimental group. Consequently, a larger sample size than this minimum required number of N = 56 is aimed for in this third consumer experiment.

However, due to time and resource constraints, this minimum number of required test participants cannot be achieved with the original laboratory approach as applied in the FUI experiment. Therefore, an alternative experimental approach is selected for this third consumer experiment. Instead of confronting participants with the failures in an actual use situation (in the laboratory), the failure scenarios are presented to participants by means of a short demo movie on a laptop. Subsequently, the participants are asked to evaluate the presented failure scenario. Figure 8.4 shows the proposed experimental setup.

This approach has several advantages. In the first place, this approach is less time consuming. Playing the demo movie, including an introduction to the TV functionalities, takes 7 minutes. Furthermore, the evaluation of the failure scenario takes another 5 minutes. So within 12 minutes, one complete test can be performed. Secondly, the required equipment (only a laptop) offers a lot of flexibility with respect to the possible locations of the experiment. Since the scenarios are only presented in a demo movie, no TV is involved in the actual experiment. This makes it possible to choose an experimental location close to potential participants (e.g. the canteen of the university) which makes the recruitment of test participants rather easy.

---

\[ f = \text{effect size} \quad f = \frac{\sigma_{\mu}}{\sigma} \quad \text{where} \quad \sigma_{\mu}^2 = \frac{\sum_{i=1}^{k} n_i (\mu_i - \mu)^2}{N} \]

\[ \sigma_{\mu}^2 = \frac{12 \times (19.50 - 20.79)^2 + 12 \times (22.08 - 20.79)^2}{24} = 1.66 \quad f = \sqrt{\frac{1.66}{3.36}} = 0.38 \]

\( \sigma_{\mu} = \text{standard deviation of the effect} \quad \mu = \text{mean of group} \ i \)

\( \sigma = \text{standard deviation of the population} \quad \mu = \text{mean of the total sample size} \)

\( k = \text{number of groups} \quad N = \text{total sample size} \)

\( n_i = \text{sample size of group} \ i \)

**Figure 8.3: Calculation of the effect size using Cohen’s f**

---

*Power and precision is a software tool to perform power analysis. A power analysis is used to anticipate the likelihood that the experiment will yield a significant effect and is based on the same factors as the significance test itself. The power and precision tool is developed by BioStat.*
Nevertheless, this proposed experimental approach also has a potential drawback. Test participants do not actually use the TV functions and do not experience the failure scenario in the experiment. Instead, they only observe the functionalities and the implemented failures in a demo movie. Consequently, the application of this approach may reduce the measured UPFS among test participants that is caused by the failure scenarios. In other words, the expected effect sizes of the FUI-UPFS and FA-UPFS relations in such an experiment are lower than the expected effect sizes of a similar test in a laboratory environment. The previous subsection illustrated that a relation with a lower effect size can still be established by increasing the sample size of the experiment. Therefore, the minimum required number of 56 test participants is considered as the absolute lower limit; a higher number of test participants is aimed for. In addition, the measured effect size of both relations in the proposed experiment is probably an underestimation of the actual effect size of these relations in actual use conditions. Nonetheless, this approach makes it possible to investigate the influence of FUI and FA on UPFS with a rather large group of test participants. Moreover, instead of rewarding all test participants with a present, a lottery is organized among the test participants. The winner of the lottery receives a LCD TV worth €500,-. This makes it possible to attract more participants at the same experimental costs.

### 8.2.4 Experimental variables

The proposed $2 \times 2$ factorial experimental design consists of two independent variables (FUI and FA), one dependent variable (UPFS) and several extraneous variables. The first section described different approaches for the variation, measurement and control of these experimental variables. This section gives a short overview of the application of these variables in the proposed consumer experiment.

**Independent variables: FUI and FA**

The previous chapter used a FUI ranking of TV functions to identify an important and unimportant TV function for the variation of FUI among the experimental scenarios (see Section 7.3.1). For the proposed FA experiment, these same TV functions will be used for the definition of two levels of the independent FUI variable in the failure scenario design, namely:

- **Important function:** “Watch the desired program”
- **Unimportant function:** “Motorized swivel”
The first section of this chapter explained that the locus of control dimension can be used to define the two levels of the independent FA variable in this research (see Section 8.1.1). The two levels of FA are defined as:

- **Intern**: the cause of the failure is attributed to the product itself or the user. In the proposed FA experiment, the failure is explicitly attributed to the TV.
- **Extern**: the failure is considered to be caused by a factor other than the product or the user. In the proposed FA experiment, the failure is explicitly attributed to the cable or electricity company.

The combination of these two levels of both independent variables results in the four failure scenarios of the hypothetical UPFS matrix (see Figure 8.2).

**Dependent variable: UPFS**

UPFS is the level of irritation experienced by the user caused by a product failure (see Section 3.4). In the experimental model of Figure 8.1, UPFS is considered the dependent variable that is influenced by FUI and FA. Section 8.1.3 already explained that in the proposed FA experiment only the self-report UPFS survey is selected for the measurement of the dependent UPFS variable.

**Extraneous variables**

Section 8.1.4 described several approaches for controlling the relevant extraneous variables in this research. In order to limit the influence of the extraneous variables on the dependent variable (UPFS), all these control approaches are applied in the third consumer experiment.

**8.2.5 Experimental tasks**

The previous subsection gave an overview of the application of the experimental variables in the third consumer experiment. For investigating the influence of FUI and FA on UPFS a 2×2 factorial design is adopted for which two levels of the independent variables are defined. Based on these results, this section describes the development of the failure scenarios for the proposed consumer experiment.

Based on the two independent variables (FUI and FA), four different failure scenarios are developed. For each scenario, a different movie is made. Each movie starts with the same introduction of the TV functions. In this introduction movie, several TV functions are demonstrated in a movie presenting the functioning of an actual TV. This movie is supported with on-screen textual information about the functioning of the TV. After these introduction movies, the participant is confronted with one of the four failure scenario movies. In these failure scenario movies, one of the two selected functions fails (“Watch the desired program” or the “Motorized swivel” function) and the failure cause is explicitly presented to be internal or external. The four scenario movies present the following failures:

- **Scenario 1** (High FUI, internal FA): a software failure in the TV is causing the complete degradation of the TV image (black screen) and sound (no sound)
- **Scenario 2** (High FUI, external FA): a disruption of the cable signal that is brought on by the cable company is causing a complete degradation of the TV image (black screen) and sound (no sound)
- **Scenario 3** (Low FUI, internal FA): a software failure in the TV is causing the complete break down of the motorized swivel function
Scenario 4 (Low FUI, external FA): a disruption in the power supply that is brought on by the electricity company is causing a complete break down of the motorized swivel function

All movies are recorded in the laboratory living room presented in Figure 4.4 using the same TV and the same implemented failures as in the FUI consumer experiment (see Section 7.4). The explanations given to the participants during the movies about the TV functions, the implemented failures and the failure causes are presented in textual form. The text is made visible on the screen during the movie (appearing in text boxes) or presented in slides before and after the movie. Figure 8.5 shows an example of the FA explanation (in Dutch) given to participants for scenarios 1 and 3 (internal attribution) after the failure scenario movie.

![FA explanation given to participants in scenarios 1 and 3 (internal attribution)](image)

8.2.6 Experimental protocol

The previous sections all described parts of the approach for the third consumer experiment. This subsection combines these parts into a final experimental protocol to investigate the influence of FA and FUI on UPFS. Figure 8.6 gives an overview of this final experimental protocol.

First, test participants are given a short introduction to the experiment. Participants are requested to sit down behind the laptop. Subsequently, the experimental set-up (laptop and headset) is explained and the experiment is started.

The actual experiment starts with a PANAS-X measurement that is performed to measure the current emotional state of the test participant (see subsection 6.2.2). This PANAS-X measurement is used to compare the irritability level of the test participants of both scenarios.

Subsequently, an introduction movie is presented to the test participant. In this introduction movie, several TV functions are demonstrated on an actual TV. This movie is followed by the completion of a supplemental FUI measurement with a restricted number of TV functions. This FUI measurement is used to verify whether a significant difference exists between the FUI ratings of the two selected TV functions among the test participants.
Because participants can rate the “Motorized swivel” function (unimportant) as more important after seeing it in the introduction movie (see Section 6.1.2), the results of this supplemental FUI measurement are used to divide the participants over the scenarios. To maintain the significant difference in FUI ratings of the two scenario functions, the participants who give a high FUI score for the “Motorized swivel” function always test a high FUI failure scenario (scenario 1 or 2).

Figure 8.6: Final experimental design

After this supplemental FUI rating, the participants are confronted with the failure scenarios in the corresponding scenario movies. Directly after the failure scenario movie, the UPFS survey measures the UPFS that is caused by one of the four failure scenarios.
This UPFS survey is followed by a so-called Failure Attribution questionnaire. This questionnaire is added to check whether the participant perceived the failure cause similar to the presented failure cause in the experimental scenario (internal or external). This questionnaire is based on Russell's Causal Dimension Scale [RUS82]. This scale has been well received in literature and has a good reliability (Cronbach's alpha: 0.867). For this research, only the “locus of causality” dimension of this Causal Dimension Scale is of interest (see Section 8.1.1). This dimension consists of three items and the Dutch version is presented in Appendix L.

Subsequently, after the completion of this FA questionnaire, the participant receives a digital list of TV functions on the laptop for which he/she should indicate whether they worked flawlessly during the experiment. This list is very similar to the Function Failure list presented Appendix H. Only the explanation of this Function Failure list is adjusted to fit the description of the FA experiment. The function failure list is used to check whether the impact of the failures on the functions is equivalent (see Section 6.3.3). In order to control for this failure characteristic, the failure impact scores should be equal for all four experimental scenarios. Lastly, each participant is debriefed and informed about the purpose of the research.

8.2.7 Participant selection

As explained earlier, only university students are selected as test subjects. Students are a relatively homogeneous group, which minimizes variability within the conditions of the experiment. Participants are selected using a convenience sample of students from the Eindhoven University of Technology. No distinction is made between students from different faculties or educational programs. Students are picked out randomly at the Auditorium Building of this University. Participants that already participated in the previous FUI consumer experiment could not participate in this third follow-up experiment. The actual recruitment of test participants took place from week 22 till week 24 of 2007. Potential participants were approached and were asked for their participation in the FA consumer experiment. This approach resulted in 149 test participants who took part in this third consumer experiment. The distribution of these participants over the four failure scenarios is presented in Figure 8.7.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest</td>
<td>Low</td>
</tr>
<tr>
<td>UPFS</td>
<td>UPFS</td>
</tr>
<tr>
<td>N=35</td>
<td>N=38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher</td>
<td>Lowest</td>
</tr>
<tr>
<td>UPFS</td>
<td>UPFS</td>
</tr>
<tr>
<td>N=37</td>
<td>N=39</td>
</tr>
</tbody>
</table>

However, a disproportionate distribution of irritableness levels among these participants may influence the relation between the independent variables (FA and FUI) and UPFS. In contrary to the previous FUI experiment, in this FA experiment average irritable test participants were selected after all experiments had been performed.

Figure 8.7: Number of participants in each scenario
In other words, students were invited to join the FA experiment and only afterwards the data of extreme (high or low irritableness level) participants were removed from the complete dataset. This selection approach of test participants results in UPFS data collection from average irritable and extreme (high or low) irritable test candidates. Although the data of the extreme participants cannot be used for testing the experimental hypotheses, this data can be used to evaluate the influence of irritableness on UPFS. In other words, by comparing the UPFS levels of these different “irritableness groups”, it can be checked whether the level of irritableness (measured by NA) really influences the dependent variable UPFS. Based on this comparison, the necessity of controlling for irritableness can be validated.

Again, the PANAS-X measurement data was used for the selection of average irritable test participants. Subsection 6.3.2 explained how this PANAS-X survey can be used to control for irritableness in the selection of participants. The results of the PANAS-X validity check (Section 6.2.3) justify the application of the General Dimension Scale Negative Affect and the Basic Negative Emotion scale for controlling irritableness in this research. The reliabilities of the Negative Affect (NA) and Positive Affect (PA) PANAS-X scales in this FA consumer experiment are presented in Table 8.1.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Cronbach's Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Affect</td>
<td>0.80</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The reliability of both scales is good (Cronbach’s α ≥ 0.80). Only the NA scale is of interest to control for irritableness. The PANAS-X data of the FUI experiment (N=200) and the PANAS-X data of this FA experiment (N=149) are combined. One participant filled in PANAS-X in both experiments. The PANAS-X data of this participant are deleted from one of the databases. Based on this sample size (N=348), the average Negative Affect (NA) score is 18.69 with a standard deviation of 5.26. Consequently, the acceptable irritableness level of test participants is determined to be NA scores from 13.43 till 23.95 (18.69 +/- 5.26). Based on these NA cut-off values, the data of 36 test participants were deleted from the FA database. Fifteen participants scored below the lower threshold value and 21 participants scored above the upper threshold value. Altogether, this resulted in a sample size of N=113 distributed over the four failure scenarios as presented in Figure 8.8.

![Figure 8.8: Participants in each scenario after controlling for irritableness](image-url)
8.2.8 Pilot experiment

For evaluating the experimental protocol of Figure 8.6, eight pilot experiments were performed. Based on the observations of these pilot experiments and the comments of the test participants, some small adjustments were made to the final experimental protocol:

- **Information overload**: The TV introduction movie contained too much information about the TV and its functions. Therefore, participants could not grasp all this information and could not focus on the movie itself. It was not clear to the test participants where they had to concentrate on: the information or the movie. Consequently, parts of the provided information were removed so that the main focus of participants was directed to the introduction movie itself.

- **Programming error**: During the pilot experiments, it became apparent that the link between the failure scenario movies and the corresponding questions was incorrect. This failure was caused by a programming error and was solved before the actual experiments started.

The other experimental settings worked out well during the pilot experiments and were thus left unchanged.

8.3 Experimental results

The actual experiments were performed in week 22 till week 24 of 2007. As mentioned above, 149 university students participated in the FA experiment. Based on the average irritableness level of all test participants, the result data of 113 participants are selected and used in the analysis of this experiment (see Section 8.2).

Twenty-eight of these participants underwent scenario 1 (failure in the “Watch the desired program” function, caused by the TV), 28 participants underwent scenario 2 (failure in the “Watch the desired program” function, caused by the cable company), 26 participants underwent scenario 3 (failure in the “Motorized swivel” function, caused by the TV) and 31 participants underwent scenario 4 (failure in the “Motorized swivel” function, caused by the electricity company).

This section deals with the results of this FA experiment. The first subsection investigates the influence of FUI and FA on UPFS by statistically testing the hypotheses formulated in subsection 8.2.2. Subsequently, in the second subsection some additional outcomes of the experiment are discussed.

8.3.1 Influence of FUI and FA on UPFS

As Figure 8.6 indicates, the UPFS of the failure scenarios is measured with the UPFS survey. This UPFS survey is a self-report survey that consists of five statements to measure the level of irritation experienced by the user as caused by the TV failure scenario (see Appendix F).

This subsection investigates the relation between FUI and UPFS and between FA and UPFS by statistically testing the earlier formulated hypotheses:

---

*N* Nevertheless, another programming error was only discovered after the execution of several experiments. Because of this error, 4 participants got the wrong failure impact questions in the function failure list in the FA experiment. Therefore, the FI data of these four test participants is deleted from the dataset.
There is no significant difference in UPFS caused by failures with different levels of function importance

The UPFS that is caused by a failure in an important function is significantly higher than by a failure in an unimportant function

And:

There is no significant difference in UPFS caused by internal or external attributed failures

The UPFS that is caused by an internal attributed failure is significantly higher than by an external attributed failure

These hypotheses are tested by using the UPFS measurement results (UPFS survey). Consequently, the results of these statistical tests again raise or lower the confidence in the construct validity of (parts of) the theory-based UPFS model presented in Figure 5.2.

Before analyzing the experimental results, the reliability of the UPFS survey in this FA consumer experiment should be determined. The Cronbach's Alpha value of this survey in this experiment is 0.80, and therefore the reliability of the UPFS survey is considered to be good [HAI05]. Subsequently, a comparison is made between the average UPFS scores of the different failure scenarios. Table 8.2 presents the results of this comparison.

**Table 8.2:** Multiple comparisons of UPFS scores between scenarios (p-values)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Average UPFS</th>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.79</td>
<td></td>
<td></td>
<td>0.20</td>
<td>0.83</td>
<td><strong>0.01</strong></td>
</tr>
<tr>
<td>2</td>
<td>17.64</td>
<td></td>
<td>0.20</td>
<td></td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>18.85</td>
<td></td>
<td>0.83</td>
<td>0.69</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>16.45</td>
<td></td>
<td><strong>0.01</strong></td>
<td>0.70</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

There is only a significant difference in UPFS score between scenario 1 and 4 (p = 0.01). Therefore, based on this comparison, it can only be concluded that the failure of a high important function that is internally attributed causes a higher UPFS level than a failure in a low important function that is externally attributed. These relations are also confirmed by the graph in Figure 8.9. Nevertheless, the earlier formulated hypotheses cannot be tested with this comparison data.

These hypotheses can be tested with a (two-way) ANOVA. However, first the assumptions underlying ANOVA have to be checked. These assumptions are:

- Independence of observations
- Homoscedasticity (homogeneity of variance)
- Normality of the dependent variable

**Independence of observations**

The independence of the observations is ensured by the fact that all participants had to do the experiment under the same conditions.
Homoscedasticity
In order to investigate the homoscedasticity of the experimental data, a Levene’s test should be performed [HAI05]. The result of the Levene’s test is not significant (p=0.03) indicating that the hypothesis “the (error) variances are equal among the groups” cannot be rejected.

Normality of the dependent UPFS variable
This assumption requires that the population distributions of the dependent variable are normally distributed for all scenarios. For all four scenarios, a Shapiro-Wilk test is performed to investigate this normality assumption [HAI05]. The results of these tests are presented in Table 8.3. Based on these results, it can be concluded that the UPFS scores from scenario 1, 2 and 3 are normally distributed (p>0.05). The UPFS scores from scenario 4 are not normally distributed. However, for a moderate or large sample size (n \(\geq\) 30), ANOVA yields relatively accurate results even when the normality assumption is violated [MON99].

Table 8.3: Results Shapiro-Wilk test UPFS scores

<table>
<thead>
<tr>
<th>Scenario</th>
<th>N</th>
<th>Average UPFS</th>
<th>St. dev.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>19.79</td>
<td>3.60</td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>17.64</td>
<td>4.19</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>18.85</td>
<td>3.93</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>16.45</td>
<td>4.30</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Consequently, based on the above mentioned analysis, it is justified to use two-way ANOVA for testing the hypotheses of the FA experiment. In this two-way ANOVA, the dependent variable is the UPFS score and the independent variables are FUI and FA. The results of this two-way ANOVA are presented in Table 8.4.

Table 8.4: Results of two-way ANOVA of the UPFS variable

<table>
<thead>
<tr>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function Importance</td>
<td>0.16</td>
</tr>
<tr>
<td>Failure Attribution</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Function Importance * FA</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Despite the results of the qualitative analysis in Figure 8.10, this two-way ANOVA demonstrates that the influence of the FUI variable is not significant. Consequently, based on this analysis, the first null hypothesis of Section 8.2.2 cannot be rejected. Nevertheless, this two-way ANOVA confirms that the influence of the FA variable is significant. Therefore, based on this analysis, the second null hypothesis of Section 8.2.2 should be rejected. The interaction effect between FUI and FA is also not significant. Consequently, this insignificance of the interaction effect between FA and FUI in this experiment validates the significance of the relation between FUI and UPFS in the second FUI consumer experiment. That is, the absence of this interaction effect implies that the lack of FA control in the second consumer experiment did not bias the experimental outcomes and the relation between FUI and UPFS (see Section 7.8).

In other words, the second FUI consumer experiment did identify a significant influence of FUI on UPFS. However, in this third FA consumer experiment this significant influence is not established. Only FA influences UPFS significantly in this third consumer experiment. A possible explanation for the absence of a significant relation between FUI and UPFS in this third experiment is the difference in experimental approach (laboratory versus laptop approach). Table 8.5 presents a comparison of the Failure Impact (FI) and UPFS scores between the two experiments.

Table 8.5: FI and UPFS scores of the FUI experiment and FA experiment

<table>
<thead>
<tr>
<th>Scenario</th>
<th>FI score</th>
<th>UPFS score</th>
<th>FI score</th>
<th>UPFS score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. dev.</td>
<td>Mean</td>
<td>St. dev.</td>
</tr>
<tr>
<td>Motorized swivel</td>
<td>3.75</td>
<td>1.76</td>
<td>19.50</td>
<td>3.26</td>
</tr>
<tr>
<td>Watch the desired program</td>
<td>9.33</td>
<td>0.88</td>
<td>22.08</td>
<td>3.06</td>
</tr>
<tr>
<td>Difference</td>
<td>5.58</td>
<td>2.58</td>
<td>6.63</td>
<td>1.17</td>
</tr>
</tbody>
</table>

From Table 8.5, it becomes clear that the difference in UPFS scores between the two FUI scenarios is larger in the FUI experiment than in the FA experiment (2.58 versus 1.17). The UPFS score of the “Motorized swivel” scenario in the FUI experiment is significantly higher than the average UPFS scores of the “Motorized swivel” scenarios in the FA experiment (p < 0.01). The UPFS score of the “Watch the desired program” scenario in the FUI experiment is significantly higher than the UPFS scores of the “Watch the desired program” scenarios of the FA experiment (p < 0.01).

In this third FA consumer experiment, the differences in UPFS scores between the two levels of FUI are not large enough to detect a significant difference. The results indicate that the influence of FUI on UPFS can be measured better in a laboratory environment in which test participants actually experience product failure. The laboratory setting of the FUI experiment resembled more a real-life use situation than the setting of the FA experiment. Therefore, in real-life the emotional reactions on a product failure (UPFS) are expected to be even stronger than the UPFS levels that were measured in the FUI experiment. So although the influence of FUI on UPFS cannot be reconfirmed in the third FA consumer experiment, this failure characteristic is still considered to influence UPFS in real life use situations.

In this third FA consumer experiment, the significance level of the relation between FA and FUI is very high (p < 0.01, see Table 8.4). Based on these experimental results, the influence of FA on UPFS is established to be stronger than the influence of FUI on UPFS.
Altogether, the results of this third consumer experiment raise the confidence in the construct validity of these parts of the theory-based UPFS model presented in Figure 5.2.

8.3.2 Other results

Control for irritableness

A disproportionate distribution of irritableness levels over the experimental scenarios is expected to bias the relation between the independent variables and UPFS. Therefore, PANAS-X is used to control for the irritableness level among test participants. The selection approach of test participants in this FA experiment resulted in UPFS data collection from average irritable and extreme (high or low) irritable test candidates. This data can be used to evaluate the actual influence of irritableness on UPFS. Since fifteen participants scored below the lower threshold value for irritableness and 21 participants scored above the upper threshold value of irritableness, the UPFS scores of these two groups can be compared with the UPFS scores of the “average irritable” test participants. With this comparison of the UPFS levels of these different irritableness groups it can be checked whether the level of irritableness (measured by NA) really influences the dependent variable UPFS. Based on this comparison, the necessity of controlling for irritableness can be validated.

For comparing the average UPFS scores of these three irritableness groups (low, average and high), a non-parametric Kruskal-Wallis test is recommended [HAI05]. Table 8.6 presents the average UPFS scores of the different irritableness groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Average UPFS</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Low irritableness level</td>
<td>15</td>
<td>20.73</td>
<td>2.84</td>
</tr>
<tr>
<td>2. Average irritableness level</td>
<td>113</td>
<td>18.12</td>
<td>4.17</td>
</tr>
<tr>
<td>3. High irritableness level</td>
<td>21</td>
<td>19.33</td>
<td>4.12</td>
</tr>
</tbody>
</table>

The Kruskal-Wallis test indicates that the UPFS scores of group 1 and group 2 differ significantly (p = 0.02). However, an unexpected result is that the UPFS score of group 1 is higher than the UPFS score of group 2 and even higher than the UPFS score of group 3 (see Table 8.6). From a consumer behavior perspective, it is difficult to explain this extreme response of the low irritableness participant group. However, a possible explanation may be found in literature on cultural anthropology but is beyond the scope of this research. The Kruskal-Wallis test results also indicate that the UPFS score of group 3 is higher than the UPFS score of group 2, but not significantly higher.

These experimental results confirm the necessity to control for irritableness. Particularly, low irritableness participants are demonstrated to bias the UPFS measurement results with their extreme responses to the experimental failure scenarios.

Control for Failure Impact

Based on the results of the Function Failure list, it is possible to examine whether the participants perceived the same Failure Impact (FI) across the different failure scenarios. The experimental scenarios were designed to cause a complete (100%) disruption of the failing function.
As mentioned before, because of a programming error, four participants got the wrong Failure Impact questions in the Function Failure list in the FA experiment. Therefore, the FI data of these four test participants were deleted from the dataset. Consequently, the data of 109 test participants are used in this FI analysis.

For comparing the average FI scores of the four experimental scenarios, a non-parametric Kruskal-Wallis test is recommended [HAI05]. Table 8.7 presents the average FI scores per scenario.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>N</th>
<th>Average Failure Impact score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25*</td>
<td>3.44</td>
</tr>
<tr>
<td>2</td>
<td>27*</td>
<td>3.41</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>2.85</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>3.32</td>
</tr>
</tbody>
</table>

* because of a programming error, 4 items had to be deleted

The result of the non-parametric Kruskal Wallis test did not demonstrate a significant difference among the FI scores of the different failure scenarios (p = 0.30). Based on these results, it can be concluded that in this third FA consumer experiment the Failure Impact was well controlled.

**Variation of Failure Attribution**

A Failure Attribution (FA) questionnaire was added to check whether the test participants perceived the failure cause similar to the presented failure cause in the experimental scenario (internal or external). In the most ideal situation, the FA questionnaire scores should be 3 (internal FA score) for the internal FA scenarios (scenario 1 and 3) and 15 (external FA score) for the external FA scenarios (scenario 2 and 4). However, there should at least be a significant difference between the FA scores of the internal and external failure scenarios. Table 8.8 presents the average FA scores for the internal and external failure scenarios.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Average FA score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Failure Attribution scenarios (1 and 3)</td>
<td>54</td>
<td>6.39</td>
</tr>
<tr>
<td>External Failure Attribution scenarios (2 and 4)</td>
<td>59</td>
<td>11.08</td>
</tr>
</tbody>
</table>

Statistical testing (Welch test and Brown-Forsythe test) [HAI05] demonstrates a significant difference in FA score between the internal and external failure scenarios (p < 0.001). Based on these results, it can be concluded that variation of the independent FA variable is performed sufficiently.

**Variation of Function Importance**

A FUI survey was added to verify whether a significant difference exists between the FUI ratings of the two selected TV functions among test participants. Table 8.9 presents the FUI scores of the test participants for the two functions of interest.
Table 8.9: Function Importance scores

<table>
<thead>
<tr>
<th>Function</th>
<th>Average FUI score</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watching the desired program</td>
<td>9.50</td>
<td>0.69</td>
</tr>
<tr>
<td>Motorized swivel</td>
<td>4.51</td>
<td>2.28</td>
</tr>
</tbody>
</table>

Statistical testing demonstrates a significant difference in FUI score between the two functions (p < 0.001). Based on these results, it can be concluded that the variation of the independent FUI variable is performed sufficiently.

8.4 Conclusions

In this chapter, the influence of Function Importance (FUI) and Failure Attribution (FA) on UPFS in the theory-based UPFS prediction model was investigated. A positive relation was expected between FUI level and UPFS. Moreover, internal FA was expected to cause higher levels of UPFS than external FA. In order to validate these relations, a third consumer experiment was performed.

The results of this third consumer experiment support the expected relation between FA and UPFS. Statistical tests confirm that a failure that is internally attributed by test participants causes higher levels of user dissatisfaction than a failure that is externally attributed. The relation between FA and UPFS is highly significant (p < 0.01).

Moreover, the insignificance of the interaction effect between FA and FUI in this experiment validates the significance of the relation between FUI and UPFS in the second FUI consumer experiment.

Conversely, the results of the third consumer experiment do not support this expected relation between FUI and UPFS. Statistical tests do not confirm the significant influence of FUI on UPFS. Therefore, these experimental results do not confirm the outcomes of the second FUI consumer experiment in which this positive influence of FUI on UPFS was still established. A plausible explanation for this difference in experimental results is the deviant experimental approach of the third FA consumer experiment. Instead of simulating an actual use situation with an implemented failure scenario in a laboratory setting, this FA experiment presents test participants a movie in which the failure scenario is demonstrated.

The experimental results indicate that the average UPFS scores per FUI scenario and the difference in UPFS scores between the two FUI scenarios are larger in the FUI experiment than in the FA experiment. These results indicate that the effect size of the influence of FUI on UPFS is much higher in a laboratory environment in which test participants actually experience the product failure. Evidently, this lower effect size cannot be compensated with the higher number of test participants in the third FA experiment (see Section 8.2.3).

The laboratory experimental approach resembled more a real-life product use situation than the experimental approach of the FA experiment. Therefore, in real-life the emotional reactions on a product failure (UPFS) are expected to be even stronger (higher effect size) than the UPFS levels that were measured in the second FUI experiment. Although the influence of FUI on UPFS cannot be reconfirmed in the third FA consumer experiment, this failure characteristic is still considered to influence UPFS in real life use situations.
The results of this FA consumer experiment also confirmed the necessity to control for irritableness in UPFS research. Particularly, low irritableness participants are demonstrated to bias the UPFS measurement results with their extreme responses to the experimental failure scenarios. Therefore, for future UPFS research purposes the participant selection approach based on irritableness level should be maintained.

Also, based on the experimental results, it can be concluded that in this third FA consumer experiment the Failure Impact (FI) was well controlled. In future UPFS research, the successful control of the FI extraneous variable can be accomplished by the complete disruption of the selected functions in the failure scenarios (full failure FI-level).

In Chapter 5, the theory-based UPFS prediction model was introduced to gain insight into the impact of quality problems on user dissatisfaction in order to reduce the number of product complaints after product introduction. The last three chapters described several contributions to the validation of parts of this UPFS prediction model and the corresponding research approaches. Eventually, this resulted in the partial validation of the theory-based UPFS prediction model. The grey blocks in Figure 8.10 give an overview of the established relations between the failure characteristics (FUI and FA), the extraneous variable Irritableness and UPFS. The white blocks in this figure represent the failure characteristics that should be investigated in future research for the complete validation of this theory-based UPFS prediction model.

Moreover, based on the results of these three consumer experiments (Teletext, FUI and FA), a validated UPFS research approach has been developed.
By the application of this validated approach, future UPFS research can be performed reliably. The building blocks of this approach are:

- **The UPFS measurement approach**: The results of the FUI and FA consumer experiment indicated that the items of the self-report UPFS measurement scale correlate very highly with each other and that the reliability of this measurement scale is good (Cronbach's alphas: 0.83 and 0.80). Moreover, the results of this self-report UPFS surveys confirmed the positive relation between FUI and UPFS and between FA and UPFS. In other words, the experimental results demonstrated the ability of this survey to measure UPFS. Based on these results, the self-report UPFS survey is considered a reliable tool for the measurement of the dependent UPFS variable in future UPFS experiments.

- **The FUI measurement approach**: The proposed FUI ranking of product functions enables the variation of the independent FUI variable by selecting product functions with different FUI levels. The experimental results demonstrated the ability of this approach to measure FUI, and therefore this FUI ranking is considered a reliable tool for the measurement of the independent FUI variable in future UPFS experiments.

- **The control for Irritableness**: The application of PANAS-X for the selection of average irritable test participants generates good results. Analysis of the experimental results indicates that no significant ($\alpha = 0.1$) difference exists between the emotional states of participants in different scenarios. Moreover, the experimental results also confirm the necessity to control for irritableness in UPFS research.

- **Other extraneous variables**: Based on literature and analysis of the different experimental results, this research identified different (groups of) extraneous variables in UPFS research (use approach, use environment, other user characteristics and other failure characteristics) and approaches to control them (see Section 8.1.4). Moreover, several tools were developed to measure the successfúlness of the control or variation of respectively the extraneous and the independent variables (Function Failure List, FUI survey and FA questionnaire).

Altogether, these UPFS research results give insight into the impact of product quality problems on user dissatisfaction in consumer electronics. Furthermore, future research into this impact of quality problems is advanced by the provision of the above mentioned research tools.

The next chapter, the last chapter of this thesis, discusses the main conclusions and contributions of this research together with some recommendations for further research.
9 Conclusions and Recommendations for Further Research

Based on literature review as well as case studies in industry, the first chapter of this thesis called for a better understanding of customer satisfaction and complaining behavior to predict the impact of product design decisions on the number of customer complaints. The research reported in this thesis aims to improve the quality of the product design decisions in the NPD process of consumer electronics. Based on the research results presented in the previous chapters, this chapter is concerned with general research conclusions and recommendations for further research. The chapter is organized as follows. Section 9.1 gives an overview of the major research findings of this thesis. In Section 9.2, the major contributions of this research are discussed. Recommendations for further research are given in Section 9.3.

9.1 Research overview

In this thesis, product improvement decisions and underlying customer satisfaction models in consumer electronics have been analyzed. Several important conclusions have been drawn that are summarized below.

9.1.1 Problem Identification

The combination of several business trends causes an increase in market uncertainty concerning the nature and extent of customer's need for new products. This uncertainty makes it very difficult for designers to assess the impact of design decisions on customer dissatisfaction. This uncertainty in the decision process eventually causes “incorrect” design decisions; decisions that have a negative impact on the customer satisfaction level. In the end, these wrong design decisions result in an increasing number of customer complaints.

Literature review indicated that the combination of lacking user-orientation and confined specificity in current quality improvement methods restricts the potential contribution to market uncertainty reduction of these methods. Particularly, the unknown impact of identified quality problems on user dissatisfaction is limiting the potential contribution of these methods to the quality improvement decision process.

The results of case studies performed at three consumer electronics services centres demonstrated that the information collected in current product service processes is far from complete and specific. Consequently, the potential contribution of these methods to uncertainty reduction and customer dissatisfaction prediction in the product design process is limited. Particularly, the logistical orientation of the service process hinders the required information collection for improved design decision making.

Consequently, in order to deal with the increased market uncertainty in consumer electronics, the existing quality methods from literature and practice should be complemented with a customer dissatisfaction prediction method. This analysis resulted in the formulation of the following research question for this thesis:
How to predict the impact of (potential) product quality problems on customer dissatisfaction in consumer electronics industry?

Subsequently, this main research question was divided into three smaller sub-questions:

Sub-question 1: How to express and measure the impact of quality problems on customer dissatisfaction in consumer electronics industry?

Sub-question 2: What are the underlying factors of the quality problems that influence this customer dissatisfaction concept?

Sub-question 3: How do (some of) these factors influence this customer dissatisfaction concept in consumer electronics industry?

In the following subsections, for each of these sub-questions conclusions are discussed. Some general conclusions and implications related to the main research question are discussed in the end.

9.1.2 Sub-question 1: Measurement of the impact of quality problems

In order to reduce the number of user complaints in consumer electronics, user dissatisfaction has been identified as a good starting point for measuring the impact of (potential) quality problems. Moreover, especially the irritation emotion was identified to have a negative impact on this user satisfaction level.

All research performed in the area of consumer behavior and emotion provides a solid theoretical foundation for the application of irritation as a response measurement dimension in this research. Based on these insights, the following concept was formulated that captures the user impact of quality problems in consumer electronics products:

User Perceived Failure Severity (UPFS) is the level of irritation experienced by the user caused by a product failure

Insight into the expected/predicted UPFS resulting from a certain design decision would reduce uncertainty in the product development decision process and decrease the number of product complaints. UPFS is an emotion measurement and therefore depends on a certain stimulus (product failure) and individual- and environmental characteristics.

In this research, UPFS caused by a product failure was measured as follows:

- The UPFS survey; a self-report survey consisting of five statements to measure the level of irritation of the user when experiencing the TV failure and five additional statements to describe the expected reactions of the user to the experienced failure scenario (see Appendix F). The five irritation statements together form the self-report UPFS measurement scale.

- PANAS-X; 60-item emotion measurement schedule (see Appendix G). The difference between the user's emotional Negative Affect states just before and after product failure can be considered as a measure for UPFS.
Statistical analyses confirmed the validity of both measurement methods. However, PANAS-X was found unable to measure user dissatisfaction when the user experienced a product failure in a laboratory environment. The subtle changes in the emotional state of the participants when confronted with a product failure in a laboratory environment instead of their natural use environment (e.g., their home) can potentially explain this inability.

Nevertheless, it was found that the UPFS survey was able to measure user dissatisfaction in the same context. In particular, the reliability of this UPFS measurement scale is good (Cronbach's alpha > 0.80). In short, the UPFS survey can be considered as a valid measurement approach to measure the impact of quality problems on customer dissatisfaction levels in consumer electronics.

9.1.3 Sub question 2: Underlying factors that influence customer dissatisfaction

The first step of identifying the underlying factors of the quality problems that influence the UPFS was to evaluate the current industry approach to user dissatisfaction prediction. This approach was translated into the first version of the UPFS prediction model: the expert-based UPFS model. This model states that product developers/testers evaluation of the user dissatisfaction of a certain failure agrees with the actual user dissatisfaction of that failure expressed by the UPFS. This expert-based UPFS prediction model was evaluated in a first consumer experiment.

The results of this consumer experiment did not agree with the expert-based UPFS model. Furthermore, several other recent sources in literature dispute the ability of product experts to predict user preferences and behavior with respect to consumer electronics. The combination of these results confirms that product developers are unable to predict the level of irritation experienced by a user when confronted with a product failure. Consequently, there is not sufficient theoretical foundation to apply this UPFS prediction model in practice (e.g. in consumer electronics industry).

An adjusted UPFS prediction model was therefore formulated based on the results of the first UPFS consumer experiment, literature from different scientific fields and an expert validation session with an academic panel. The starting point of this so-called theory-based UPFS prediction model is the emotional response model from consumer behavior theory.

The theory-based UPFS prediction model consists of two groups of variables that may influence UPFS:

- **Failure characteristics**: failure attribution, failure frequency, failure impact, failure moment in use process, failure reproducibility, failure solvability, failure work around and function importance
- **Extraneous variables**: use conditions and user characteristics

Furthermore, behavioral research indicates the need to control the environment and the individual characteristics of test participants, when investigating the influence of the stimulus (quality problem) on the emotional response (irritation) in this research. The required control of the environmental and the individual characteristics in this research was performed by the definition of the following groups of extraneous variables:

- **The use environment**: the physical conditions in which the product is used
- **The use approach**: the way the product is utilized by the users

149
- Irritableness: quick-temperedness of users
- Other user characteristics: individual traits of the users

In this research, methods have been developed to control the different groups of extraneous variables. In addition, the results of this research confirmed the necessity to control for user irritableness in UPFS research. The PANAS-X measurement scale can be successfully used for the selection of average irritable participants. Consequently, the combination of the UPFS measurement approach and the effective control of the extraneous variables is considered a valid tool for the impact measurement of product quality problems on customer dissatisfaction.

In summary, the expected underlying factors that influence UPFS are failure characteristics together with user characteristics and use conditions. This research mainly focused on investigating the influence of failure characteristics on UPFS.

9.1.4 Sub question 3: Influence of factors on customer dissatisfaction

Given the time and resource constraints of this research project, the complete validation of the theory-based UPFS model within this project was unfeasible. Therefore, this research started with investigating the influence of one of these failure characteristics, Function Importance (FUI), on UPFS. In order to validate the relation between FUI and UPFS, the second consumer experiment was performed.

The results of this second consumer experiment confirmed the positive relation between FUI and UPFS. Although the statistical tests showed that a failure in an important function causes higher levels of user dissatisfaction than a failure in an unimportant function, the unsuccessful control of the extraneous Failure Attribution (FA) variable made it impossible to accept this experimental result at this point in time.

Consequently, in a third consumer experiment the combined influence of FUI and FA on UPFS was evaluated. The results of this third consumer experiment confirmed the expected relation between FA and UPFS. Statistical tests showed that a failure that is internally attributed by test participants causes higher level of user dissatisfaction than a failure that is externally attributed. The relation between FA and UPFS is very significant (p < 0.01).

Moreover, the insignificance of the interaction effect between FA and FUI in this third experiment validated the significance of the relation between FUI and UPFS in the second consumer experiment. Conversely, statistical tests did not confirm the significant influence of FUI on UPFS. A plausible explanation for this difference in experimental results is the different experimental approach (failure observation instead of actual failure experience) of the third consumer experiment.

Altogether, the second and third consumer experiment contributed to the validation of the theory-based UPFS prediction model. The influence of two failure characteristics (factors) on UPFS (user dissatisfaction) has been validated:

- Function Importance: A failure in an important function causes a significantly higher level of UPFS than a failure in an unimportant failure
- Failure Attribution: An internal attributed failure causes a significantly higher level of UPFS than an external attributed failure
In addition, the influence of FA on UPFS was demonstrated to be stronger than the influence of FU1 on UPFS.

9.1.5 General conclusions and implications

Den Ouden states that in the consumer electronics industry, more than 50 percent of the returned products are in reality in full working order [OUD06]. This situation is, however, not unique for consumer electronics products. Davidson indicates that No Fault Found figures have increased to extraordinary high levels (50-60 percent) in commercial airline industry, military repair depots and microsystems [DAV05A].

This research demonstrates that increasing market uncertainty makes it very difficult for designers to assess the impact of design decisions on customer dissatisfaction. This is the case for early decisions in the concept selection process as well as for later quality improvement decisions in the product validation phase. This uncertainty in the decision process eventually causes “incorrect” design decisions; decisions that have a negative impact on the customer satisfaction level and increasing NFF levels.

In addition, this research also demonstrated that, especially under increasing market uncertainty, product designers often fail to predict the impact of their design decisions on user dissatisfaction. Moreover, they often fall short in accounting for users' views on the product and foreseeing their preferences [KAR07]. Market uncertainty can therefore be potentially reduced when applying the theory-based UPFS prediction model that incorporates the user perspective into the design decision process. In other words, the prediction of the impact of (potential) product quality problems on user dissatisfaction would eventually result in less “incorrect” design decisions and lower NFF levels.

9.2 Research contributions

Based on the conclusions presented in the previous section, the main contributions of this research are summarized below.

Lacking user orientation and specificity of quality improvement methods in literature

One important observation in the literature is that the combination of lacking user-orientation and confined specificity restricts the potential uncertainty reduction of current quality improvement methods. There is lack of impact prediction of quality problems and the related improvement decisions on user dissatisfaction in almost all methods. Moreover, the methods that do predict the impact of the identified quality problems do not apply a user-oriented approach. In these methods, the impact prediction is based on the general appraisal of the involved (product) engineer(s).

Limited contribution of service feedback information to uncertainty reduction

The complete service process is designed for managing technical quality problems. The only non-technical quality problems that are dealt with are issues with so-called ignorant customers. In addition, the information collected in current product service processes is far from specific to potentially contribute to uncertainty reduction in the product design process. The logistical orientation of the service process hinders the required information collection for improved design decision making. As a result, the service feedback process in its current form does not generate suitable information for uncertainty reduction in the design process of consumer electronics.
The inability of experts to predict users' preferences and behavior

The expert-based UPFS model stated that product developers/testers evaluation of user dissatisfaction with a certain failure agrees with actual user dissatisfaction with that failure, expressed by the UPFS. However, the results of a first consumer experiment do not support the validity of the expert-based UPFS model. In addition to this experiment, several other recent sources in literature disputed the ability of experts to predict user preferences and behavior in consumer electronics. The combination of results demonstrated that product developers are unable to predict the level of dissatisfaction experienced by a user caused by a product failure. Consequently, the theoretical foundation for the application of this UPFS prediction model in practice (e.g. consumer electronics industry) is lacking. The limited validity of this practical approach might well contribute to the increasing number of customer complaints about consumer electronics products.

The theory-based UPFS prediction model

The theory-based UPFS prediction model is introduced to gain insight into the impact of quality problems on user dissatisfaction in order to reduce the number of product complaints after product introduction. Literature review related to consumer behavior and emotion provided the basis for the UPFS model. In this model, user dissatisfaction (UPFS) is modeled as a function of not only the product related characteristics (failure characteristics), but also of the user related characteristics (use conditions and user characteristics). This model was partially validated by conducting two consumer experiments. The application of these parts of the model in the NPD process of consumer electronics can potentially reduce market uncertainty in the design decision process.

The UPFS experimental approach

Based on the results of the three consumer experiments (Teletext, FUI and FA), a UPFS experimental approach has been developed. By the application of this approach, future UPFS research can be performed reliably. The building blocks of this approach are:

- **The UPFS measurement approach**: Experimental results confirmed the reliability and validity of the self-report UPFS survey as a tool to measure the dependent UPFS variable in future UPFS experiments.

- **The FUI measurement approach**: Experimental results demonstrated the ability of the FUI ranking approach to measure FUI and therefore this approach is considered a reliable tool to measure the independent FUI variable in future UPFS experiments.

- **The control for Irritableness**: The application of PANAS-X for the selection of average irritable test participants generated good results. Analysis of the experimental results indicated that no significant ($\alpha = 0.1$) difference exists between the emotional state of participants in different scenarios. Moreover, the experimental results also confirm the necessity to control irritableness in UPFS research. Therefore, for future UPFS research purposes, the PANAS-X approach for the selection of participants on irritableness level should be maintained.

- **Other extraneous variables**: Based on literature and analysis of the different experimental results, this research identified different (groups of) extraneous variables in UPFS research and approaches to control them.

Altogether, future research on analyzing the impact of quality problems on user dissatisfaction is advanced by the provision of the above mentioned research tools.
Valorization
Both the UPFS prediction model and the UPFS experimental approach can add value to the product development processes of companies within consumer electronics industry. Although valorization of the research results is not the primary goal of this research, the business context provided by the Trader project required constant consideration for the business perspective in this research. In order to confirm the business contributions of this academic work, a workbook will be composed in which the practical application of the UPFS model and the accompanying experimental approach are explicated.

9.3 Generalization and further research

The previous section presented the main contributions of this research. In addition, this section deals with the generalization of these research results and some suggestions for further research.

9.3.1 Generalization
To control the influence of user characteristics on the experimental results, it was decided to select a homogeneous group of Dutch university students to participate in the last two consumer experiments. In general, students are considered to be a sound, homogeneous respondent group in consumer (survey) research [PET01]. Stangor states that although the use of college students poses some limitations, it must be realized that any sample of research participants is limited in some sense [STA98]. Moreover, there are no concrete indications to assume that university students respond differently to product failure behavior than other user groups would do. However, generalization of the research results to other user groups should be done with care.

All consumer experiments were carried out with a single type of consumer electronics product: a TV set. Again, it would be statistically invalid to simply generalize the research results to consumer electronics products in general. Nevertheless, it is possible to do some generalization based on the structural similarities between TVs and other consumer electronics products. The product TV, its technology, the structure of its NPD process and its intended users are not very different from other consumer electronics products. Moreover, all these consumer electronics products operate in a market that is influenced by similar business trends; time-to-market, globalization, increasing complexity and decreasing product understanding [BRO04]. Based on these structural similarities, it is quite likely that the results of this research can be extrapolated to other consumer electronics products than just TV.

9.3.2 Recommendations for further research
The previous sections summarized the main results of this research. Notwithstanding the important contributions of this thesis, the research results also trigger the formulation of new research directions. Therefore, in the following, recommendations for future research are discussed.

Further validation of the UPFS model
The validation of the complete UPFS model requires a gradual approach in which the (combined) added values of all failure characteristics to the model are evaluated. This research has investigated the influence of two of these failure characteristics (function importance and failure attribution) on UPFS.
However, in future research the influence of the other failure characteristics on UPFS should be investigated together with the interaction effects between these failure characteristics.

Moreover, the influence of user characteristics on UPFS was left out of consideration in this research. User characteristics were controlled by the selection of a homogeneous group of participants (university students) with an average level of irritableness. However, the emotional response model from consumer behavior theory indicates the expected influence of individual characteristics on the emotional response (UPFS) that is caused by a stimulus (product failure). In future research, the influence of these user characteristics on UPFS should be analyzed. This requires the following research steps:

- Identification of relevant user characteristics to predict the UPFS
- Formulation of the influence (direction) of these user characteristics on UPFS
- Validation of these hypothetical relations between user characteristics and UPFS

Currently, another PhD student within the Trader project, Jeroen Keijzers, is working on the investigation of the influence of relevant user characteristics on UPFS [KEI07].

Lastly, the interaction effects between the determinant failure- and user characteristics should be evaluated. This validated UPFS model can then be used to predict the UPFS for different user groups that is caused by specific product failures.

**Different products, users and cultures**
As mentioned before, a single group of participants (Dutch university students) and a single product (TV) were applied in the UPFS consumer experiments. Although it is possible to do some generalization based on the structural similarities, future research should validate the relations in the UPFS prediction model with different user groups and different products. Moreover, consumer behavior is becoming more heterogeneous because of cultural differences [HOF02], [MOO04]. The influence of cultural differences on UPFS should therefore also be analyzed. This would further increase the generalization of the above mentioned research results.

**Formalization of the UPFS experimental protocol**
Based on the results of these three consumer experiments, a UPFS research approach has been developed. This research approach consists of validated tools to measure UPFS, function importance, irritableness and other extraneous variables. Currently, the formalization of this research approach is improved by generating a ready-for-use experimental protocol. Such a protocol should consist of a practical experimental handbook complemented with the required variable measurement tools and (straightforward) variable analysis tools. Eventually, the measurement and analysis tools should be ready-for-use digital forms that do not require in-depth knowledge on experimental design or statistical testing.

This ongoing further formalization of the UPFS approach into a full experimental protocol makes in easier to implement in actual product development processes. Moreover, this ready-to-use protocol should lower the threshold for product designers to evaluate the influence of certain design decisions on user dissatisfaction. Eventually, the application of this protocol in different consumer electronics companies should result in a large knowledge base on the impact of different failure- and user characteristics on UPFS.
Implementation of the UPFS model in the NPD process
The above mentioned results illustrate that market uncertainty can be potentially reduced when applying the theory-based UPFS prediction model in the NPD process of consumer electronics products. However, the actual implementation of the theory-based UPFS model and the accompanying experimental approach into the NPD process of consumer electronics requires some additional research.

The first version of this theory-based UPFS prediction model is mainly theory-based. The previous subsection suggested the improvement of the practical applicability of the model and the experimental approach: the UPFS experimental protocol. The actual implementation of the UPFS model in the NPD process requires several additional research steps. Therefore, in future UPFS research, the following questions should be answered:

- In what phase of the NPD process of consumer electronics products is the application of the UPFS model most useful?
- Which people involved in the NPD process of consumer electronics products should adopt the UPFS model in their daily practice?
- Who should be held responsible for the incorporation of the user perspective into the development process?
- How can the UPFS knowledge base be made accessible for all these involved people?

These theoretical and organizational issues should be dealt with in future UPFS research.

Implementation of the UPFS model in the test strategy in the NPD process
Besides the general adoption of the UPFS model by product designers in the NPD process, the UPFS model can be potentially applied for early product evaluation in consumer tests. In these consumer tests, the UPFS can be used to predict the impact of the different design decisions and quality problems on user dissatisfaction. Eventually, this should result in the development of a user-oriented impact assessment tool for product design decisions. The actual implementation of this model into the consumer test protocols requires the further validation of the variables in the UPFS model and the earlier mentioned formalization of the UPFS experimental protocol.

Implementation of the UPFS model in the service feedback process
The last suggestion for future research concerns the adoption of the UPFS prediction model for the improvement of the quality of the feedback information in the service process.

In order to reduce the number of future problems with consumer electronics products, Den Ouden suggests improving the decision processes in the product creation process by supporting it with up-to-date and rich information about customer use preferences [OUD06]. One approach to acquire this information is to collect high quality feedback information in the aftermarket/service processes. Subsequently, this information can be used for deciding on the different product improvement options based on the impact on consumer satisfaction and improvement implementation effort.

Future research should investigate how the theory-based UPFS prediction model can be applied in the aftermarket/service processes for this prioritization of product improvement options.
In other words, future research should evaluate how this application of the UPFS prediction model in the aftermarket can increase the quality of service feedback information by providing the user impact of the different identified product failures to the NPD process of the involved companies.

Application of the UPFS prediction model and experimental approach in the product aftermarket/service processes could potentially reduce the market uncertainty in consumer electronics. Furthermore, this UPFS model application would imply a higher level of user orientation of the involved service centres, which is one of the major barriers in the current application of service information. Moreover, ongoing research efforts into the enrichment of this market feedback information could be supported by the application of the UPFS model variables [LU07]. In this research, the presence of information about the different failure- and user characteristics in the feedback information can be considered criteria for the evaluation of these information sources and databases.
References


[BER04]  Bergamaschi, R., “System-level design: only the radical will survive”, *IBM T.J. Watson Research Center*, 2004


[DAV05A] Davidson, S., “Towards an understanding of no trouble found devices”, 23rd *IEEE VLSI Test Symposium*, 147-152, 2005


[KET02] Ketola, P., Integrating Usability with Concurrent Engineering in Mobile Phone Development, PhD-thesis, University of Tampere, Department of Computer and Information Sciences, Tampere, 2002


164


[PDA04] The PDMA glossary for new product development, Product Development and Management Association, 2004


Appendix A: Paper - Understanding Failure Severity

This paper was published in 2006. It is based on three case studies performed at the European service centres of a large consumer electronics company (company X). This paper illustrates that the recent trends in the field of consumer electronics have resulted in an information gap between developers and users of consumer electronics products. Reproduced from “Understanding failure severity in new product development processes of consumer electronics products”, by I.M. de Visser, Y. Lu and G. Nagappan in the proceedings of the 2006 IEEE International Conference on Management of Innovation Technology.

A.1 Abstract

Reliability feedback information from the field is essential for product improvement. Traditionally, consumer electronics service centres have provided the product designers with reliability information about product generations on the market at that time. However, within the field of high volume consumer electronics, movement of the business processes to emerging markets, increasing time-to-market pressure, increasing product complexity and increasing customer requirements have led to an increasing number of non-technical failures. Moreover, these business trends have resulted in higher requirements on the information that is needed in the development process. Particularly, information about the root-causes of non-technical failures is required to determine the impact of these failures on the overall product reliability. This study investigates the suitability of current service processes for the collection of information about non-technical failures and their root-causes in order to prioritize them within the development process of consumer electronics. Three case studies, performed at service centres for consumer electronics products, indicate that the recent trends in the field of consumer electronics have resulted in an information gap between developers and users of these products. In order to close this gap, the authors suggest the development of a new failure prioritization model for consumer electronics products.

A.2 Introduction

Reliability feedback information from the field is essential tool product improvement [PET05]. Traditionally, consumer electronics service centres provided the product designers with reliability information about product generations on the market at that time. This information mainly consisted of number of repaired parts within the different consumer electronics products [PET03]. This information was used by the product designers to identify and prioritize major reliability issues in current product generations or to prevent this problem from appearing in future product generations. This failure prioritization aims at assuring that these product failures are dealt with in accordance to their negative impact on overall product reliability.

However, within the field of consumer electronics, manufacturers are currently working under strong pressure, because they have to deal with four different and often conflicting, business trends, namely globalization and segmentation of the business processes, increasing time-to-market pressure, increasing product complexity and increasing customer requirements [PET03]. These business trends directly have led to a situation where products are rejected even when they still perform according to the technical specification. To reach
customer satisfaction, it is necessary to extend the definition of reliability to include these non-technical failures as well. Moreover, these business trends also impose new requirements for reliability feedback processes in service centres: collect reliability feedback information that can be used to prioritize these failures in the development process. The main purpose of this study is to investigate the suitability of current service processes for the collection of sufficient and relevant information to perform failure prioritization within the development process of consumer electronics.

The first step in this investigation is presented in section A.3. In this section, a more extensive description of the business trends within consumer electronics industry is given together with the implications of these trends for the requirements of reliability feedback processes in service centres. Subsequently, in section A.4, the results of three case studies are presented. These case studies were performed at three service centres for consumer electronics products. Based on the results of these three case studies, some conclusions can be drawn about the suitability of current service processes for the collection of information to perform failure severity prioritization within the development process of consumer electronics. These conclusions are summarized in section A.5.

A.3 Business Trends and Implications

The first market trend is the globalization and outsourcing of the business processes. In their paper “foundation and funding opportunities for globalization”, Gerstenfeld and Njoroge define globalization as “the construction of a global economy largely through the activities of private firms that move their economic activities around the world” [GER04]. As alternative approach for managing product creation activities globally, companies can focus on their core-business and outsource the other activities globally. Another advantage for the outsourcer is the reduction in overhead costs. Nevertheless, there are some serious risks attached to the outsourcing process that should be considered. Outsourcing can cause unwanted dependencies of suppliers [PET03]. Moreover, effectively managing outsourced relationships becomes a core competence itself. Insufficient communication and unclear interfaces between the products of outsourcing and supplying companies may result in serious product failures. Some of these failures may only be discovered after market introduction, generally resulting in serious financial and technical consequences.

The second trend, time-to-market, has become a central point in many industries recently [LU02]. Development time determines how responsive a company can react to competitive forces and to technological developments, as well as how quickly a company will receive returns form their development effort [ULR00]. In order to keep up with competitors, it has become essential for companies to introduce more products to the market faster.

The third market trend is the increasing complexity of products and the accompanying product creation processes. Probably the best illustration of the ongoing increase in product complexity is the growing software content of consumer electronics products. Around 1995, an average consumer electronics product contained around 100 000 lines of code. In 2000 this number had already increased to an average of 1 000 000 lines of code. This implies a tenfold increase in software content within 5 years. It is expected that this rapid growth in software complexity for consumer electronics will continue [BER04]. The increase in software content of consumer electronics products has resulted in a corresponding increase in state space of these products. The interaction between the different software and hardware components makes it almost impossible to determine all possible failure mechanisms of these products.
Not only is the complexity of the technical content of products increasing, but the diversity and the variety of products as well. Together with that, the “openness” of consumer electronics products is continuously increasing. This means that these products become involved in networked environments that affect them in ways that were not foreseen during their creation. It is not difficult to see that in the situation of open systems, product interactions are more difficult to predict. Eventually this may also result in unexpected field failures and higher levels of warranty costs.

Companie s have to deal with this increasing complexity in their product creation processes by delivering products that satisfy customer requirements [LU02]. However, at the same time, meeting customer requirements has become more difficult as a result of the increasing customer expectations. Besides that, customers often do not realize the complexity behind the systems they use, and therefore they do not see the difficulties that come with complex systems and just expect them to work [BRO01]. The trend of increasing customer requirements is also expressed in the warranty period and warranty coverage. Nowadays, a warranty period of one year is normal and two years quite common. In the past warranties only covered the products that did not comply with the product specifications, like replacement of defective components. These days, most manufacturers tend to follow a “no questions asked policy”. This means that a product is considered to have failed if customers are simply dissatisfied with its performance.

**Implications of the Business Trends**

The above described business trends directly affect the quality and reliability research area and also the requirements for reliability feedback information. One important development in this research area, resulting from these business trends, is the extension of the reliability definition. In 1996 Lewis defined reliability as “the probability that a system will perform its intended function for specified period of time under a given set of conditions” [LEW96]. In this definition a product is said to fail when it does not perform its intended function. However, these business trends have led to the situation in the field where products can be rejected despite the fact that they do comply with their technical specifications [OUD05]. These products are rejected because they do not perform according to the expectations of the user. This is especially true in high volume consumer electronics industry. Therefore, in order to deal with reliability in consumer electronics the classical definition needs to be extended. In this paper, the extended reliability definition is adopted in which a product is said to fail when it does not comply with the expectations of the user. The extended reliability definition has led to the denomination of another category of failures namely, non-technical failures [GEU05].

Complains about the (lack of) functionality are caused by a mismatch between customer expectations of the product and the actual experience of the customer with the product. As a result of an increasing reliability of components together with an increasing product complexity and customer requirements, component related reliability problems have become a minority of current field complaints [PET00]. On the other hand, the increase in customer requirements has resulted in more non-technical reliability problems entering the product service process. It is very questionable whether current service processes are designed in such a way that these reliability problems can be identified and prioritized. Literature review indicates that earlier described business trends have resulted in higher information needs [PET05]. Information about the root-causes of failures is necessary to distinguish between technical failures and non-technical failures and to determine possible solutions for the failure
in current or future product generations. This root-cause information is also required for the determination of the impact of the failure on the overall product reliability.

Therefore, the service processes that are suitable for failure severity prioritization in consumer electronics industry should (at least) be able to:

- Identify and prioritize non-technical reliability problems in the product.
- Perform root-cause analyses for the determination of the impact of the failures on the overall product reliability.

### A.4 Case Studies

As mentioned earlier, the main goals of this section is the investigation of the suitability of current service processes for the collection of sufficient and relevant information to perform failure prioritization within the development process of consumer electronics. Based on the identified business trends, this investigation will concentrate on service centres ability to perform root-cause analyses and to identify and prioritize non-technical reliability problems.

#### Introduction to case studies

The case studies described in this section are all performed in the field of high volume consumer electronics. In order to get a complete overview of current service processes, three third party European service centres of a large consumer electronics company (company X*) were visited. These service centres are located in the Netherlands, Germany and France and are responsible for most of company X's repairs in these countries. However, these three service centres work independent of each other. The service process is one of the many business/consumer processes that connects the consumer with the product. After market introduction, consumers buy company X's products and start to use them. When the products do not meet their requirements, products are send to the service centres via a number of different channels. After repair, products are returned to the customers and reliability information with various levels of detail on the repair-action performed is send back to company X.

#### Case study approach

The case studies consisted of the following activities:

- **The case study preparation**
  - Selection of a carrier product. This carrier product was followed during the service process in the different service centres. The selection of one carrier product makes it easier to trace the different service activities and to compare service centres.
  - Definition of key persons, operational and managerial, who represent the different activities within the service process.
  - Acquisition of full commitment of the involved service centres. This commitment was acquired with the help of company X.

- **Interviews and cross check**
  - Operational process mapping by interviewing people involved in the service process.

*The name of the company, presented in these case studies, cannot be disclosed due to reasons of confidentiality. This information, however, is available to the authors.*
- Operational process verifying by cross checking with on site observations in the service process.

**Case study results: the service process**

Although the three service centres worked independent of each other, the service processes of these service centres appeared to be quite similar. An overview of the general service process is given in figure A.1. The service process consists of five phases, namely:

- Reception, unpacking and identification
- Diagnosis, repair and correction
- Quality check
- Packaging
- Dispatching

The German service centre in some cases applied, in addition, the so-called “swap procedure”. Subsequently, an explanation of the different process phases will be given.

![Diagram of the service process]

**Figure A.1: Overview of the service process**

**Reception, unpacking and identification**

As part of the service agreement between company X and the service centres, the three service centres collect field returns of company X's products from various dealers and distributors. The service centres can also receive direct returns from the end users. After reception, all returned products are unpacked and visually inspected. The identification is based on product type and serial number. At the request of company X, for all products the basic customer information (e.g. name, address, date of purchase) is logged in a computer program. Subsequently, the different product types are sorted out and forwarded to the involved repair station.

**Diagnosis, repair and correction**

In Germany and the Netherlands, the people repair only a restricted number of product types. As a result of that, the people can be considered true experts regarding these product types. For several product failures, the repair engineers have executed complete root-cause analyses and designed repair processes or work-around solutions. In the case of a product entering the service process for the second time within three months without a clear problem, the repair engineer will contact the customer to ask for extra information about the failure behavior of the device. Some failures are then solved by giving some extra use-information to the customer. Sometimes they write down instructions themselves and in other cases they copy
pieces of the user manual. For other failures, the repair engineer will start a root-cause analysis based on the customer information of the failure, in order to solve it.

In France, two levels of repair processes are defined by the service centre. The first level of repair is concerned with common failures in the products. These repairs can be executed by non technical repair personnel according to a standard repair protocol. The second level of repair is similar to the standard diagnosis, repair and correction process of the Dutch and German service centres. In this second level of repairs, all less common failures are dealt with by technically educated repair personnel. Again, these repair people can be considered as true experts with regard to the involved product types. However, this French service centre does not contact the customers in case of repeating repairs.

**Quality check, packaging and dispatching**

In the Quality Check phase, a number of tests are performed. The test process is defined by the service centres with the requirements from company X. About 50 percent of the products that have passed the repair process will be tested at the Quality Check. Products that do not pass the tests at the Quality Check are sent back to the repair process. Products that initially did not pass the quality check are always tested a second time after the second repair process. Subsequently, all repaired products are packed properly. All necessary documents will be attached and than the product will be sent to the customers directly or via distributors and dealers.

**Swap procedure**

In Germany, some product types are immediately replaced by new products. For these products no repair takes place. The customers of these products always receive an alternative product instead of their own product. This process is called: the swap procedure. This swap process is also used for some immature products for which no root-cause analysis has been performed to identify the main causes of failure. Important information about failure mechanisms in these products is discarded together with the products. In future product generations, these failures will not be prevented due to this loss of information about the root-causes of these failures.

**Root-cause analysis**

As was indicated in the previous section, service processes suitable for failure severity prioritization in consumer electronics industry should (at least) be able to perform root-cause analyses on failed products. In these three service processes failure analysis is performed as follows:

For every new product introduction, company X defines that a number of first returned sets are analyzed in detail and repaired in specialized service centres. All three repair centres are indicated as specialized repair centres for new product introduction. In these specialized service centres the technical department at the service centres will repair the products with the support from company X. Subsequently, detailed root-cause information is send back to company X. This information collection process is called Introduction Repair Process (IRP).

For some of these newly introduced products, company X requires that a certain number of these products are to be exchanged and directly sent back to the factory for further analysis. This process is called the Introduction Factory Process (IFP). Company X decides for every product type whether the sets are to be changed or repaired. The IRP and IFP approach are part of a new procedure within Company X to improve production quality. For other repairs (not new product introduction), the repair centres send repair job sheets back to company X via the central service organization. From these job sheets, company X can learn what failure
symptoms the returned products have and what repair actions were taken. Based on this information, company X calculates the amount of service costs that should be paid to the service centre. The content of this information source is logistically oriented. As a result of this logistical orientation of this information, this repair information is hardly suitable for reliability analysis and reporting.

Moreover, company X requires from the service centres that they use so-called IRIS code to describe the failures. IRIS code is a standard failure description code for consumer electronics. However, the IRIS failure descriptions are very general. All service centres indicated that this IRIS code does not specify the failures symptoms specific enough. As a result, the service centres internally use their own failure description to classify the failure symptoms more specifically. But this information is not sent back to company X because it is not requested for.

Another important observation was that the service centres often classify the failure symptoms on the basis of the repair actions that were taken. This is quite understandable considering the logistical orientation of this feedback information. However, the repair actions aim to remove the failure symptoms instead of solving the failures from their root-cause level. That means that the failure classification does not provide any information about the root-causes of the different failures.

In general can be concluded that feedback information based on IRP data is very detailed and does contain root-cause information. However, the number of sets used in IRP is very limited. It is very uncertain whether all failure root-causes are identified in the IRP process. Feedback information based on the normal repairs is of limited quality since it describes failure symptoms using IRIS code. The service centres collect more root-cause information than company X requires. Within the current logistical organization of the service process failure severity prioritization using root-cause information is impossible.

**Non-technical failures**

The other requirement for the service process, in order to be suitable for failure severity prioritization in consumer electronics industry, is that the process should (at least) be able to identify and prioritize technical and non-technical reliability problems in the product. In these three service processes all non-technical failures follow the so-called No Fault Found (NFF) process. The NFF process is a special case of the repair process. This describes the situation in which no clear failure can be identified during the repair process in the service centre.

When a product enters the service centre for the first time and the repair engineers can not find a problem, the device software will be updated. By updating the software, the service centre hopes to solve the intermittent failure or the interaction failure that was possibly present in the product. However, the non-technical reliability problems can not be solved by this software update. After this software update, the product is packaged and dispatched to the customer. All these product failures are reported as being software failures to company X.

The service centres receive a fixed amount of money per repair for every product type. When a product enters the service process for the second time within three months, company X will only pay material costs for this second repair. The procedures for the second NFF-service for the same product within the first three months differs between the three service centres. In France the procedure is exactly the same as for the first NFF service. After a new software update the product is dispatched to the customer. Only at the third request for service for the
same product, this service centre will ask company X to offer this customer a so-called “commercial solution”. This means that company X will offer this customer another product or (part of) their money back.

In the Netherlands and Germany, the procedure for the second NFF service is different. Both service centres will do another software update but also provide extra use-information to the customer. As mentioned before, they will type their own instructions or copy parts of the user manual. By providing this information to the customers, the service centres try to solve non-technical failures that are caused by incomplete customer understanding of the product. However, another group of non-technical failures, caused by customer's dissatisfaction with the functionality of the product, is not given any attention not to mention solved by the service processes.

The complete service process is designed for managing technical reliability problems. The only non-technical reliability problems that are (sometimes) dealt with are the issues of so-called ignorant customers. These actions are initiated by the service centres and not by company X. Company X does not pay these service centres for these activities. As a result, the customers depend on the goodwill of the service centre whether they will provide this extra information. No formal procedures are formulated for the management of customer requirement problems. No information about non-technical reliability problems is provided to company X. Consequently, company X does not use any non-technical reliability information in their failure prioritization process.

A.5 Conclusions and Recommendations

The main goal of this research was to investigate the suitability of current service processes for the collection of sufficient and relevant information to perform failure severity prioritization. The results of the case studies indicate that the recent trends in the field of consumer electronics have resulted in an information gap between the developers and users of consumer electronics products. The current logistical structure of the service organization is not suitable for the collection of information about root-causes of failures and about non-technical reliability problems. As a result of that, product developers are thrown back on their own intuition with respect to failure prioritization. Moreover, the failure severities of technical reliability problems are being over-emphasized as a result of neglecting non-technical reliability problems.

In order to close this information gap, the authors propose the development of a new failure prioritization model for consumer electronics products. This model should support product developers to collect and organize the appropriate failure and customer information in order to make founded decisions on product failure severities. Furthermore, this model should recognize the new information needs of the development process as a result of earlier described business trends. This model will contribute to a reliability improvement process in which product failures are dealt with in accordance to their negative impact on overall product reliability. Eventually, this should add to the improvement of current and future generations of consumer electronics products.
Appendix B: Task list Teletext Experiment

This Appendix contains the task list of the teletext experiment (in Dutch).

B.1 Task list

Takenlijst Trader consumententest teletekst gebruiksvriendelijkheid
Hieronder staan drie scenario's beschreven waarin informatie van teletekst nodig is. Wij willen je vragen om in de onderstaande volgorde de benodigde informatie op te zoeken in teletekst op Nederland 1, 2 of 3. Probeer hierbij net te doen alsof je je TV thuis gebruikt.

Wij willen je vragen om alles wat je doet en denkt hardop te zeggen zodat wij kunnen volgen wat je doet en denkt. Mocht je er in het uiterste geval niet uit komen kun je de observatoren vragen om hulp.

Om de informatie met betrekking tot het experiment op te kunnen slaan willen wij je erop wijzen dat je je TV net als thuis kunt gebruiken behalve dat je de TV niet uit mag zetten en niet aan de kabels van de TV mag komen!

Probeer om de volgende informatie zo goed en volledig mogelijk op te zoeken in teletekst:

1. PSV scenario
Je bent een grote fan van het voetbalteam van PSV Eindhoven. Je bent net terug van vakantie en je wilt graag weten:
   • Wanneer PSV de volgende wedstrijd moet spelen
   • Op welke positie PSV op de ranglijst staat

2. TV-gids scenario
Je hebt thuis geen TV-gids en je wilt graag weten of er vanavond leuke films zijn te zien op de televisiekanalen RTL5 en SBS6. Zoek op welke films er vanavond te zien zijn op RTL5 en SBS6.

3. Schiphol scenario

Bedankt voor je tijd!
Appendix C: Survey Teletext Experiment

This Appendix contains the survey of the teletext experiment that measures UPFS and participant reaction (in Dutch).

C.1 Survey

Naam: ...............................................................
Datum: ..............................................................
Scenario: ...........................................................

De volgende 10 stellingen hebben betrekking op de teletext opdracht die je zonet hebt uitgevoerd. Geef aan in welke mate je het eens bent met onderstaande stellingen.

<table>
<thead>
<tr>
<th>Stelling</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. De teletext functie van deze televisie werkt foutloos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Het probleem in de teletext functie vind ik relatief onbelangrijk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Als dit probleem thuis op zou treden zou ik mijn omgeving daarover vertellen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Thuis, zou ik voor dit probleem de helpdesk van de fabrikant raadplegen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Dit probleem in de teletext functie zou ik omschrijven als &quot;erg irritant&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Dit soort problemen in de teletext functie van mijn televisie zou ik accepteren</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. De teletext functie van deze televisie vertoont problemen in het gebruik.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Thuis zou ik deze fout beschouwen als een kleine tekortkoming in de teletext functie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Als dit probleem thuis op zou treden zou ik een familieled/kennis om hulp vragen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Thuis, zou ik de televisie teruggebracht hebben naar de dealer voor reparatie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Nogmaals bedankt voor je medewerking!
Appendix D: Function Importance Survey

This Appendix contains the Function Importance survey to determine the FUI ranking of the TV functions (in Dutch).

D.1 FUI Questionnaire

Het kopen van een nieuwe TV is zowel een tijdrovende als een moeilijke beslissing. Stel je voor dat je, na een uitgebreide oriëntatie, een aantal potentieel geschikte TV’s hebt gevonden in jouw prijsklasse. Hieronder staat een tabel met functies, die gebruikt kunnen worden om de TV’s te beoordelen. Geef elke functie een score tussen de 0-10 in overeenstemming met hoe belangrijk die functie is in jouw keuze voor een bepaalde TV. De scores hoeven niet op te sommen tot 10, je kunt de eerste functie een rating van 8 geven en de tweede een rating van 5, etc. Een hoge score betekent dat je die functie belangrijk vindt in jouw keuze voor aanschaf.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Functies</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Kijken van het gewenste programma</td>
<td>Kijk het gewenste programma op TV, zonder mankementen in beeld en geluid.</td>
</tr>
<tr>
<td>2.</td>
<td>Gebruik van teletekst</td>
<td>Gebruik teletekst en alle functies die ertoe behoren.</td>
</tr>
<tr>
<td>3.</td>
<td>Automatische zender installatie</td>
<td>Automatisch instellen van de tv-zenders op kanalen, zonder je eigen voorkeuren in acht te nemen.</td>
</tr>
<tr>
<td>4.</td>
<td>Handmatige zender installatie</td>
<td>Handmatig instellen van de tv-zenders op kanalen die door jou gekozen zijn.</td>
</tr>
<tr>
<td>5.</td>
<td>Instellen van voorkeuren</td>
<td>Stel jouw voorkeuren voor beeld, geluid en diversen in. Voorbeelden hiervan zijn, contrast, helderheid, hoge en lage tonen, balans, surround sound mode, ondertitel opties.</td>
</tr>
<tr>
<td>6.</td>
<td>Switch on timer</td>
<td>Schakel de TV automatisch in op een ingesteld tijdstip op de gewenste zender vanuit de stand-by stand.</td>
</tr>
<tr>
<td>7.</td>
<td>Sleep timer</td>
<td>Stel een periode in waarna de TV automatisch naar de stand-by stand overgaat.</td>
</tr>
<tr>
<td>8.</td>
<td>Kijk twee programma’s tegelijk</td>
<td>Kijk twee programma’s tegelijk met Picture in Picture (PiP). In het hoofdscherm wordt een subscherm opgeroepen waardoor je naar twee tv-zenders tegelijk kunt kijken, of naar één zender en teletekst. Voor een grafische weergave, zie bijlage 1 op blad 5.</td>
</tr>
<tr>
<td>9.</td>
<td>Kinderslot</td>
<td>Zet een kinderslot op bepaalde programma’s of zenders (met programmering ongeschikt voor kinderen) om ze te vergrendelen.</td>
</tr>
<tr>
<td>Nr.</td>
<td>Functies</td>
<td>Score</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>10.</td>
<td><strong>Mute</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schakel het geluid van de TV uit, met één druk op de knop.</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td><strong>Veranderen van het beeldformaat</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selecteer het gewenste beeldformaat. Verschillende beeldformaten zijn: Automatisch, Super zoom, 4:3, 14:9, 16:9, Ondertitel zoom en Breedbeeld. Voor een grafische weergave, zie bijlage 2 op blad 5.</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td><strong>Vorige TV-zender</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schakel tussen de TV-zender waar u nu naar kijkt en de zender waar u het laatst naar gekeken hebt door één druk op de knop.</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td><strong>Gebruik van de gemotoriseerde draaivoet</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gebruik de gemotoriseerde draaivoet om het scherm te draaien tussen + en – 30 graden, voor een optimale kijkhoek. Voor een grafische weergave, zie bijlage 3 op blad 5.</td>
<td></td>
</tr>
</tbody>
</table>

### Algemene vragen

1. **Wat is je leeftijd?**
   - < 20
   - 20-25
   - 26-30
   - 30 >

2. **Wat is je geslacht?**
   - man
   - vrouw

3. **Ben je een student?**
   - ja
   - nee

4. **Zo ja, welke faculteit en generatie?**

5. **Hoe duur is de duurste TV in jouw huishouden?**
   - < € 500
   - € 500 - € 1000
   - > € 1000
   - weet ik niet

6. **Hoe oud is die TV?**
   - < 2 jaar
   - 2 – 6 jaar
   - > 6 jaar
   - weet ik niet
7. Welke van de volgende functies weet je zeker dat die TV bevat?
   a. Teletext
   b. Automatische zender installatie
   c. Handmatige zender installatie
   d. Instellen van voorkeuren voor beeld, geluid en diversen
   e. Switch-on timer
   f. Sleep timer
   g. Picture in Picture
   h. Kinderslot
   i. Mute
   j. Veranderen van beeldformaat
   k. Vorige zender
   l. Gemotoriseerde draaivoet
Appendix E: Data Analysis FUI Survey

This Appendix contains the data analysis results of the FUI survey.

E.1 Normality test

The Kolmogorov-Smirnov test was used to determine if the observed distribution corresponds with a normal distribution. The statistics of this test are presented in Table E.1.

<table>
<thead>
<tr>
<th>Function</th>
<th>Number of respondents</th>
<th>Normal Parameters</th>
<th>Asymptotic Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch the desired program</td>
<td>121</td>
<td>9.3</td>
<td>0.000</td>
</tr>
<tr>
<td>Teletext</td>
<td>121</td>
<td>6.1</td>
<td>0.002</td>
</tr>
<tr>
<td>Automatic channel installation</td>
<td>121</td>
<td>5.6</td>
<td>0.004</td>
</tr>
<tr>
<td>Manual channel installation</td>
<td>121</td>
<td>7.0</td>
<td>0.000</td>
</tr>
<tr>
<td>Set-up of preferences</td>
<td>121</td>
<td>6.9</td>
<td>0.000</td>
</tr>
<tr>
<td>Switch on timer</td>
<td>121</td>
<td>2.7</td>
<td>0.000</td>
</tr>
<tr>
<td>Sleep timer</td>
<td>121</td>
<td>3.1</td>
<td>0.010</td>
</tr>
<tr>
<td>Watch two programs</td>
<td>121</td>
<td>4.0</td>
<td>0.083</td>
</tr>
<tr>
<td>Child lock</td>
<td>121</td>
<td>3.1</td>
<td>0.000</td>
</tr>
<tr>
<td>Mute</td>
<td>121</td>
<td>7.1</td>
<td>0.000</td>
</tr>
<tr>
<td>Change picture format</td>
<td>121</td>
<td>6.3</td>
<td>0.002</td>
</tr>
<tr>
<td>Previous channel</td>
<td>121</td>
<td>6.2</td>
<td>0.006</td>
</tr>
<tr>
<td>Motorized swivel</td>
<td>121</td>
<td>3.9</td>
<td>0.046</td>
</tr>
</tbody>
</table>

The analysis shows that, all but one scores obtained in the FI measurement have small significance values (< 0.05) and therefore do no correspond to a normal distribution. The only rating that corresponds to a normal distribution is the rating of the function “Watch two programs” (significance value: 0.083), as can be seen in Table E.1.

E.2 Influence of possession on FUI rating

The Mann-Whitney U test is performed to check on differences in the mean score for a function between people who own a TV that possesses certain functions and people who own a TV that does not possess those functions. The observation from both groups are combined and ranked, with the average rank assigned in case of equal scores. The test calculates the number of times a score from group 1 (in possession rating) precedes a score from group 2 (not in possession rating) and the number of times a score from group 2 precedes a score from group 1. The Mann-Whitney U statistic is the smaller of these two numbers.
The Wilcoxon rank sum \( W \) statistic is the rank sum of the smaller sample. Small asymptotic significance values (<0.05) indicate that the two groups have different locations. As can be seen from Table E.2, the analysis shows that there are two functions with a significantly different mean for “possession” and “no possession”.

**Table E.2: Mann-Whitney U test statistics**

<table>
<thead>
<tr>
<th>Function</th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Asymptotic Significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teletext</td>
<td>155.5</td>
<td>161.5</td>
<td>0.717</td>
</tr>
<tr>
<td>Automatic channel instal.</td>
<td>1 002.5</td>
<td>1 302.5</td>
<td>0.288</td>
</tr>
<tr>
<td>Manual channel install.</td>
<td>441.5</td>
<td>496.5</td>
<td>0.274</td>
</tr>
<tr>
<td>Set-up of preferences</td>
<td>361.0</td>
<td>427.0</td>
<td><strong>0.024</strong></td>
</tr>
<tr>
<td>Switch on timer</td>
<td>1 032.5</td>
<td>5 497.5</td>
<td>0.136</td>
</tr>
<tr>
<td>Sleep timer</td>
<td>1 748.5</td>
<td>3 764.5</td>
<td>0.681</td>
</tr>
<tr>
<td>Watch two programs</td>
<td>1 076.0</td>
<td>1 329.0</td>
<td>0.930</td>
</tr>
<tr>
<td>Child lock</td>
<td>1 324.5</td>
<td>2 114.5</td>
<td>0.100</td>
</tr>
<tr>
<td>Mute</td>
<td>526.5</td>
<td>581.5</td>
<td>0.783</td>
</tr>
<tr>
<td>Change picture format</td>
<td>1 442.5</td>
<td>2 222.5</td>
<td>0.378</td>
</tr>
<tr>
<td>Previous channel</td>
<td>1 140.0</td>
<td>2 736.0</td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td>Motorized swivel</td>
<td>22.5</td>
<td>23.5</td>
<td>0.280</td>
</tr>
</tbody>
</table>

The functions “Set-up of preferences” and “Previous channel” have small significance values (<0.05), which indicates a significant difference in mean average scores between respondents who own the considered function and respondents who do not own this function. The function “Set-up of preferences” has a significantly higher mean score for respondents who possess that function compared to respondents who do not possess that function (sign. 0.024). Respondents with a TV that possesses that function evaluate the function with a 7.0. The mean score given by respondents with a TV that does not have that function is only 5.7. This indicates that people might not see this function as valuable until they have had experience with it. The function “Previous channel” has a significantly higher mean score for respondents who possess that function compared to respondents who do not possess that function (sign. 0.000). The mean score with possession is 6.9, while the mean score without possession is only 5.4. This again indicates that people might not feel the advantage of the function unless they have tried it for themselves.

There were three functions that have three or fewer respondents in either the category “in possession” or “not in possession” (see Table 6.4). Because of the small numbers it is difficult to determine a difference in rating between “in possession” and “not in possession” for these functions. The function “Motorized swivel” was in possession of only one respondent and the functions “Watch the desired program” and “Teletext” were not in possession of respectively any and only three respondents. Therefore FUI ratings for these functions cannot be compared among the two groups.

190
Appendix F: Adjusted UPFS measurement

This Appendix contains the survey of the adjusted approach to measure UPFS and participant reaction (in Dutch).

F.1 Survey

Naam: ...............................................................
Datum: ..............................................................
Scenario: ...........................................................

Stel je nu de situatie voor waarin je deze televisie hebt gekocht. Hij staat sinds een paar dagen bij jou in de huiskamer. Vervolgens treden de problemen op in de televisie die jij net ook opmerkte. Hoe zou jij reageren?

<table>
<thead>
<tr>
<th></th>
<th>Zeer onwaarschijnlijk</th>
<th>Zeer waarschijnlijk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Ik zou deze fout enorm irritant vinden.
2. Ik zou de klantenservice van bedrijf X raadplegen voor hulp.
3. Ik zou me behoorlijk opwinden over dit probleem.
4. Ik zou gewoon wachten totdat het probleem verdwenen was.
5. Over een dergelijke fout zou ik zeker niet boos worden.
6. Deze televisie zou ik direct terugbrengen naar de winkel!
7. Ik zou op het Internet gaan zoeken naar mogelijke oplossingen voor dit probleem.
8. Thuis zou ik deze fout beschouwen als een kleine tekortkoming in de teletekst functie
9. Als dit probleem thuis op zou treden zou ik een familielid/kennis om hulp vragen
10. Aan een dergelijk probleem zou ik met niet ergeren...
Appendix G: PANAS-X

This Appendix contains the PANAS-X measurement scale to determine UPFS (in Dutch). This scale is used twice per experiment: just before and just after the failure occurs. The difference between these two measurements is considered a measure for UPFS.

G.1 PANAS-X measurement scale


| 1 | vrolijk | 31 | actief |
| 2 | walgen van | 32 | schuldbewust |
| 3 | oplettend | 33 | opgewekt |
| 4 | beduusd | 34 | nerveus |
| 5 | lui | 35 | eenzaam |
| 6 | moedig | 36 | slaperg |
| 7 | verbaasd | 37 | uitgelaten |
| 8 | sterk | 38 | vijandig |
| 9 | minachtend | 39 | trots |
| 10 | ontspannen | 40 | zenuwachtig |
| 11 | prikkelbaar | 41 | levendig |
| 12 | opgetogen | 42 | beschamel |
| 13 | geïnspireerd | 43 | op je gemak |
| 14 | onbevreesd | 44 | angstig |
| 15 | walgend van jezelf | 45 | suf |
| 16 | verdrietig | 46 | boos op jezelf |
| 17 | kalm | 47 | enthousiast |
| 18 | bang | 48 | terneergeslagen |
| 19 | moe | 49 | bedeeds |
| 20 | verwonderd | 50 | van streek |
| 21 | rusteloos | 51 | schuldig |
| 22 | gelukkig | 52 | aandachtig |
| 23 | timide | 53 | bevreesd |
| 24 | alleen | 54 | versteld staan |
| 25 | alert | 55 | geïnteresseerd |
| 26 | overstuur | 56 | afkeer hebben van |
| 27 | boos | 57 | zelfverzekerd |
| 28 | dapper | 58 | energiek |
| 29 | zwaarmoedig | 59 | geconcentreerd |
| 30 | verlegen | 60 | ontvreden met jezelf |
Appendix H: Function Failure List

This Appendix contains the function failure list of the FUI consumer experiment (in Dutch).

H.1 Function failure list

Hieronder staat een tabel met functies die de door jou geteste TV bezit. Geef voor al de door jou geteste functies aan of ze naar jouw mening foutloos werkte (zet dan een kruisje in het vakje foutloos) óf juist faalde tijdens het gebruik van de TV. Indien jij van mening bent dat (een) bepaalde functie(s) faalde(n), kun je dan een score tussen de 0-5 geven voor de ernst van dit falen. Een 0 score betekent dat de functie nog goed te gebruiken was en een 5 score betekent dat de functie bijna onbruikbaar was geworden. Als je een bepaalde functie niet hebt gebruikt kruist kruis dan NVT aan.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Functies</th>
<th>Foutloos</th>
<th>Fout, Score:</th>
<th>NVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Kijken van het gewenste programma</td>
<td>Kijk het gewenste programma op TV, zonder mankementen in beeld en geluid.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Gebruik van teletekst</td>
<td>Gebruik teletekst en alle functies die ertoe behoren.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Kijk twee programma’s tegelijk</td>
<td>Kijk twee programma’s tegelijk met Dual Screen. In het hoofdscherm wordt een subscherm opgeroepen waardoor je naar twee tv-zenders tegelijk kunt kijken, of naar één zender en teletekst.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Mute</td>
<td>Schakel het geluid van de TV uit, met één druk op de knop.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Vorige Tv-zender</td>
<td>Schakel tussen de Tv-zender waar u nu naar kijkt en de zender waar u het laatst naar gekeken hebt door één druk op de knop.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix I: Acquaintance Task List

This Appendix contains the acquaintance task list of the FUI consumer experiment (in Dutch).

I.1 Acquaintance task list

Hieronder staan een aantal taken beschreven die je kennis laten maken met de TV en de verschillende functies ervan. Wij willen je vragen om alles wat je doet en denkt hardop te zeggen zodat wij kunnen volgen wat je doet en denkt. Mocht je er in het uiterste geval niet uit komen kun je de observatoren vragen om hulp.

Het geluid van de TV staat nu uit, zet het geluid aan door op de mute toets in het midden van de afstandsbediening.

Eerst willen we je kennis laten maken met Dual screen. Met Dual screen kun je het hoofdscherm indelen in twee subschermen, waardoor je naar twee TV-zenders tegelijk kunt kijken, of naar een TV-zender en een externe bron (DVD-speler, PC, etc.) of naar een TV-zender en Teletekst.

Beeld/Beeld

- Druk op de toets om Dual screen aan of uit te schakelen.
- Druk op de cursortoets links/rechts om het linker/rechterscherm in Dual screen te doen oplichten.
- Druk op de toets of de cijfertoetsen om de TV zender, de externe bron of de teletekstpagina te veranderen in het scherm dat oplicht.

Opmerking: De scherminformatie die verschijnt wanneer u van zender verandert heeft altijd betrekking op het scherm dat oplicht.

- Druk opnieuw op de toets om een volledig hoofdscherm te krijgen. De TV stemt dan af op het beeld dat in de Dual screen-mode oplichtte.

Opmerking: Alleen het geluid van het linkerscherm wordt gehoord als Dual screen wordt geselecteerd.

Beeld/Teletekst

- Druk op de toets. Druk op de toets om teletekst op de rechterkant van het scherm te laten verschijnen.
- Druk op de toets om terug te keren naar een volledig teletekst scherm.
- Druk op de toets om terug te keren naar een volledig beeldscherm.
Gemotoriseerde draaivoet
Nu je kennis hebt gemaakt met Dual screen functie willen we je graag kennis laten maken met de gemotoriseerde draaivoet. De gemotoriseerde draaivoet kan 30 graden draaien en kan met de afstandbediening van het TV-toestel worden bediend. Er zijn ook TV’s te verkrijgen die een draaivoet hebben die handmatig te bedienen is.

Druk eerst op de **SWIVEL MOT** toets aan de zijkant van de afstandbediening en houd hem ingedrukt. Terwijl je deze toets ingedrukt houdt:
- gebruik de cursor links/rechts om de draaikop te doen draaien, draai het scherm van links naar rechts om te zien hoe dit werkt;
- beweeg de cursor naar beneden om de draaikop in de centrale stand te brengen;
- beweeg de cursor naar boven om de draaikop in de laatst geselecteerde stand te zetten.

Nu je kennis hebt gemaakt met deze functies zouden we willen vragen om deze functies een function importance rating te geven. Stel je hierbij weer voor dat je na een uitgebreide oriëntatie, een aantal potentieel geschikte TV’s hebt gevonden in jouw prijsklasse. Hieronder staat een tabel met functies die gebruikt kunnen worden om de TV’s te beoordelen.

Geef elke functie een score tussen de 0-10 in overeenstemming met hoe belangrijk die functie is in **JOUW KEUZE VOOR AANSCHAF** voor een bepaalde TV. Het gaat hierbij dus **NIET** om de kwaliteit van de functie op **DEZE** TV maar om hoe belangrijk die functie is in jouw keuze voor aanschaf.

De scores hoeven niet op te sommen tot 10, je kunt de eerste functie een rating van 8 geven en de tweede een rating van 5, etc. Een hoge score betekent dat je die functie belangrijk vindt in jouw keuze voor aanschaf.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Functies</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Het kijken van het gewenste programma met goede beeld en geluidskwaliteit</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Het gebruik van teletekst</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Het instellen van voorkeuren voor beeld, geluid en diversen</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Het gebruik van de Mute knop om het geluid uit de te schakelen</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Het gebruik van Dual screen voor het kijken van twee programma’s tegelijk</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Het gebruik van Dual screen voor het kijken van één programma en teletekst tegelijk</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Het gebruik van een handmatige draaivoet voor het instellen van de kijkhoek</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Het gebruik van een motorische draaivoet voor het instellen van de kijkhoek</td>
<td></td>
</tr>
</tbody>
</table>

Appendix J: Scenario 1: Task List

This Appendix contains the task list scenario 1 (important function) of the FUI consumer experiment (in Dutch).

J.1 Task list scenario 1 (important function)

Nu je kennis hebt gemaakt met een aantal verschillende functies van deze TV, gaan we verder met het tweede deel van het experiment.

Wij willen je vragen om alles wat je doet en denkt hardop te zeggen zodat wij kunnen volgen wat je doet en denkt. Mocht je er in het uiterste geval niet uit komen kun je de observatoren vragen om hulp.

Om de informatie met betrekking tot het experiment op te kunnen slaan willen wij je erop wijzen dat je de TV net als thuis kunt gebruiken behalve dat je niet aan de kabels van de TV mag komen!

Voer de volgende opdrachten uit alsof je thuis in de woonkamer voor de televisie zit.

- Begin op Nederland 1 en zap met de afstandsbediening van het ene naar het andere programma om te kijken wat er uitgezonden wordt.
- Probeer ook enkele andere functies van de TV uit waar je nog geen kennis mee hebt gemaakt. Je kunt gerust de knoppen van de afstandsbediening uitproberen.
- Vul de volgende tabel in, zet achter het kanaalnummer in eigen woorden het programma dat uitgezonden wordt op dat kanaal. Gebruik hiervoor één woord of een korte zin.

<table>
<thead>
<tr>
<th>Kanaal</th>
<th>Programma</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Nadat je de bovenstaande tabel hebt ingevuld zet je het geluid weer uit met de  toets. Vul nu de vragenlijst in die je van de observator krijgt.
Appendix K: Scenario 2: Task List

This Appendix contains the task list scenario 2 (unimportant function) of the FUI consumer experiment (in Dutch).

K.1 Task list scenario 2 (unimportant function)

Nu je kennis hebt gemaakt met een aantal verschillende functies van deze TV, gaan we verder met het tweede deel van het experiment.

Wij willen je vragen om alles wat je doet en denkt hardop te zeggen zodat wij kunnen volgen wat je doet en denkt. Mocht je er in het uiterste geval niet uit komen kun je de observatoren vragen om hulp.

Om de informatie met betrekking tot het experiment op te kunnen slaan willen wij je erop wijzen dat je de TV net als thuis kunt gebruiken behalve dat je niet aan de kabels van de TV mag komen!

Voer de volgende opdrachten uit alsof je thuis in de woonkamer voor de televisie zit.

- Je neemt plaats op de bank voor de televisie en zap t met de afstandsbediening van het ene naar het andere kanaal om te kijken wat er uitgezonden wordt.
- Je stopt met zappen als je een programma hebt gevonden dat jou aanspreekt.
- Probeer ook enkele andere functies van de TV uit waar je nog geen kennis mee hebt gemaakt. Je kunt gerust de knoppen van de afstandsbediening uitproberen.
- Nadat je nog een aantal van de functies van de TV hebt geprobeerd ga je achter de laptop zitten. Op de laptop vind je een opdracht die je nu uit moet voeren. Je kunt het jezelf gemakkelijk maken door het scherm van de TV naar de laptop toe te draaien.
- Nadat je de opdracht hebt uitgevoerd ga je weer terug op de bank zitten, om verder TV te kijken. Om zo relaxed mogelijk TV te kijken kun je het scherm weer terug draaien naar de beginpositie.

Vul nu de vragenlijst in die je van de observator krijgt, en zet het geluid weer uit met de toets.
Appendix L: Failure Attribution Questionnaire

This Appendix contains the Failure Attribution Questionnaire of the FA consumer experiment (in Dutch).

L.1 Failure Attribution Questionnaire

Denk aan de oorzaak van de zojuist getoonde fout. Wie denk je dat verantwoordelijk is voor het ontstaan van de zojuist getoonde fout? Hieronder staan steeds twee tegengestelde stellingen. Vink de cirkel aan die het beste de juiste situatie weergeeft.

Bijvoorbeeld: Ben je het geheel eens met de linker stelling vink dan de linker cirkel aan. Ben je het enigszins eens met de rechter stelling vink dan de tweede cirkel van rechts aan.

De fout gebeurde door iets wat de TV deed
De fout is een aspect van de prestatiekwaliiteit van de TV
De fout werd veroorzaakt door een onderdeel van de TV

De fout gebeurde door iets wat andere dingen of personen deden
De fout geeft een aspect weer van de prestatiekwaliiteit van ander dingen
De fout werd veroorzaakt door een iets anders
Ilse de Visser was born in Eindhoven, the Netherlands, on 31 December 1980. In 2004, she received her Masters degree (cum laude) in Industrial Engineering and Management Sciences from the Eindhoven University of Technology (TU/e). The topic of her graduation project was reliability optimization of photocopiers using degradation analysis.

In September 2004, she started her Ph.D. work at the sub-department Quality and Reliability Engineering of the department of Technology Management at the TU/e. The research was performed in cooperation with various institutes (Embedded Systems Institute and Design Technology Institute), companies (NXP, TASS, Philips Consumer Electronics and IMEC), and universities (Delft University of Technology, Leiden University and University of Twente).

Since June 2007 she is working as teacher/researcher at the sub-department of Business Process Design of the department of Industrial Design at the TU/e. Her research interest lies in incorporating the user perspective into the development process of highly innovative products. She teaches several courses in the area of new product development to students of different backgrounds and levels.