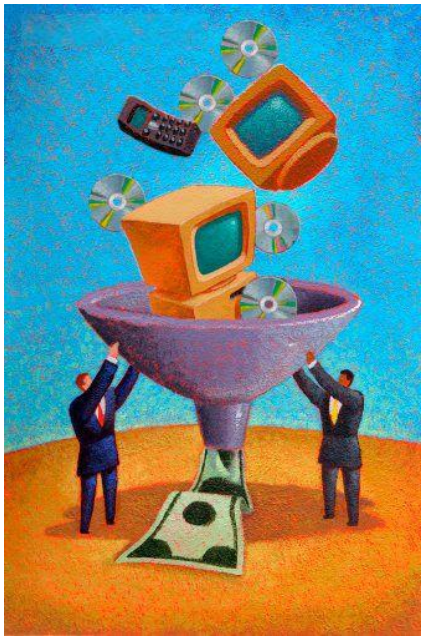
Towards the market

Improving the Testing & Validation process of new products for Company X



Master Thesis

Author:	Ruben Velstra s0069124 r.t.velstra@student.utwente.nl
University:	University of Twente School of Management & Governance Business Administration
Company:	X
Supervisors:	Dr. Ir. Jeroen Kraaijenbrink (UT) Dr. Chintan Amrit (UT)

Acknowledgements

This master thesis is the final product of my master program Innovation Management and Entrepreneurship at the University of Twente and was conducted at Company X in Amsterdam.

At this point I would like to seize the opportunity to thank everybody who was involved in this research project. First of all, I would like to thank my supervising manager at Company X, who provided me the opportunity to finish my studies in exactly the field of my interest. In our short, but inspiring meetings he provided me with support and deeper insights. I would also like to thank my colleague for his help with getting familiar with the practice at X and for his committed mentoring during the process. Furthermore, I would like to thank all the company X employees that helped me during the project and who made my internship a valuable and enjoyable experience.

Next, I would like to thank my first supervisor of the University of Twente, Jeroen Kraaijenbrink, for his support and advice during my research. I really appreciated his fast replies on my questions and the motivating meetings. I would also like to thank Chintan Amrit, my second supervisor, for his detailed feedback and support.

My aim was to provide a practical report for the Company X, while at the same time delivering a profound piece of scientific work. Therefore I hope this thesis report will prove useful and an enjoyable reading.

Ruben Velstra

Amsterdam, 2009

Management Summery

The subject of this research study is the testing and validation process of new products at Company X. Company X is responsible for bringing new products and is therefore involved in several testing and validation processes. The need to structure this process formed the main motive for this research study. As X is currently leading a consortium, which is planning to set up a large validation test of several products, the development of an execution model for this process is now of extra importance. The main research question in this research study aims at developing this model:

"How should a model to execute the testing and validation stage be structured for Company X?"

The research started off with an extensive theoretical analysis to obtain understanding of the testing & validation stage within a new product development process. During this analysis is turned out that also in the literature a model to execute this stage does not exist, which underlined the relevance of the research. By defining the purpose and boundaries of this stage, identifying overall test dimensions and identifying guidelines for designing tests within this stage, a theoretical model could be developed. This model is still relative general, because from a theoretical point of view the testing and validation stage can comprise a variety of testing types and has no unconditional form.
In order to specify and structure the model for Comp. X, the practice of Comp. X was analyzed on two points:
-The extent to which the aspects of the theoretical model were already applied
-The design requirements of Comp. X itself

The findings on both points were combined and analyzed together, to adjust the theoretical testing and validation model on. This way an effective model was developed which suited the work methods of Comp. X, while at the same time having its basis in theory.
Important aspects in the model are: proper evaluation during and after validation tests, specific focus on certain test dimensions in the different test phases and guidelines on the selection of participants, roll out plan and decisions making processes in tests.
The next step in this research was to apply the developed model on the test Comp. X is planning with the consortium, to develop a proper research design and determine the model's value and relevance in practice. The findings showed that even though not all the aspects of the developed model could be applied exactly as intended, each one of them proved to be relevant and valuable. It is insurmountable that practical dependencies or pressure from higher management will lead to compromises between an ideal and actual test. However, Comp. X should always try to apply the developed model as precisely as possible. The developed model is reflected in figure 15.
To make sure that the aspects of the developed model are not overlooked or overruled in a project, there should always be made someone specifically responsible for guarding the research requirements.
Recommendations for the implementation of the model include: Explicitly hold to all elements, Formalize the evaluation process, Integrate effective portfolio management, Realize organizational commitment and Evaluate the new model with the whole department.
As the developed (theoretical) model can also be applied at other organizations, the research study is also interesting from a scientific perspective and made a start in filling in a blank spot in literature. Furthermore it provided interesting starting points for further research.

Table of Contents

Page:

Table of figures

Page:

Figures

Tables

1. Introduction to the research

This chapter begins with a paragraph on the research background of this study, in which the main motive for conducting the research is identified (1.1.1). Subsequently a specific project is described which formed an additional motive for this research (1.1.2) and the theoretical context of the research topic is explained (1.1.3). In the second paragraph the research objective is formulated and the practical and scientific relevance of the research study is discussed. The chapter ends with paragraph 1.3 in which research questions are established to structure the research study.

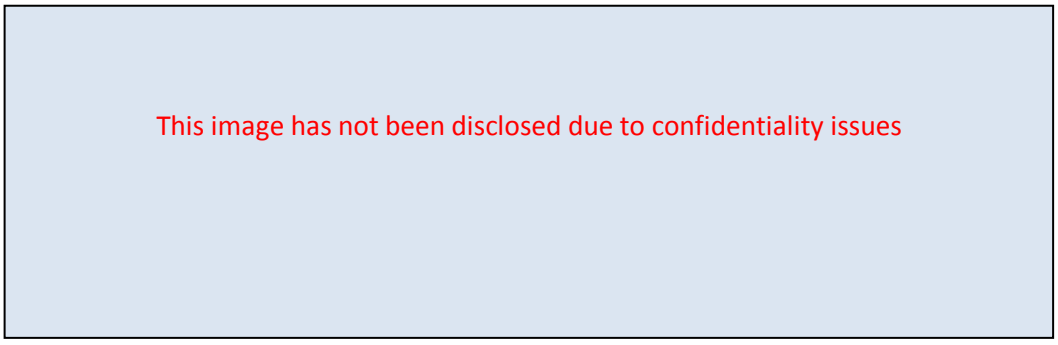
1.1 Research background

1.1.1 Comp. X Retail

Comp. X Retail is the business unit responsible for selling gas and electricity to consumers and small and medium-sized enterprises. Additionally the business unit tries to improve customer service and margins by bringing new energy related products and services to the market. This is in line with the strategy of Comp. X to transform its traditional position as an energy provider into an energy advisor, which actively facilitates consumers in their energy choices. The mission of Comp. X Retail "to be an inspiring energy company that creates sustained value and growth by establishing new norms" emphasizes this transformation.

The New Business Development (NBD) department plays an important role in this transformation. NBD can be considered the knowledge centre of Comp. X with respect to new products and technologies, energy saving and environmental sustainability. It scans the international energy market for new developments and stimulates the innovative development of the company. NBD is responsible for the development of new products, services, markets and channels with a market potential of ten percent or more and a positive contribution to Comp. X's Retail margin.

While in the past Comp. X has been focusing mainly on the development of new services and markets, they have been spending considerable amounts of time and resources on the development of new products over the last two years. Because the development of new products calls for a different approach than developing services, Comp. X is currently trying to structure this development. They developed a stage gate model (Figure 1) for their innovation process, in which the development of a new product is divided in several stages. These are visualized by the blue blocks.



The general idea is that before a project can enter a new stage, certain criteria need to be met and a gate must be passed. These gates are the circles in figure 1. Although this model is already developed, not all characteristics of the different stages are agreed upon. This model is derived from the stage gate model of Cooper (2001) and will be discussed in more detail in paragraph 1.1.3.

Figure 1: NPD stage-gate model Company X

An important stage in this model is the testing and valuation stage. This is the final stage before a product or technology is launched into the market. While Comp. X has developed models and formats to structure the other stages, the execution of the testing and development stage remains vague. Comp. X did subdivide this stage into three new phases. An alpha, beta and gamma phase which represent three tests ascending in size. In an alpha test 1 prototype is used, in a Beta test approximately 20 and a Gamma test comprises at least 100 prototypes. However, this division is still general and exact characteristics of the stage and its phases need to be developed, just as a model to execute them. This is the most important motive for this research assignment.
The specification and clarification of how the testing and development process should be structured is currently of extra importance, as a large validation project of several products is scheduled. This project is executed by a consortium of organizations with Comp. X in the lead. This project forms an additional motive and will further be explained in the following section.

1.1.2 The West Orange project

A particular field of interest of the NBD department is “Home automation technologies”. Home automation technologies cover the broad domain of domestic electronics and concern the application of computing and information technology in residential homes (Aldrich 2003).
Consumer needs for home automation technologies on the areas of energy saving and environmental sustainability have been demonstrated in various studies (Barlow 1997, Gann 1999, Ilmer, 2009).
General interest in home automation technologies is currently increasing as the Dutch government is planning to roll-out smart energy meters in all Dutch households the coming years. A smart meter provides near real time data about energy and gas consumption in a household and can therefore serve as a platform for other smart energy home solutions.
Not surprisingly, the development of activities on home automation is high on the NBD agenda. In order to fulfil the identified consumer needs and act on the opportunities provided by the smart meter, Comp. X co-developed four new home automation products: An energy display, an E-wizard an online thermostat and a central on-off button.
These products follow the line of reasoning that measuring energy leads to insight in the energy consumption, which can lead to the actual control of energy consumption. In figure 2 the four developed products are reflected as well as the way they relate to each other.

Figure 2: Energy management products

This section has not been disclosed due to confidentiality issues

The consortium that is executing this project consists of Comp. X, Y, Z, O. Apart from the consortium partners there are also governmental institutes involved and a university. Because all these organizations have different learning goals in the project there is a great need for a proper research design. Because Comp. X is in the lead, they are responsible for the development of this design. This is the reason that the West Orange project can be viewed as an additional motive for this research

study. Comp. X needs a model to structure the testing and development stage, in order to develop a proper research design.

1.1.3 Theoretical context of the research topic

The topic of this research, the testing and validation of new products, is a part of the new product development process. In this section the new product development process will be explained together with a common method to structure it.

The essence of new product development is to create or establish something new. Since this process necessarily involves risk, innovating companies require a strategy not of risk avoidance, but of early diagnosis and management (Keizer, Halman, & Song, 2002)

One means for managing the risks of new products is implementing decision gates in the new product development process, at which point a go or no-go decision is made. These points provide an assessment of the quality of the project and ensure that the company is working on the right projects in a correct way (Crawford, 1989).

Nowadays many companies are using stage models in which these decision gates are incorporated (Griffin, 1997). In such a model the new product development process is structured and divided in different stages which are separated by the decision gates.

A well known example of such a model, and also the basis of Comp. X's model, is the Stage-Gate system of Cooper (2001).

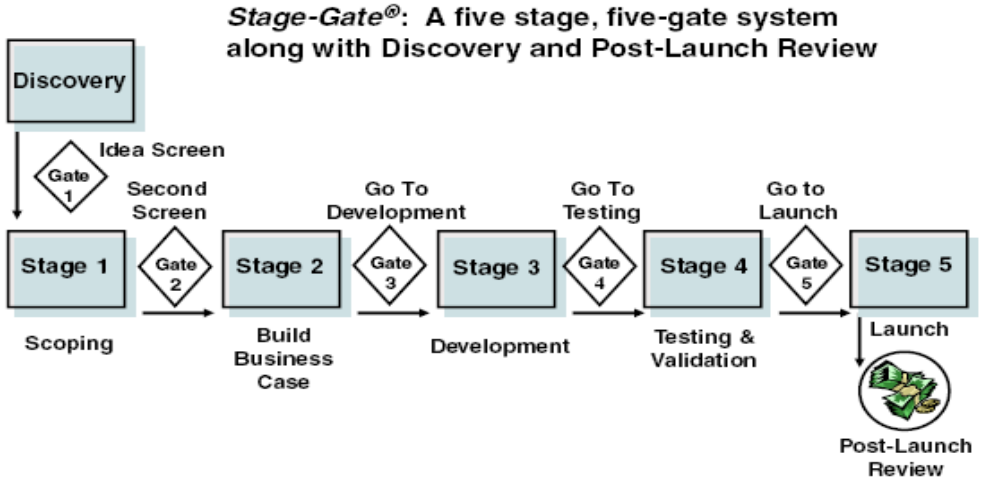


Figure 3: Stage-Gate model of the innovation process

In figure 3, the stage gate model of Cooper (2001) is visualized. The innovation process is divided into a series of five stages: Scoping, Build Business Case, Development, Testing & Validation and Launch. Discovery and Post-Launch review are also important but do not belong to the main process. Before the stage-gate process is explained, a short summery of the different stages is given:

Scoping: The first stage of the product development process is scoping. During this step the main goal is to evaluate new product ideas and their corresponding market. The researchers must recognize the strengths and weaknesses of a product and estimate what it is going to offer to the potential consumer.

Build Business Case: Together with the scoping stage this stage belongs to the "concept development" part of process. Since the actual development of a product is both time and money consuming it is necessary for organizations to perform a solid market- and competition analysis. Furthermore the design, marketing and business feasibility of the product are estimated. Based on the acquired information a business case and project plan can be drawn up.
Development: During this stage, plans made in the previous stages are executed. The product's design and development is carried out and some early tests are performed. Furthermore marketing and production plans are developed. The completion of the first working prototype marks the end of this stage.
Testing and Validation: The purpose of this stage is to provide validation for the entire project by testing the product in an actual customer environment.
Launch: The organization starts with the full production of a product and must come up with a marketing strategy to generate customer demand for the product.

Together these five stages comprise the entire NPD process.
Each stage is designed to gather information to reduce key project uncertainties and risks and each stage costs more than the preceding one. This way each increase in project costs is accompanied by a decrease of the unknowns and uncertainties, so that risks are effectively managed. Each stage in the stage-gate model is composed of a set of required or recommended best-practice activities needed to progress the project to the next gate or decision point (Cooper 2001).

The gates between the stages serve as quality-control check points. Go/kill/Hold/Recycle and prioritization decisions are made here, as well as a plan for the path forward (Cooper 2008). The structure of each gate is similar.
The current interest of Comp. X is limited to the testing and development stage as they lack, unlike the other stages, a suitable execution model. This is not strange as the testing and validation phase is still a field of research which is less investigated and there is no standard model or framework that outlines this process (Ozer, 1999). Regarding the West-Orange project, Comp. X wants to validate the four different products in a proper and structured way. Based on the findings of the West Orange project, refinements can be made to the products and/or the business case, a marketing strategy can be developed and the products can be launched into the market.

1.2 Research Objective and relevance of the research study

Following from the research background, the interest of Comp. X's management can be described as two folded. The first objective is to clarify and determine the characteristics of the testing and validation stage and to develop a standard model to execute this stage, suitable for Comp. X.
The second objective is the application of this model on the West Orange project in order to create a research design in which the different interests and learning goals of all stakeholders are represented. This design should furthermore include data collection methods and guidelines how to organize the process.
The interest of the management can be summarized into the following research objective:

> *"Develop a model for the testing and validation stage which can be used to validate new products, and apply this model on the West Orange project in order to create a research design that covers the information needs of all stakeholders and provides guidelines how to organize the process"*

Relevance

The relevance of this research study for Comp. X is evident. The developed model can be used as a standard tool for executing the testing and development stage and developing research designs for tests within this stage. An additional advantage of the developed model is that it leads to a better overview on the projects Comp. X is working on. Most businesses have too many development project underway (Cooper et al., 1999) and the new model will enable the management of Comp. X to compare products and/or projects with each other. They can then apply portfolio management to steer the resources towards the most promising projects.

The research study is also relevant from a scientific perspective. Since uncertainties and risk are inextricably linked in the development of new products, authors agree that these uncertainties should be managed by testing and validating new products. However, compared to the other stages of the NPD process the testing and validation stage is still a field of research which is less investigated (Ozer, 1999). There is no standard model or framework that outlines this process and it remains vague how tests within this stage should be executed. Cole (2002) therefore states that the challenge for quality practitioners and scholars is to develop a set of tools that allow us to improve the deployment and optimization of probe-and-learn strategies.

The contribution the model developed in this research study makes to the scientific base of knowledge of testing and validating products, consists of several points. At first it is likely that the developed model is also valuable for other organizations who are having problems with executing this stage. In any case the theoretical model that forms the basis of the developed model will be of use for other organizations. This theoretical model is more general then the final model, which is adjusted to fit the practice at Comp. X, and is therefore easier to apply at other organizations. Other organizations can adjust this theoretical model to their own work methods the same way as was done for Comp. X in this study. Furthermore the way different academic findings are combined, to build the theoretical and final model, are important for other academics involved in the testing and validation of new products. It can lead to a different perspective on this process and will provide directions for further research. The same holds for the analysis of the current testing and development practice at Comp. X. Since little is known about how organizations execute this stage, insight in problems an organization encounters is of great value.

1.3 Research Questions

In the previous paragraph it became clear that the research objective in this study is two folded.
First the development of a general model that can be used to execute the testing and validation stage and second the application of this model on the products within the West Orange project.

The fulfillment of the second research objective depends on the realization of the first one. Only when a suitable model is developed it becomes possible to apply it on a specific project. The main research question is therefore established to meet the first objective.
As the second objective implies the use of the developed model in practice, this provides an opportunity to determine its value and relevance. The second objective will therefore be addressed in a supporting research question and gives further substance to the first research objective. Based on these findings the developed testing and validation model can be made definite for Comp. X.

Main research question:

"How should a model to execute the testing and validation stage be structured for Company X?"

Supporting research questions:

1. How should a model to execute the testing and validation stage be structured according to theory?

2. How does Comp. X currently organize the testing and validation process of new products and what are their design requirements for an execution model of this process?

3. What are the implications of the practice at Comp. X for the testing and validation model which is being developed?

4. How does using the developed testing and validation model in practice, affect its value and relevance?

Research structure

The first supporting research question is used to build a theoretical model for the testing and validation stage. It consists of four steps; 1) Defining the purpose and boundaries of the testing and validation stage and the criteria for product tests within this stage, 2) Identifying the overall test dimensions in the testing and validation stage, 3) Identifying guidelines for designing tests within the testing and validation stage, 4) Integrating the findings of the first three steps into a theoretical model. This model reflects on how the testing and validation stage "should be executed" in general. These four steps are taken in chapter 2, the theoretical framework.
The second supporting research question is addressed in chapter 4; Practice at Comp. X. This question aims to verify to what extent the aspects from the theoretical model are already applied by Comp. X. It furthermore explores the requirements and design goals Comp. X has of its own for the

development of a testing and validation model. How these topics are explored will be explained in the methodology section, chapter 3.
The findings of chapter 4 are used in the third supporting research question to adjust and specify the theoretical model to match the practice and work methods of Comp. X. This is done in chapter 5; Combination of theory and practice. The outcome of this chapter is an adjusted testing and validation model in which practice and theory are combined.
Subsequently this new model is applied on the West Orange project in chapter 6; Application of the developed model on the West Orange project. Research question 4 reflects on this process and addresses the implications of using the developed model in practice, in terms of its value and relevance. Based on these insights a definite testing and validation model suitable for Comp. X is presented is chapter 7. This chapter furthermore addresses the contributions of the findings and recommendations for implementing the developed model in practice. The chapter, and research study, concludes with the limitations of the research study and directions for further research.
The steps that are taken in this research and the corresponding chapters are reflected in figure 4, the research structure:

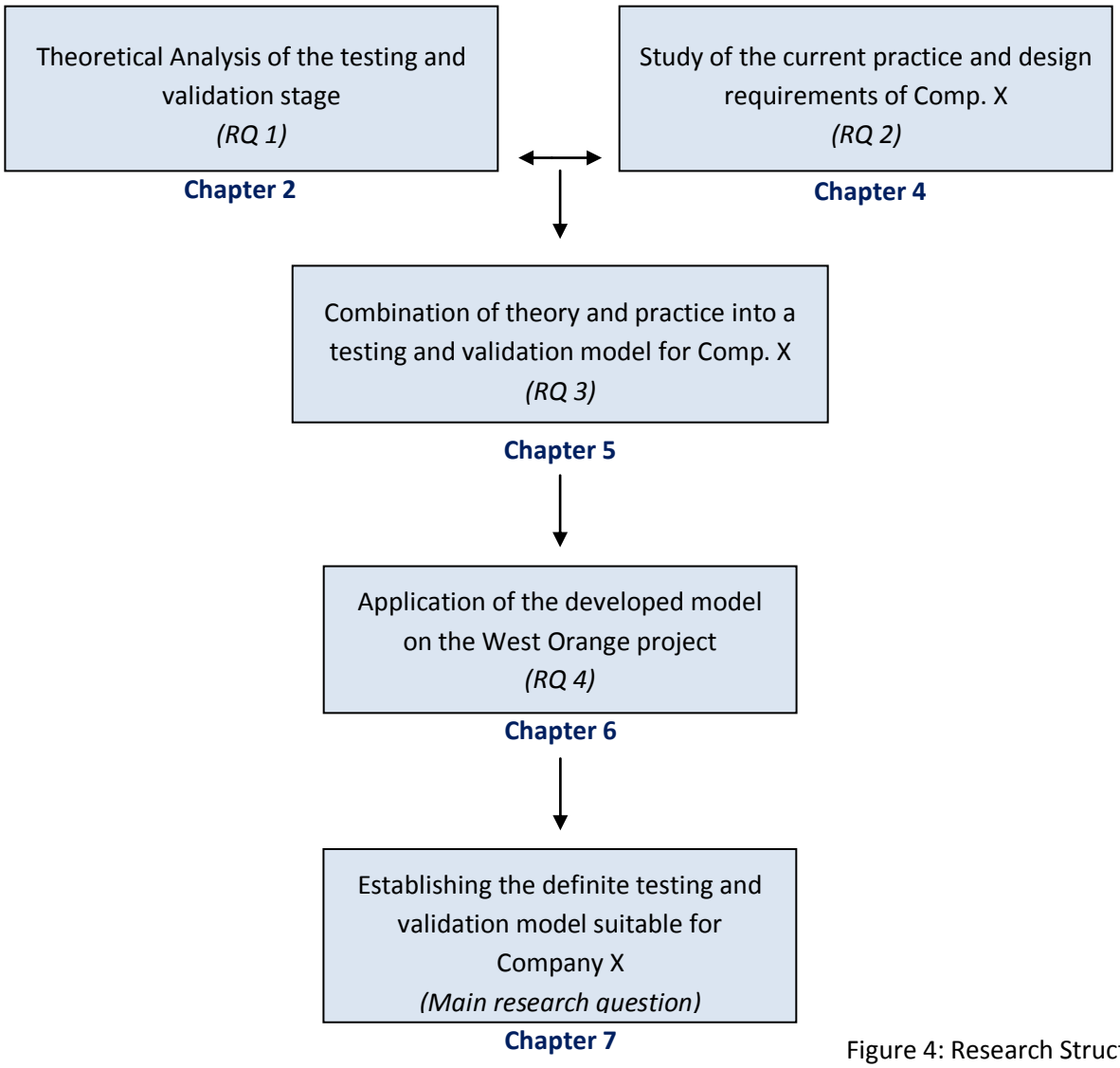


Figure 4: Research Structure
(RQ = Research Question)

2. Theoretical Framework

The result of this theoretical framework is an execution model for the testing and validation stage. That is, a general model of how the testing and validation stage should be executed according to theory. The building blocks of this model are formed by the first three paragraphs of this chapter. The purpose and boundaries of the testing and validation stage and the criteria for product tests within this stage are identified in paragraph 2.1. Subsequently overall testing dimensions for the testing and validation stage are identified in paragraph 2.2. The last step is the identification of guidelines for designing tests within the testing and validation stage in paragraph 2.3. In paragraph 2.4 the findings of these paragraphs are combined into a model.

2.1 The testing and validation stage

In the stage model of Cooper (2001) (and Comp. X) the testing and validation stage is the final phase before a product is launched into the market. The beginning of this phase is marked by the completion of a prototype that resembles all the functions of the final product. This prototype is the output of the stage 3, the development stage.
The main purpose in the testing and validation stage is to further develop the prototype into a "market ready product", by placing the prototype in a real world setting and validating its requirements. Since this is a very visible process the testing and development stage as a whole can be used for more purposes then just further developing the prototype.
Ullrich and Eppinger (2000) identify four purposes for the testing and validation stage: Learning, Communication, Integration and Milestones. While the purposes Learning and Integration are related to the development goal, the other two purposes are not. In the purposes Communication and Milestones the testing and development stage is viewed as a communication instrument. Working prototypes in a customer setting enrich the communication with all stakeholders (top management, vendors, partners, extended team members, customers and investors) as a physical prototype is much easier to understand than a verbal description or sketch of the product. It also provides a tangible goal and demonstrates progress which can serve to enforce the schedule.

In order to maximize the utility of the testing and development stage and the development process as a whole, product designers want to learn from consumer feedback in order to improve their product. Therefore, they will move through several testing and improvement "loops" before finalizing the product (Minot and Wood, 2003). Cooper (2005) also recognized this need and updated his model in order to accommodate such iterative processes. He build in a spiral or agile development in which project teams move through series of "build, test, feedback and revise" loops. Figure 5 shows an example of these spirals.

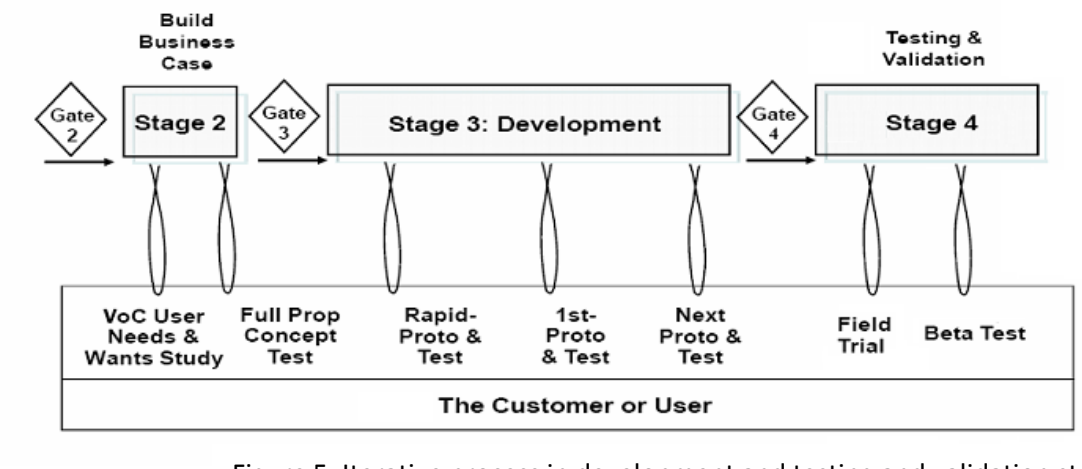


Figure 5: Iterative process in development and testing and validation stage

In figure 5 it becomes clear that this iterative process already starts in the beginning of the new product development cycle and becomes an integral part of the entire development process, including the testing and validation stage. The testing and validation stage should thus be seen as a way to develop a prototype in a market ready product and does not have an unconditional form. Depending on the characteristics of a product and working methods of an organization, the testing and validation stage can comprise one or more testing practices. This lack of definite form is reflected by the absence of a widespread structure for testing and validation in organizations (1999).

Another important aspect in Figure 5 are the different testing types that are distinguished. Definitions of these types differ in the literature and also other testing types are used. Common names for testing practices are concept testing; prototype testing; alpha testing; beta testing; pre-test market; Field testing; market testing; gamma testing or preproduction prototypes (Ulrich and Eppinger, 2008; Crawford and Benedetto, 2006; Ozer 1999; Vellandi, 2008).
The absence of a commonly used testing structure and the variety of testing types complicates the selection of appropriate literature. However there are two clear criteria for tests within the testing and validation stage:

- Physical prototypes which resemble the commercial model are used
- The testing is done outside the organization in a users' actual environment

These criteria follow from the starting point of the testing and validation stage: a working prototype that resembles the final product, tested inside the organization.
Now that the purpose and goal of the testing and validation stage is clear and the criteria for test within this stage established, the next step is to further specify the dimensions which needs to be addressed in a test.

2.2 Test dimensions in the testing and validation stage

In the previous paragraph it became clear that there are several purposes of the testing and development stage. The primary goal is to further develop the prototype into a "market ready product", through an iterative process. Since the market readiness of a product is determined by several aspects, there are also several dimensions on which a product should be tested in the testing

and validation stage. In this paragraph the key testing dimensions needed to fulfil the Learning- and Integration purpose (Ullrich and Eppinger 2000) and the primary goal of the testing and validation stage will be identified. This is done by discussing the view that Ullrich and Eppinger (2000), Cooper (2001) and Yang & El-Haik (2009) have on these test-dimensions and comparing their view with several studies, in which testing dimensions are used. This way overall testing dimensions can be derived.

Ullrich and Eppinger

According to Ullrich and Eppinger two types of questions are used to fulfil the learning purpose: "Will it work?" and "How well does it meet the customer needs?" Will it work questions are related to the product functioning in which technical requirements are validated. "How well does it meet the customer needs" is related to the customer acceptance of the product. What is the customer opinion on appearance, usability, etc. They thus identified the test dimensions *Product functioning* and *Customer acceptance.*

The integration purpose (Ullrich and Eppinger 2000) aims at validating whether components and subsystems of the product work together as expected. This way problems with integrating the product in the organization or specific infrastructure can be detected in an early stage. Also communication links with important parties can already be established which can be a great help when actually launching the product into the market. Since all aspects which this purpose includes are important to test, *Integration* in the view of Ullrich and Eppinger can also be seen as test dimension.

Cooper

In the stage gate methodology the purpose of the testing and validation stage is to provide validation for the entire project (Cooper 2001). Just as Ullrich and Eppinger, Cooper identified testing the *Product function* and testing the *Customer acceptance* as important test dimensions. However there is another important test dimension Cooper identified: *Validating the business case*. Figure 3 showed that a complete stage in the stage gate model is aimed at building a business case, the so called build business case stage. The assumptions which were made in this stage (e.g. price setting and potential market size) need to be validated in the testing and development phase. A product can function perfectly, as well as the organization around it, but when the customer is not prepared to buy it, the launch will be a failure (Cooper, Edgett and Kleinschmidt, 2002). In total, Cooper identified three test dimensions:

1. Product function
2. Customer acceptance
3. Validating the business case

Yang & El-Haik

Yang & El-Haik (2009) are two authors who are specialized in implementing six sigma and lean methodologies in NPD processes. In this context they also studied the testing and development process of new products and identified several dimensions for testing physical prototypes. The two dimensions they identified are:

1. Design analysis and Validation testing
2. Interface and compatibility testing

They further specified Design analysis and Validation testing into 6 sub categories :

Functional performance validation: This verifies whether the product can deliver all its functional requirements.
Operation environmental requirements validation: This verifies whether the product can deliver its function in diverse environmental conditions.
Reliability requirements validation: This verifies whether the product can perform its functions in an extended period of usage.
Usage requirements validation: This verifies whether the product can deliver its functions under various usage conditions.
Safety requirements validation: This verifies whether the product can meet the safety requirements.
Maintainability requirement validation. This verifies when maintenance is necessary and if it can be performed conveniently.

The above requirements and performance validation are all related to technical aspects and thus the functioning of the product. This corresponds with the view of Cooper and Ullrich and Eppinger. The second dimension identified by Yang & El-Haik, Interface and compatibility testing, also matches the view of Ullrich and Eppinger and stresses the importance to verify whether a product can work together with its sub-systems.

Dolan and Matthews (1993), Ozer (1999) and Carbonell-Foulquie et al (2004) carried out research studies in which they identified or used several testing dimensions for the testing and development stage. By comparing their findings with the view of the authors mentioned above a more profound comparison can be made.

Dolan and Matthews

Dolan and Matthews (1993) studied 21 beta-test programs of which they identified the purposes and test dimensions. As Dolan and Matthews view beta tests as 'one of a variety of procedures by which a firm has potential users "try out" a product and report on their experience', this corresponds with the criteria of the previous paragraph; "Psychical prototypes which resemble the commercial model are used" and "The testing is done outside the organization in a users' actual environment". The examined beta tests are therefore placed in the testing and validation stage.

Dolan and Matthews revealed that the beta tests they studied where used for three mayor purposes:

1. Product Function
2. Product support/marketing mix
3. Sales promotion

Project Function
The core of each test is to check if the product does what it is designed to do. While most companies do extensive testing prior to a beta test they recognize that that the demands put on the system by external users cannot always be anticipated. In addition to the basic function check, data can be obtained pertaining to desired refinements of the product and/or added features.

Product support/marketing mix

A commonly useful practice is to extend the test beyond the product itself to support elements. This can include training and documentation accompanying the product but also the communication and integration with other products and/or systems.

Beta tests are also used to determine the optimal positioning of a product for the actual launch. Important aspects to validate are target market selection, pricing and the economic value to the customer. Marketing policies for the introduction can also be tested.

Sales promotion

A successful beta test can lead to actual sales among the test group but also reduces the uncertainty of potential buyers about the product. This uncertainty reduction occurs as publicity about the successful test spreads.

The purposes Product Function and Product support/marketing correspond to a large extent to the Learning and Integration purposes of Ullrich and Eppinger (2000). The difference is that while the purposes of Ullrich and Eppinger are formulated in a general way, the ones identified by Dolan and Matthews are more specific. Therefore it is not necessary, unlike the purposes of Ullrich and Eppinger, to make them more profound and derive test dimensions. *Product Function, Product support* and M*arketing mix* are already dimensions on which a product can be tested. The last two will be treated as single test dimensions as they cover different areas of interest.

The Sales promotion purpose of Dolan and Matthews is linked to the Communication purpose of Ullrich and Eppinger (2000). As explained in paragraph 2.1, the testing and validation stage as a whole is viewed as a communication instrument in this purpose. Therefore sales promotion is not a dimension which can be tested.

Ozer

The goal of the study Ozer carried out in 1999 was to review the widely cited new product evaluation models since, according to him, many have been underutilized. Several of the explored evaluation models meet the criteria for the testing and development stage established in paragraph 3.1 (prototype testing, pre-market test and test-market). In his attempt to provide a survey of these models Ozer also addresses the test domains of the models.

Along with the other authors Ozer (1999) acknowledges the importance of testing:

1. Product performance
2. Consumer preference.

The third test domain Ozer identifies is :

3. Marketing variables.

Organizations in his study use a test population to validate their sales forecast and marketing plans. This corresponds with the Business case validation dimension identified by Cooper (2001).

Carbonell-Foulquie et al

Carbonell-Foulquie, Munuera-Aleman and Rodriquez-Escudero (2004) examined the relative importance of a set of go/no go criteria at gates in the NPD process by studying the development process of 77 highly innovative products. Their findings reveal that the go/no go criteria for products can be grouped into five dimensions (test dimensions on which a new product is scored):

1. Strategic fit
2. Technical feasibility
3. Customer acceptance
4. Market opportunity
5. Financial performance

The next step they took was to verify which dimension is important in which part of the NPD process. They discovered that the strategic fit test dimension is important in the beginning of the new product development process as it is used to approve a new product concept. Since this consideration is a part of the 'scoping'- and 'build business case' stage of the NPD process it is not appropriate anymore in the testing and development stage. The other four identified dimensions, however, are of importance during the whole NPD process or near the end. This makes them relevant for the testing and development stage.

Combined authors

When the identified test dimensions of the different studies are compared it becomes clear that, while different authors use different terms for test dimensions, four overall test dimensions emerge.

Overall test dimensions	Ullrich and Eppinger	Cooper	Yang & El-Haik	Dolan & Matthews	Ozer	Carbonell-Fouliquie et al.
Product Function	Product function	Product Function	Design analysis and validation testing	Product Function	Product performance	Technical feasibility
Customer acceptance	Meeting customer needs	Customer acceptance			Consumer preferences	Customer acceptance
Business Case and Plan		Business Case validation		Marketing mix	-Sales forecast -Marketing variables	-Financial performance -Market opportunity
Integration & Support	Integration		Interface and compatibility testing	Product support		

Table 1: Overview test dimensions

In table 1 the identified dimensions of the different authors are listed and grouped together based on their content. All of the four identified overall test dimensions are supported by several studies.

Product function

All authors and studies agree on the fact that testing the product function is an important dimension in the testing and development stage. Demands put on the system by external users cannot be always anticipated, which is why technical requirements need to be validated. In general, this test dimensions verifies whether the product does what it is designed to do.

Customer acceptance

The first of the established criteria for tests in the testing and development stage (The testing is done outside the organization in a users' actual environment) already stresses the importance of customer input.

Before a product is launched into the market the customer acceptance should be validated by measuring the usability and the extent to which users' needs are met. Yang & El-Haik and Dolan & Matthews view this dimension as a part of the product function dimension but this does not reflect its importance. In both the studies of Ozer and Carbonell-Fouliquie et al, customer acceptance emerged as a vital dimension on which organizations focus, what does the customer think of the usability and of the product in general? This is further supported by the view of Ullrich and Eppinger and Cooper and also adopted in table 1.

Business case and plan

This is the last test dimension for the testing and validation stage. Validating the targeted market selection, pricing, economic value to the customer and marketing policies for the introduction are essential to move to the launch stage. In the business world a product can be great but when it does not add value (profit usual), it will not be produced. Therefore it is not strange that organizations in the studies stress the importance of this dimension.

An explanation for the fact that Ullrich and Eppinger and Yang & El-Haik neglect to identify this dimension could maybe be found in their focus on the product development cycle. Business case and plan aspects can be viewed as constrains for this process rather than directly influencing factors, this might be the reason that Ullrich and Eppinger overlooked them.

Integration & support

This dimension is a somewhat complicated as it goes further then the actual product. Since the importance of this dimension is also product specific, it depends on the scope of a researcher or study whether they identified it. However, the aim of this dimension can be a crucial aspect for many products. Validating the success of integrating a product with other products, services and work systems/methods and test the organizational process to effectively manage the process, can save a lot of time in the launch stage. Future problems can already be identified and work processes can be adjusted in advance.

The four identified testing domains for the testing and validation stage are visualized in figure 6. The complete circle symbolizes a test within the testing and validation stage and the four identified test dimensions each form a part of it.

Since the testing and validation stage is fulfilled in a different way by diverse organizations and comprises a variety of testing types, the importance of each of the four testing dimensions is not standard. Depending on the purposes and characteristics of a specific test, the importance of each of the different testing dimensions should be agreed upon.

While the importance of the four testing domains can differ between single tests, all four testing dimensions should be addressed in the testing and validation stage as a whole.

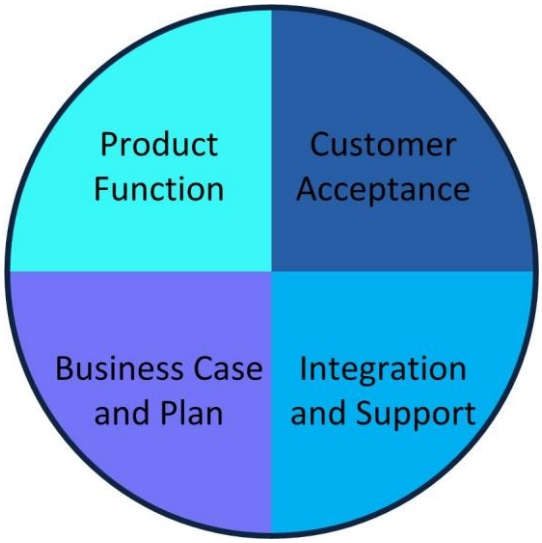


Figure 6: The four identified test dimensions

2.3 Guidelines for designing a test in the testing and validation stage

Dolan and Matthews (1993) discovered in their analysis of 21 test programs that the design of a test is crucial for the usefulness of the outcomes and a careful process is therefore necessary.
In this paragraph guidelines for designing tests within the testing and validation stage are identified, in order to help organizations to optimize their testing strategies.
The first guideline followed from the findings of the previous paragraph, the deriving of four testing dimensions:

1. *Effective test programs recognize the full set of benefits of all four test dimensions.*

Dolan and Matthews (1993) argue that researchers need to be aware of the importance of all test domains and incorporate them in the design. This way the usefulness of a test is maximized.

The second guideline is on the selection and approach of participants in tests. Cole (2002) states that a good selection can lead to rapid learning curves while the selection of "wrong" customers can result in false inferences and erroneous decisions. Furthermore it is important to prevent developers from influencing the test population with their enthusiasm since they are naturally impassioned by their product (Crawford & Benedetto, 2000). Useful results are objective results and striving for a close approximation to a real world experience is therefore important:

2. *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience.*

In order to limit the risk of negative exposure when problems occur, Dolan and Matthews (1993) argue that it is smart to start a test with customers who have good relations to the organization.

Additionally they argue that when an organization grows the number of testers over time, it becomes possible to pre-solve problems identified at the first users for the rest of the participants. Bonner (1997) even goes further and recommends to look for customers who will have a need for a product if it proves to be successful. According to her it gives the customer incentive and provides a better test. These findings led to the third guideline:

3. *Grow the number of testers over time and begin with sophisticated customers with good relations to the firm*

The initial design of the test is important but not enough for a successful test. Monitoring and evaluating the test in a proper way is just as essential. According to Dolan and Matthews (1993) it is not sufficient to only monitor a product in the beginning and the end of a test period. Keeping up with changes in all test domains provides a much richer data set and enables a deeper analysis. When the test is also evaluated on a regular basis, flaws can be identified and solved in such a way that the remaining test period will provide the needed results.
Crawford & Benedetto (2006) furthermore discovered that many organizations have not enough internal capacity to test the performance of a product at the required levels and lack the funding to hire an outside firm to do the test. This results in developers who are testing their own products with limited means. Since the developers are very close to the product they are often not suitable to critically test it and to find problems, which reduces the usefulness of the outcomes. Additionally developers sometimes ignore early negative results, hoping that the product will improve by itself during the testing process.
To tackle these pitfalls the fourth guideline is:

4. *Monitor and evaluate the product performance on a regular basis in an objective way.*

Another factor influencing the usefulness of tests within the testing and development stage is whether the design process of the product stops or runs parallel with the test (Dolan and Matthews, 1993).
The aim of a test is to learn from customer experience and feedback, identify technical problems, validate the business case/plan and test the integration and support of a product in order to adjust or improve the product.
When the design process of a product continues during the test and the product is altered parallel with the test, the findings do not apply anymore. To prevent that the results of a test become outdated, the fifth guideline was established as follows:

5. *Product design should be frozen prior to a test*

A remark on designing product tests is provided by Vellandi (2008). He points at the expensive nature of product tests and argues that organizations need to be sure of the benefits a test will deliver before starting. Especially large scale product tests should be considered optional according to Vellandi. Its costs versus its benefits depend on the uniqueness of the product category, operational flexibility and channel strategy. This remark forms the final and sixth guideline:

6. *Be certain of the benefits a test will deliver before starting*

2.4 Theoretical model for the testing and validation stage

By integrating the findings of the previous paragraphs it became possible to establish a theoretical model for the testing and validation stage; figure 7.

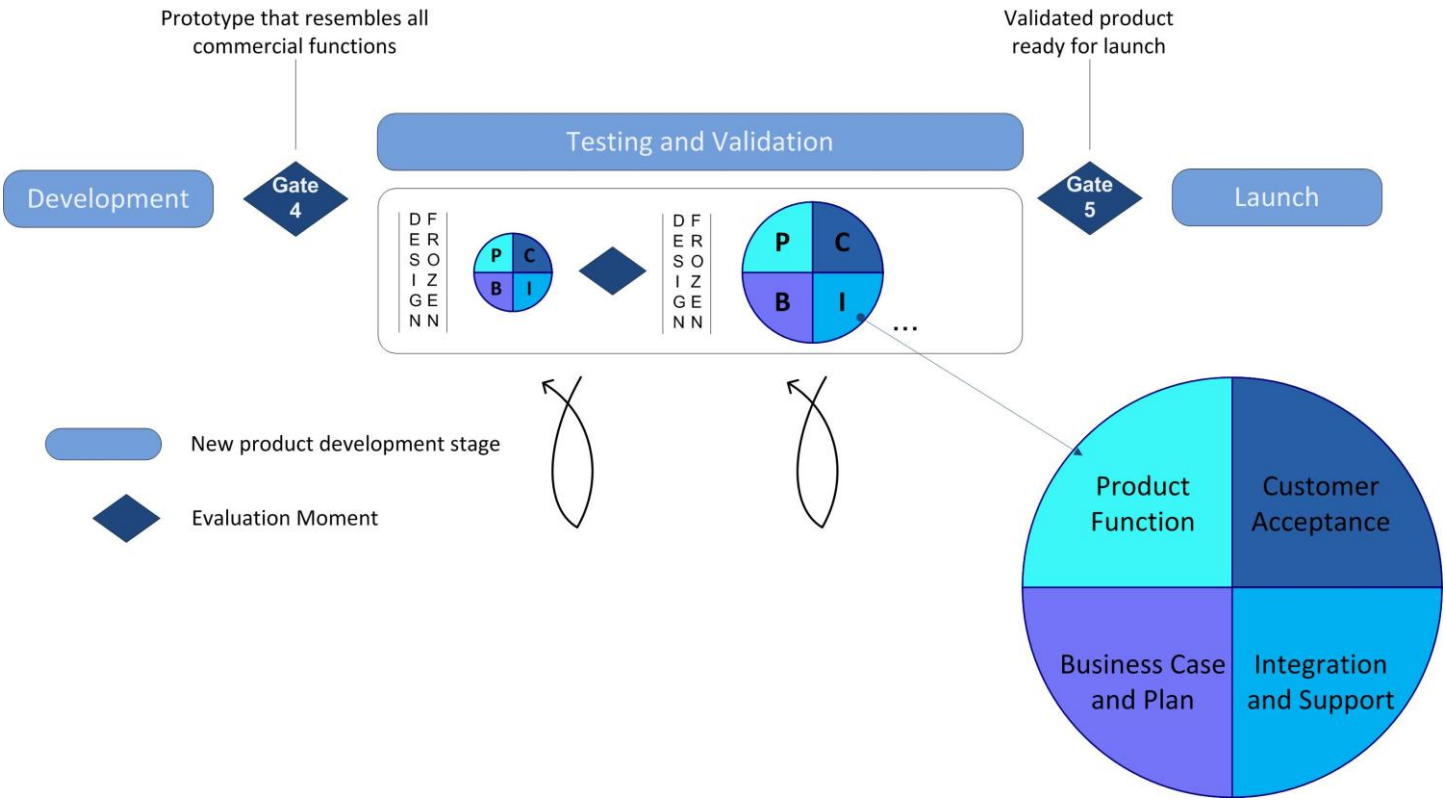


Figure 7: Theoretical testing and validation model

While there were eight important aspects identified in the previous paragraphs it was not possible to visualize and embed each one of them completely. The ones who could not be visualized complement the model.

The first two aspects of this model are the criteria for tests within the testing and validation stage identified in paragraph 2.1:

1. *Physical prototypes which resemble the commercial model are used*
2. *The testing is done outside the organization in a users' actual environment*

In figure 7 can be seen that the first criterion was used to mark the beginning of the testing and validation stage. Before a product should be allowed to pass the gate between the development stage and the testing and validation stage (the blue diamond) it should technically be completed and resemble a commercial model. The second criterion was hard to visualize and therefore accompanies the model.
In paragraph 2.2, four test dimensions were derived which should all be addressed in the testing and validation stage to achieve maximum utility. These test dimensions were reflected in figure 6 as parts of a test within the testing and validation stage. The same image was used in figure 7 to picture the

tests this stage comprises. From a theoretical point of view the testing and validation stage has no defined amount of tests it comprises (Minot and Wood 2003), this depends on the product and organization characteristics. This is further supported by the fact that the stage should be viewed as an iterative process, (Cooper 2005) which makes the determination of a vast amount of tests incorporated in this stage even harder. The iterative nature is visualized in figure 7 by the two iteration loops. The lack of a specific amount of tests is reflected by the line of dots in figure 7.
The importance of addressing all test dimensions was further underlined by the first guideline identified in paragraph 2.3, this is the third aspect of this theoretical model:

3. *Effective test programs recognize the full set of benefits of all four test dimensions.*

The other aspects of the theoretical model are formed by the remaining five guidelines:

4. *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience.*

This guideline could not be visualized and therefore complements the model.

5. *Grow the number of testers over time and begin with sophisticated customers with good relations to the firm*

This guideline was visualized to some extent in figure 7 by making the second test within the testing and validation stage bigger than the first one. This symbolizes an increase in the number of participants over different tests. However also within a test the number of participants should be increased gradually.

6. *Monitor and evaluate the product performance on a regular basis in an objective way.*

While this guideline is meant to be applied in tests within the testing and validation stage, proper evaluation between stages or tests is also relevant. The blue evaluation diamond between the two tests in figure 7 symbolizes that after every test a proper evaluation should be held.

7. *Product design should be frozen prior to a test*

This guideline was visualized completely in figure 7 by the |Design Frozen| sign before each test within the testing and validation stage.

8. *Be certain of the benefits a test will deliver before starting*

This guideline could not be visualized and therefore complements the model.

3. Methodology

3.1 Research Strategy

In this chapter an outline is given on how the study of the current testing and development practice and design requirements of Comp. X NDB was carried out.
When the research model of paragraph 1.3 is recalled (figure 4) it becomes clear that, as the theoretical analysis was carried out in the previous chapter, this is the next step in realizing the purpose of this research study "To develop and structure an execution model for the testing and development stage, suitable for Comp. X NDB"

This step was necessary as testing and validation stages can differ to a great extent among organizations. Depending on the characteristics of a product and company work methods, organizations can use one or more testing practices. From a theoretical point of view the testing and validation stage thus comprises a variety of testing types and has no unconditional form. This was also reflected by the outcome of the theoretical framework: a relative general theoretical model.
By analyzing the work methods of Comp. X, verifying to what extent the identified key dimensions and guidelines in theory are already applied and identifying the design requirements of Comp. X it became possible to adjust and specify the theoretical model.

While the theoretical model could be developed by desk research alone, the study of the current testing and development practice and the design requirements of Comp. X could not. There was no sufficient documentation, and intended practices often differ from actual practices. Empirical data was needed in order to provide a good understanding on these topics. In collecting this data a choice had to be made between a quantitative or qualitative data collection method.
Quantitative data is usually presented as numerical data and can be collected through questionnaires, surveys or large scale observations. The advantage of this method is that the analysis can be done easily and fast and it is suitable for a large group of respondents. The disadvantage is that you are limited to closed types of questions or need to code open questions which is a very time consuming process. This limits the possibility to create a deeper understanding of the issues in a reasonable amount of time. This deeper understanding can be obtained by using qualitative techniques. They do not involve measurement or statistics, the most common form of qualitative techniques is interviews. As the questions asked in these interviews are open, respondents have the possibility to give an argumentation for their opinions and elaborate on a topic (Gramsbergen & van der Molen, 1992).

For studying the current testing and development practice of Comp. X a qualitative method was preferable for two reasons:

- A testing and development practice is a complex and company specific process and respondents need to have room for elaborating on- and explaining their view in detail. Additionally the aim of the interviews is an exploration of the current practice which stresses the need for a qualitative approach.
- As the outcomes of the data collection were used to adjust the theoretical testing and validation model specifically for Comp. X, the potential group of respondents consists of the

people working in the Comp. X department. This is only a small amount so an in depth method is preferred over a method suitable for a large group of respondents.

Semi-structured interviews

Semi-structured interviews were selected as the most suited approach to provide insight in the intended practice of the management, actual used practices by the employees and the design requirements. Totally structured interviews would not be convenient since both the interviewer and respondent need to have the possibility to bring up new questions or elaborate on a certain question. An unstructured interview however, was also not appropriate as the scope of the interview is limited to a specific goal, insight in the current practice and design requirements. Therefore the respondents should not be allowed to move the conversation in any direction of interest they come up with (Gramsbergen & van der Molen, 1992).

This led to the choice for semi-structured interviews as this technique finds itself between a structured and unstructured interview and combines both strengths.

Observational research and document analysis

In section 1.1.2 was explained that an important motive of Comp. X for this research was the need for a research design for the West Orange project. This project comprises 4 products (Product H, Product J, Product K and Product L) which are going to be tested on a large scale.

Smaller tests (20 consumers) of the Product H and Product J were in process when this research study was started at Comp. X. Because of the relevance for this study, the evaluation sessions of both products were used as observation research to observe the applied research strategies and testing dimensions. Furthermore the files in which both tests were documented and evaluated, both tests were finished during this research study, were also studied. Through the combination of observational research and document analysis it was possible to develop a comprehensive view on both tests which makes them a valuable information source.

However, the Product H and Product J are not the only tests that were discussed during the interviews. Additionally a small and large scale test of Product L and a test with a Product G were discussed. Documentation on these tests (e.g. surveys that were used, evaluation reports, etc) was also studied.

By combining these three research techniques; semi-structured interviews, observational research and document analysis the results became more substantial.

3.2 Data collection

To complete the goal of this research, the development of a testing and development model suitable for Comp. X, the outcomes of the interviews had to be used to adjust the theoretical testing and validation model specifically for Comp. X. Therefore the people working in the Comp. X department formed the potential group of respondents for the semi structured interviews.

There is one department manager and two team managers who are each responsible for the work of five employees. Managers as well as normal employees were important to include in the sample as the practice as intended by the management can differ from the practice actually applied by the employees. Furthermore also their view on the design requirements of Comp. X for a testing and valuation model can differ.

While all three managers have at least some experience with product test, not all employees have. At Comp. X new services as well as new products are developed and some employees have never participated in an actual product test. While employees with no test experience will not be able to provide input on the actual testing practice, their view on design goals and requirements of Comp. X could be valuable since they did participate in evaluation sessions. As the model that is developed will be the new "way of working" for Comp. X, and affects all people working in the department, they were all selected to be interviewed except one. One of the employees is the subsidy specialist of Comp. X and is only involved in specific part of a project, the subsidy track. As he has no affection with developing products or services and the outcome of this research study will not affect his work, he was excluded from the sample. Table 2 shows the selected and interviewed respondents compared to the departments total. One employee who was selected to be interviewed accepted a new job before the interview took place and as he went on holiday before the interviews started it was not possible to include him in the sample. Therefore 8 employees instead of nine were actually interviewed.

Category	Comp. X Total	Selected respondents	Interviewed respondents	Percentage
Managers	3	3	3	100%
Employees	10	9	8	89%
Total	13	12	11	93%

Table 2: Interview Respondents

To make sure that the answers on the interview questions and the view of the respondents in general was interpreted in the right way, each interview was summarized. These summaries were sent to the respondents to provide them with the opportunity to correct or specify them.

3.3 Interview protocol and points of interest for the observational research and document analysis

Current testing practice of Comp. X

The first purpose of the empirical research was to explore the work methods of Comp. X in previous tests and verify to what extent the identified key dimensions and guidelines in theory were already applied.

While in theory there are very clear criteria for the beginning and end of the testing and validation stage, Comp. X does not necessarily view this the same way. The testing and validation of new products as a whole was therefore the topic of the interviews, to prevent that vital information was overlooked due to differences in point of view. After the whole testing process was explored a differentiation of the actual practices could be made between the NPD stages.

The topics that were addressed in the interviews, and also formed the points of interest for the observational research and document analysis, were derived from the theoretical framework.

Topics that were addressed:

1. The characteristics and purposes of the applied testing types
2. The learning goals and testing dimensions of the tests
3. The selection of participants in the tests
4. The composition of testing groups
5. The size of testing groups and the roll out process
6. The people in charge of the setting up, running and evaluation of tests
7. The evaluation of tests and the variables on which tests are scored
8. The way evaluation and feedback is incorporated in the design process
9. The existence of Go/No go gates and the corresponding criteria
10. The time schedule of the design process in relation to the consumer tests
11. The existence of a definite testing structure and the existence of criteria on which a decision is made to initiate a test.

Topics 1 and 8 focused on the first step in the theoretical framework, the identification of the characteristics of the testing and development stage in theory.
Two criteria were established for test within the testing and development stage:

- Physical prototypes which resemble the commercial model are used
- The testing is done outside the organization in a users' actual environment

These criteria were verified with topic 1 and the goal of the testing and development stage according to theory, "To further develop the prototype into a market ready product, trough an iterative process", was explored with topic 8.

The second step in the theoretical framework was the identification of test dimensions (table 1). This step was again be dealt with by topic 1 together with topic 2 and 7. By addressing the characteristics and goals of tests, the intended test dimensions and the criteria on which tests are scored the actual used test dimensions could be identified. This way the first Guideline "*Effective test programs recognize the full set of benefits of all four dimensions"* was also addressed

Topic 3, 4 and 5 were on the selection, composition and size of test group and hereby addressed the second and third guideline: "*Grow the number of testers over time and begin with sophisticated customers with good relations to the firm"* and "M*aximum utility of a test is derived from recognizing the segmentation of the market and devising ways of making tests a close approximation to real world experience"*

The fourth guideline was on the evaluation and monitoring of a test, "*Monitor and evaluate the product performance on a regular basis in an objective way".* The combined findings of topic 6, 7, 8 and 9 provided a good view on how this process was carried out at Comp. X.

As topic 10 identifies whether consumer tests run parallel or successively the guideline *"Product design should be frozen prior to a test"* was addressed. Topic 11 finally focused on the validation of the last guideline: *"Be certain of the benefits a test will deliver before starting"*

Identification of requirements and design goals of Comp. X

The second purpose of the empirical research was the identification of design goals and requirements for the model that is developed. Topic 1-11 already revealed some limitations within Comp. X and guidelines for the execution of tests. By projecting these topics to the future, and ask respondents what they would change or do differently when they were asked to set up the test again, these aspects became more clear. Additional topics that were addressed in order to identify requirements and design goals of Comp. X were:

12. The relevance of a testing methodology in general
    -The way it should be incorporated in the NPD funnel

13. Their view on the alpha, beta, gamma methodology (1, 20, > 100 products)
    -What should be the goals of the different tests?
    -Are all three tests necessary for every new product?

14. Boundary conditions for tests within Comp. X

Topic 13 referred to the division Comp. X made in the testing and validation stage, three tests ascending in size. While the exact characteristics of these phases was not yet agreed upon, the division already proved to be of great value in the communication to the rest of the organization. Therefore it was treated as a design restriction from Comp. X.
The established topics for the interview protocol were further processed into specific open questions in Dutch. A list of these questions can be found in the Appendixes.

3.4 Data analyses

While the previous paragraph explained how the current testing practice of Comp. X was explored in the interviews, the question how these outcomes were interpreted was not discussed. That is the topic of this paragraph.

The subjects that were addressed in the interviews were derived from the outcomes of the theoretical framework. The eight aspects which formed and complement the theoretical testing and validation model were used as criteria on which previous tests Comp. X executed were scored. The aim of this scoring was not to criticize or judge the previous tests but to provide insight in how Comp. X is currently working, as input for the new model developed in this research study. Therefore it was not necessary to score each criterion on a very specific level and the choice was made to use a three point colour scale. Each criterion was scored as follows:

Score Colour	Meaning
	Yes, Comp. X applied or followed the criteria
	Comp. X applied or followed the criteria to some extent
	No, Comp. X didn't apply or follow the criteria

The first two criteria of analyses were the criteria which were established for test within the testing and development stage:

1. Physical prototypes which resemble the commercial model are used
2. The testing is done outside the organization in a users' actual environment

For criterion 1,a green score was given when the product did resemble all commercial functions and a red score was given as the product used in the test did not resemble a commercial model at all. An orange score was given when the product was meant to resemble the commercial model but during the test it turned out that fundamental changes were necessary.

For criterion 2 only a red or a green score was given. Green when the product used in the test, was indeed tested outside the organization in a users' actual environment. A red score was given when the product was tested internally.

The 3th criterion on which the previous executed tests were scored is a mix between a guideline and the four identified test dimensions:

3. Effective test programs recognize the full set of benefits of all four test dimensions.

This theoretical guideline stresses the importance of testing on aspects of all four test dimensions. Previous executed tests were therefore scored on each test dimension individually (Product function, Customer acceptance, Business case and plan and Integration and support)
A green score indicates that in that test all relevant aspects of a test dimension were addressed while a red score was given when a test dimension was not addressed at all. An orange score was given when the importance of a test dimension was recognized, but not every relevant aspect measured.
Criteria 4,5,6,7 and 8 the previous executed tests were scored on, are the outstanding guidelines from the theoretical model:

4. Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience.

When both aspects from this guideline are followed (participants are actual customers who resemble the market and the test tried to become a real world experience) a test was given a green score. When only one of them was followed, or both to some extent, the score was orange. When the guideline was completely neglected (e.g. instead of customers own employees are used) the score was red.

5. Grow the number of testers over time and begin with sophisticated customers with good relations to the firm

Similar to criterion 2 there were only two scorings options on this criterion. Green when they did grow the number of testers overtime and began with "friendly" users in a previous test and red if they did not.

6. Monitor and evaluate the product performance on a regular basis in an objective way.

This criterion includes two aspects which are relevant for a proper evaluation of all aspects; regular and objective. When both aspects were present in the evaluation of a test a green score was given. When only one of them was applied the score was orange and when both aspects are not in place the score given was red.

7. Product design should be frozen prior to a test

This criterion will be scored in a similar way as criterion 1. A green score was given when the product development process was frozen completely and a red score was given as the product development process continued parallel with the test. An orange score was given when the product development process was meant to be frozen but was unfrozen to solve redundant technical problems which came to light in the test.

8. Be certain of the benefits a test will deliver before starting

A green score was given for this criterion when the benefits of a test for Comp. X were clear before starting, and when they could also be met within the test design. A combination of clear benefits and the possibility to realize them, form the actual benefits of a test. When both aspects applied on a test a green score was given. When the benefits in a test were unclear or when they could not be met within the scope of the test, a red score was given. When the goals of a test were clear and they could also be met, but the role Comp. X played in the test was not adjusted accordingly the score was orange. In other words, when Comp. X only had some minor gains from a test but put in lots of effort or resources, there was a mismatch between benefits and effort. If this was the case a test was scored orange.

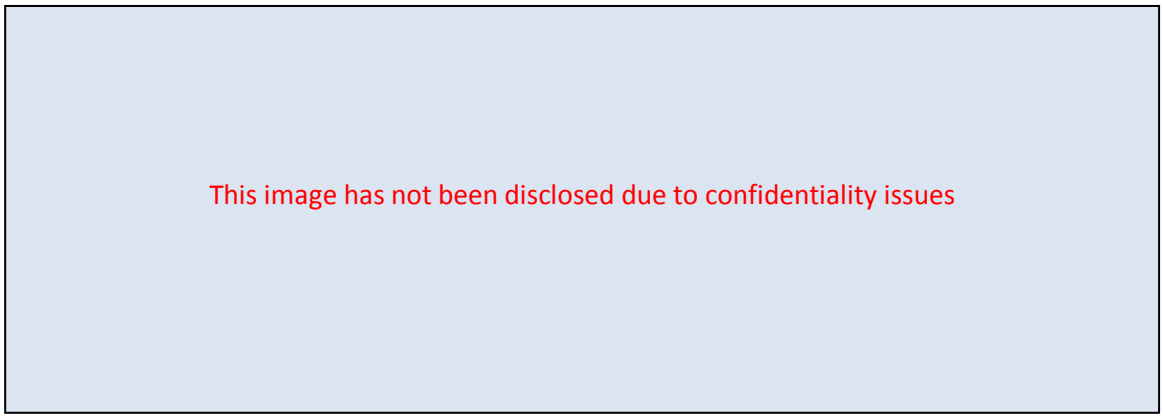
3.5 Application of the developed model on the West Orange project

The analyses of the current testing and development practice of Comp. X in chapter 4 is used to adjust and specify the theoretical model in chapter 5. This way a model for the testing and validation stage of Comp. X is established which suits their work methods and design requirements. While this is an important step towards answering the main research question, it is not the final step in this research study. The developed testing and validation model is applied on the West Orange project in order to develop a research design for this test (recall figure 4). The experiences with the actual application of the model are used to determine the value and relevance of the developed model in practice.

An important aspect in the application of the model was the identification of the learning goals in the project. There are many organizations involved in the project and it is important that the learning goals of each one of them are embedded in the final research design. In order to identify these learning goals, semi structured interviews were held with the representatives of all participating organizations.

Semi structured interviews were selected as the most appropriate data collection method for this identification because respondents needed to have the freedom to bring up new topics or elaborate on a certain question, but as it had to be prevented that topics were overlooked a general structure for the interviews was necessary.

In total twelve people were interviewed from eight different organizations. The respondents from Comp. X were the project initiators from NBD and two managers from Marketing and Product and Services Management (PMS). Because the project has to land in the organization when it becomes a success it was important to already involve persons from relevant departments.
The respondents of the other organizations were the people responsible for the project within their organization. This ranged from project manager to director or even CEO. An exception is the professor of the University who is not really responsible for West Orange within his organization, but is interested in the research itself and helps to make sure that all learning goals will be met.



The interview was structured with the help of the four test dimensions on which a testing and validation program should focus, identified in the theoretical framework (figure 6). Since all organizations have a different point of view in the project, this helped to make sure that all relevant aspects were addressed.

Topics that were addressed in the semi-structured interviews:

- *Learning goals on the Product Function dimension*
- *Learning goals on the Customer Acceptance dimension*
- *Learning goals on the Business Case and plan dimension*
- *Learning goals on the Integration and Support dimension*
- *Other learning goals not belonging to one of the above dimensions*

The respondents were informed of the structure of the interview beforehand so that they could prepare themselves and discuss the learning goals of their organization with their colleagues. To clarify each of the test dimensions some examples of learning goals Comp. X already formulated were also provided. To make sure that no learning goal would be overlooked there was also asked if an organization had other learning goals, not belonging to one of the identified dimensions.

The list of learning goals that emerged from these interviews was matched to appropriate data collection methods with the help of data marketers from Comp. X and the University

Since the application of the developed model mainly consisted of following the prescribed guidelines it was also necessary to identify the limitations in the project. Limitations like available participants/houses, available data collection methods and roll out restrictions can have serious consequences on the possibility to follow a guideline.
This information was not gathered in a structured way but more on the run through meetings and desk research. Based on the developed model an ideal research design was build which was adjusted by the project management office to fit to the actual test possibilities.

4. Practice at Company X

The theoretical testing and validation model, developed in chapter 2, is a good start to structure the testing and validation process at Comp. X. However, it is still rather general and it therefore needs to be adjusted and specified in order to be useful and applicable for Comp. X.
This chapter explores to what extent the identified aspects from the theoretical model are already applied by Comp. X and which design requirements Comp. X has of its own.
In the next chapter the findings on these topics are combined and analyzed together, to determine what the implications of the practice at Comp. X are for the model that is developed.
Before the practice at Comp. X is described in detail their general view on the testing and validation process is clarified.

In paragraph 1.1.1 was explained that Comp. X is currently implementing a stage gate model, in order to structure the development of new products and technologies (Figure 1). In this stage gate model Comp. X subdivided the testing and valuation stage into three new phases, an alpha, beta and gamma phase. This division is mainly based on an increase in the number of participants or prototypes in tests. In an alpha test 1 prototype is used, in a Beta test approximately 20 and a Gamma test comprises at least 100 prototypes.
The general idea of Comp. X behind this division is that the alpha phase serves as a proof of concept. The main question during this phase is: will this product work? The beta phase is used to focus on the customer. What is the initial customer interest? How does a customer use the product and what is the products durability in a customer setting.
The gamma phase is the last testing phase and is aimed to validate the organizational capacity to effectively install, manage and support the new product on a larger scale. It can be viewed as a preparation for an actual market introduction. In figure 8 this process is visualised. The three phases are succeeding and together form the complete testing and validation process.

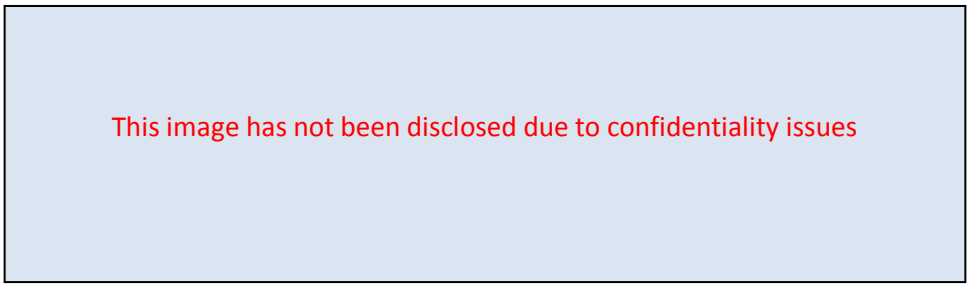


Figure 8: Current Testing and Validation process Company X

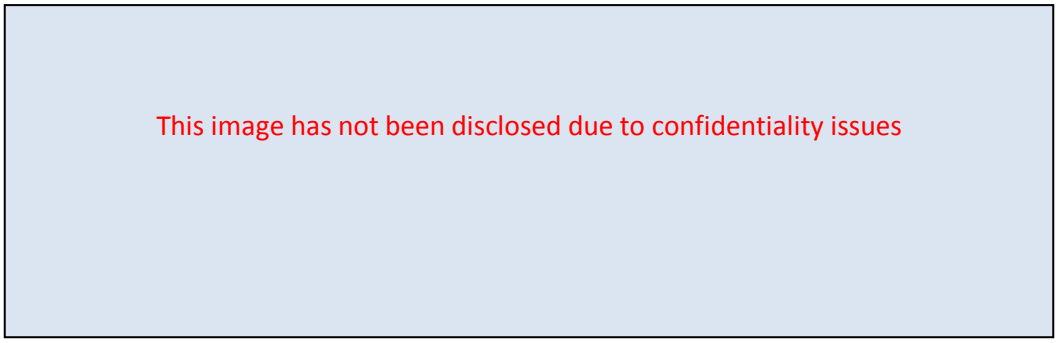
As this division in sub phases is still rather general, specific goals and characteristics are not yet established. However, all respondents agree on the fact that it is valuable to have a standard testing methodology.
In the opinion of several employees, previous executed tests were often designed too fast and uncoordinated, which sometimes resulted in poor results. They feel that a standard testing methodology, and a more project based way of working, can improve this. Especially since the products that are tested by Comp. X have many similarities.
Furthermore all respondents agree that a standard testing methodology is very valuable for communication purposes. Comp. X has been using the terms alpha, beta, gamma for almost a year

now as a way to explain to other departments/organizations were they stand in the testing process and what they are doing. This led to more openness in the projects and understanding from other departments.
The fact that the terms alpha, beta and gamma are already embedded in the organization led to the most prominent design requirement of Comp. X: the preservation of this subdivision in the model that is developed in this research study. That is, only the general division needs to be preserved. The fulfillment and characteristics of these phases are open for discussion.
Before other design requirements are identified, the current work methods of Comp. X and the extent to which the identified key dimensions and guidelines in theory are already applied are described.

4.1 The current work methods of Company X

The current work methods of Comp. X are described by analyzing several tests which were executed the last two years. This was done with the aspects of the theoretical model as criteria. Each test was scored on a three point scale (red, orange and green). Specifics on how the scoring was performed for each criterion were provided in paragraph 3.4 and is summarized before each test criteria.

During the interviews it turned out that Comp. X executed five product tests over the last two years. Comp. X classified them as one alpha, three betas and one gamma test. Table 4 provides an overview of these tests together the abbreviations which are used from now on.



The scores of the previous executed product tests are discussed separately for each indentified criteria. Each section on a criterion ends with a table in which the scores are reflected. When all criteria are addressed and scored, these tables are combined into an overview of all scores. At that point, problem areas are identified which form the main point of analysis in chapter 5.

Test Criteria:

1. *Physical prototypes which resemble the commercial model are used*

Green score:	The product did resemble all commercial functions
Orange score:	The product was meant to resemble the commercial model but fundamental changes were necessary
Red Score:	The product did not resemble a commercial model at all.

This section has not been disclosed due to confidentiality issues

2. *The testing is done outside the organization in a users' actual environment*

Green score:	The testing was done in a users' actual environment
Red Score:	The testing was done internally

This section has not been disclosed due to confidentiality issues

3. *Effective test programs recognize the full set of benefits of all four test dimensions.*

The guideline underlying this criteria aims to maximize the usefulness of a test by emphasizing the need for organizations to address the four identified test dimensions in paragraph 2.2 However, only recognizing the importance of a test dimension is not enough. A differentiation was made between the identified test dimensions at the beginning of a test and the test dimensions on which a product was actually evaluated. The last one was the basis for the scores. Apart from the interviews especially the evaluation documents on the previous tests proved very useful to score them, as these documents described the findings on each of the tested aspects.

Green score:	All relevant aspects of a test dimension where addressed
Orange score:	The importance of a test dimension was recognized, but not every relevant aspect was measured
Red Score:	The test dimension was not addressed at all

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

4. *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience.*

Green score:	Participants were actual customers who resemble the market segmentation and the test tried to become a real world experience
Orange score:	Only one of the two aspects of the criteria was followed or both to some extent
Red Score:	Participants did not match the segmentation of the market and the test was not a real world experience

This section has not been disclosed due to confidentiality issues

5. *Grow the number of testers over time and begin with sophisticated customers with good relations to the firm*

Green score:	The number of participants was grown in the test and the test started at “friendly” users
Red Score:	The number of participants was not grown in the test and the test did not started at “friendly” users

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

7. Product design should be frozen prior to a test

Green score:	The product development process was frozen completely
Orange score:	The product development process was meant to be frozen but was unfrozen to solve redundant technical problems
Red score:	The product development process continued parallel with the test.

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

8. Be certain of the benefits a test will deliver before starting

Green score:	The benefits were clear before starting and could be met within the test design
Orange score:	The benefits were clear before starting and could be met within the test design but the role of Comp. X played in the test was not adjusted accordingly
Red score:	The benefits of a test were unclear before starting or could not be met within the test design

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

Combination of the scores per test on all criteria

Table 5i provides an overview of the scores per test on all criteria. This overview was used to identify problem areas.

	Alpha	Beta			Gamma
	Test G	Test H	Test J	Test K	Test L
Product used is a full working prototype				2	
Testing is done in a users' actual environment					
Tested on Product Function dimension					*
Tested on Customer acceptance dim.					*
Tested on Business case and Plan dim.	1	3	3	3	*
Tested on Integration and Support dim.	1	3	3	3	*
Test design and participant selection match the real world					
Number of testers is grown over time, starting with friendly participants					
Evaluation is done regularly and objective				4	*
Product design is frozen during the test				5	
Certain of benefits before starting					

=Problem Area

Table 5i: Combination of the scores per test on all criteria

**Estimated scores based on test outlook*

The first ting which was noted when looking at these scores in table 5i was that Test G scored eight times red on a total of eleven scoring criteria. While this could mean that this was simply a very bad test, this is not likely. Nuon was certain of the benefits of the Test G before starting and they could also be met within the test design (green score on the last criterion). It is therefore more likely that there is another reason for the bad scores of the Test G, possible related to the view Comp X has on alpha tests in general. The Test G as a whole is therefore marked as problem area 1 and is analyzed in chapter 5.

With all scores of the Test G marked as a specific problem area, there are four scores (Test H, Test J, Test K and Test L) on each criterion left. These scores were used to identify the other problem areas. The other problem areas were defined as a "criterion on which the majority of scores was orange or red". In these situations there can be concluded that Comp. X underestimates the importance of the criterion or that it is hard to apply it in practice. In both cases this can have implications for the model that is developed and these problem areas are therefore analyzed in chapter 5.
By applying the above definition, four additional problem areas were identified:

Problem area 2:	The products used are often not full working prototypes
Problem area 3:	The test dimensions Business case and plan and Integration and Support are poorly addressed
Problem area 4:	Evaluation is not performed regularly and objective
Problem area 5:	The product design is not frozen during the tests

4.2 Design requirements of Company X

By discussing the previous test with the respondents and talking about the things they would change or do differently in future tests, several points that could (or need) to be improved were identified. These points were adjusted into design requirements of Comp. X for the testing and validation model that is build in this research study. One of these design requirements was already mentioned in the beginning of this chapter: The preservation of the alpha, beta and gamma subdivision in general. In this paragraph the other design requirements that were identified will be addressed, but first a minor remark will be made to the first one.

1. *Alpha, Beta and Gamma division should be preserved*

While all respondents agreed on the fact that a standard test methodology is very useful, not everyone agreed on the necessity to execute a gamma test.
Respondents in support of this large scale test (>100) argued that many of the products that are being developed are so new that a good integration and support of these products by the organization will require a change in the current working methods of Comp. X. Apart from influence on the current logistic and billing systems, some products will also increase the knowledge of customers about their energy consumption. This can affect other departments of Comp. X, for example the call centre can be confronted with other and more difficult (specific) questions. In a gamma test the organization can try whether this new role fits them and can therefore be viewed as a risk management device.
Opponents on the other hand, were concerned about the extra time and costs that are related to a gamma test and the impact it has on the "time to market". They argued that Comp. X can achieve the same goals in a staged market introduction.

2. *Larger focus on the customer in tests*

Many respondents felt that the opinion of the customer and their experience with the product has sometimes been underexposed in previous tests. Because the strength of Comp. X lies in its large

customer base and marketing power, they feel that this should be reflected by a large focus on the customer in tests.
Both managers and employees of Comp. X agreed that in previous test technical issues played the most important part while this is not in Comp. X's interest. Comp. X is not the one who will actually build the new products (at least not on their own) and should therefore emphasize the importance of the customer acceptance dimension in tests. Related to this point, is the question whether Comp. X should be involved with technical validation at all. Some respondents state that Comp. X should only be involved in a validation process when the technical validation of a product is completed. This relates to the suggestion of one of the managers to more often perform a market/competition analysis to compare a product Comp. X is developing with other available products. This way Comp. X can decide whether to go forward with their own development process or simply buy the product from another organization.

3. *More thorough business case and plan validation*

Next to the desire to get more insight in the actual customer experience and customer acceptance, several respondents stressed the importance of a proper validation of the business case and plan.
While Comp. X develops new products and services, they hand them over to other departments within Comp. X, such as marketing, when they plan to actually launch the product into the market. As a consequence the importance and participation of other departments should grow along with the maturity of a product. The acceptance of a product by other departments is for a large part based on the business case and plan of the product; therefore it is important to validate this as soon as possible. Of course this is also convenient for the NBD department itself to adjust their resources and business planning on.

The right moment for a proper business case and plan validation is the beta phase. In this phase the product is technical completed and actual customers are involved. Waiting until the gamma phase is too late. In the gamma phase the other departments already need to be involved as the product is tested on its integration with other systems. Furthermore it often happens that an organization anticipates on a success of the gamma phase and is already scaling up. In addition a gamma test is very time and money consuming itself, so having good insight in the financial potential of a product before starting a gamma test is required.
Validation the business case in the alpha stage, on the other hand, is too early. This phase mainly focuses on the technical aspects. Based on the experience in this phase the product can still change fundamentally.

4. *Better structured and objective evaluation:*

Managers from Comp. X indicated that problems with evaluation often occur because there is no clear planning for a test and no structured way of working. Especially intermediate evaluation is overlooked this way.
While it is important to limit the evaluation moments and keep a healthy balance with one's own decision authority and the projects pace, proper evaluation is performed too little at the moment. Many respondents therefore think that more people should be involved in the evaluation process and also from other departments such as marketing. This way evaluation becomes more objective than when only one NBD employee is responsible. It also generates internal support for the project,

which is important since Comp. X needs to hand over the product to other departments before the actual marked introduction.
Input from other departments could also be useful at other moments then only at the evaluation. In the form of a steering committee representatives from other departments within Comp. X could be a sparring partner as well as an evaluation board. Some projects already have such a committee in place but in a more advisory role.
All respondents agreed that a better structured and more objective evaluation will be a large improvement.

Test restrictions

Apart from the design requirements there were also some test restrictions identified.
The first, and obvious one, is the available funding and FTE of Comp. X. The amount of employees available for tests is limited and the NBD budget is not large enough to simply fund each desired test. Executing alpha and beta tests is not so much of a problem, but gamma tests are generally too expensive to carry out on their own.

Another factor related to the available funding for projects is the fit of a product with the strategic vision of the higher management. This is not a factor within the new product development process itself but it shapes the context around it. It often happens that a trend or hype is recognized on a higher management level and Comp. X is dictated to explore it. This also works the other way around. When the higher management does not see the added value of a specific product they can pull the plug out of it.
When a project exceeds a predetermined funding need, it needs to be approved by control divisions within Comp. X. Capacity of these departments and the decision processes within Comp. X are therefore also factors which can complicate a development process. This is linked to the strategic fit in a way that when the fit (and thus higher management support) is larger, decision and approval procedures run smoother.

An external test restriction is the dependence of many product tests of seasonal influences. This dependence holds for products that aim to save energy or generation energy and was also mentioned in point eight of this chapter. To properly test and validate assumptions on these products, including winter months in the test period is necessary. This is the time in which the energy and especially gas consumption of Dutch households is the highest. Minimal test periods of these products are therefore relative long and starting the test on the right moment (before the winter) is very important.

Integration of the findings

In this chapter the practice at Comp. X was studied. A verification was made to what extent the identified test dimensions and guidelines for tests within the testing and validation stage in theory, are already applied by Comp. X. The problem areas that were identified are:

Problem area 1:	The Alpha tests in total
Problem area 2:	The products used are often not full working prototypes
Problem area 3:	The test dimensions Business case and plan and Integration and Support are poorly addressed
Problem area 4:	Evaluation is not performed regularly and objective
Problem area 5:	The product design is not frozen during all tests

The second step in this chapter was the identification of design requirements Comp. X has of its own for a testing and validation model. Four requirements were identified:

1. The division of the testing practice in three tests ascending in size (Alpha, Beta and Gamma phase) should be preserved
2. The focus on the customer in tests should become larger
3. Business case and plan should be validated more thorough
4. Evaluation should better be structured and become more objective

When looking at these findings it becomes apparent that design requirement 3 and 4 exactly match problem area 3 and 4. This means that Comp. X confirms that these aspects are of importance and need to be improved.
Design requirement 1 is also related to a problem area, area 1, but conflicts with it. The alpha FCBS test does not fit within the described theoretical criteria but Comp. X on the other hand wants to preserve this phase.
Design requirement 2 is not related to a problem area but has to do with the validation of the customer acceptance dimension, a part of criterion 3 "*Effective test programs recognize the full set of benefits of all four test dimensions".*

In the next chapter the findings from this chapter on all criteria will be analyzed in order to derive the implications for the model that is build. Criteria who comprise a problem area or to which a design requirement is linked will be studied more in depth. As problem area 1 and design requirement 1 are not related to a specific criterion, their implications will be discussed at the beginning of the next chapter

5. Combination of theory and practice

In the previous chapter the scores on the theoretical criteria of different tests Comp. X carried out in the last two years were described. Criteria which comprised problem areas where identified as well as Comp. X's own design requirements for a testing and validation model. The chapter concluded with linking these design requirements to specific criteria.

In this chapter the findings for each criterion are analyzed, together with the corresponding design requirement, in order to determine if there are implications for the model that is developed. This way the theoretical model can be adjusted and specified to match the practice of Comp. X.
Since problem area 1 and design requirement 1 could not be related to a specific criterion their impact will be discussed first.

5.1 Implications of the practice at Comp. X for the theoretical model

The first problem area which was identified comprised the complete score set of the FCBS alpha test. This test scored 8 times red out of eleven criteria and among the red scores where the two criteria which are distinctive for tests within the testing and validation stage:

1) *Physical prototypes which resemble the commercial model are used*
2) *The testing is done outside the organization in a users' actual environment*

The fact that these criteria did not hold for the FCBS test means that this test is not a part of the testing and validation stage from a theoretical point of view.
This is an important remark as Comp. X did identify the alpha phase, together with the beta and gamma phase, as a part of the testing and validation stage. Additionally their first design requirement was the preservation of these three phases in the model that is developed.
In Comp. X's view, the purpose of the alpha phase is to proof that the concept technically works. While this does not match the theoretical view on the testing and validation stage, it does look very similar to the purpose of a preceding stage in the stage gate model (Cooper, 2001): the development stage.
During the development stage, plans made in the previous stages are executed. The product's design and development is carried out and some early tests are performed. Furthermore marketing and production plans are developed. The completion of the first working prototype marks the end of this stage (Cooper 2001).
The idea that the FCBS test and alpha tests in general do not belong to the testing and validation stage is further supported by the many red scores of the test on the other criteria. Table 5i showed that the test only focused on the product function dimension and while Comp. X was certain of the benefits and held a proper evaluation, all other criteria were neglected.
The reason for this inconsistency is likely to follow from a different view on the product development process. Comp. X views all tests as part of the testing and validation stage while the development stage also comprises testing according to literature. While placing the alpha phase in one of the two stages looks like a minor detail it has important consequences.

By following the view of Comp. X, gate 4 (in which all technical requirements need to be evaluated) is skipped. Only when a product completely resembles all commercial functions it may be allowed to pass. By skipping this gate there is a large chance that (redundant) technical problems emerge in a later stadium. This exactly happened in the OT, MCHP-B and MCHP-G tests. In the MCHP-B and MCHP-G test, technical adjustments were necessary during the tests and the OT only worked on a specific type of condensing boiler. To prevent these kinds of problems from happening in future tests the alpha phase is placed in the development stage, before gate 4.
By placing the alpha phase in the development stage of the stage gate model (Cooper, 2001) the design requirement of Comp. X, *The division of the testing practice in three tests ascending in size (Alpha, Beta and Gamma phase) should be preserved,* is respected and the mismatch with the theoretical criteria of the testing and validation stage is solved.

Placing the alpha phase in the development stage also means that the theoretical guidelines which were used as criteria are not applicable anymore. They were specifically identified for test within the testing and validation stage. Also the four test dimensions which should be addressed in the testing and validation stage do not apply anymore. In order to produce a full working prototype it is not necessary to test on the customer acceptance dimension or on the integration and support one. A proper validation of the business case and plan is also too early.
However when the components within these dimensions are completely ignored during the development stage, it is not likely that the prototype will live up to the market requirements. For example production costs are a factor which should already be taken into account in the development stage. When the production costs of a product are higher than the estimated selling price in the best case scenario, it has no use to go forward with the development process. Aspects from the other three dimensions are therefore certainly important in an alpha test but not as specific test goals. Instead they should be threatened as design requirements (e.g. the product should be easy to use, it must have the possibility to connect to internet, the production costs may not exceed 200 Euros, etc).
The new position of the alpha phase is reflected in figure 9. Just as in figure 6, the circle symbolizes a complete test but now the red section indicates on which test dimensions the focus should be.

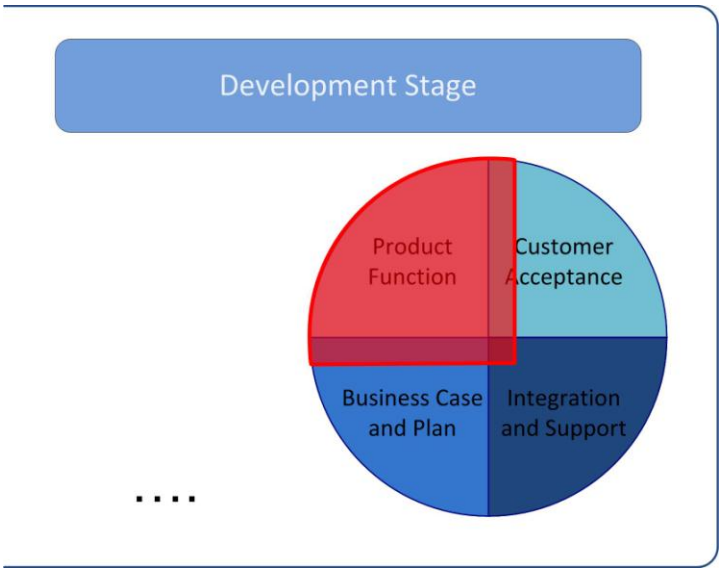


Figure 9: Alpha phase

Now that a solution for Problem area 1 and design requirement 1 has been established, the theoretical criteria are discussed. This discussion is limited to the scores of the ED, OT, MCHP-B and MCHP-G since the FCBS scores do not apply anymore.

Criterion 1 and its corresponding problem area, products that are used are often not full working prototypes, were already mentioned in the above section.
The OT, MCHP-B and MCHP-G tests all scored orange on this criterion, which means that Comp. X meant to test a prototype that resembled all commercial functions but during the test it turned out that fundamental changes were necessary. This leads to the conclusion that Comp. X did not evaluated the product properly before the test started, at that point these technical issues should already been recognized. The three tests which scored orange furthermore demonstrated that not applying this criterion can have a negative influence on the usefulness of a test. The positive side of the orange scores however is, that Comp. X did try to follow this criterion and acknowledges its relevance. The criterion therefore remains a part of the model that is developed. In order to make sure that similar problems as in the OT, MCHP-B and MCHP-G tests do not happen again the evaluation of the new product development process in general also needs to be improved. This will be discussed at point 6 (regular and objective evaluation) in this chapter.

Criterion 2 was also already mentioned above. Unlike criterion 1 this criterion was applied in all four tests and does not comprise a problem area or relates to a design requirement. Therefore this criterion should remain a part of the model that is developed. However in order to fit into that model it should become a guideline instead of a condition. The new guideline is:

2) *The testing must be done outside the organization in a users' actual environment*

Criterion 3 was on including all four identified test dimension in a test and because each of the four test dimensions was scored individually they will also be analyzed individually:

3) *Effective test programs recognize the full set of benefits of all four test dimensions.*

Product function dimension
In all tests there was a lot of attention for the Product Function dimension. In the MCHP-B and MCHP-G the main focus was on this dimension and while the focus was divided among the test dimensions at the start of the OT, the occurrence of technical problems led again to a lot of attention for product functioning.
Only in the ED test the importance of this dimension was overshadowed by the focus on the customer acceptance dimension in both design and evaluation.
In addition to the scores of the tests all respondents also indicated that technical issues often play the most important part in tests. This emphasis partially follows from the fact that many of the products that are tested in customer environments do technically not yet match the commercial standards and technical adjustments are necessary during the test. This was also the case in the OT, MCHP-B and MCHP-G test. Although the theory also states that a proper validation of this dimension is desirable, it should be in balance with the other dimensions. Especially because Comp. X is not the one who will actually build the new products (at least not on their own). This means that Comp. X

should make sure that a product resembles all commercial functions before starting a beta or gamma phase so that they can validate whether the product works in a users' actual environment instead of validating if the product works at all.

Customer acceptance dimension

This dimension was not identified as a problem area because the main purpose in the ED- and OT-test was to validate this dimension, and both tests scored green. However, Comp. X did formulate a "*larger focus on customers in tests"* as a design requirement. This means that this dimension is important for them. In the MCHP-B test this dimension was only slightly addressed. Obvious questions were asked but many opportunities to learn from how the customer experiences the product were overlooked. The same holds for the MCHP-G test in which this dimension is neglected completely.

The concentration on the customer acceptance dimension in the ED- and OT-test in relation to the lack of attention in the MCHP tests can be found in the nature of the products. The ED and OT aim to change and/or influence the behavior of customers and in order to substantiate this, a large focus on the customer is necessary.

Because the strength of Comp. X lies in its large customer base and marketing power, Comp. X feels that the focus on the customer acceptance dimension should be independent from the nature of the product and also be applied in tests such as the MCHP. It is important for Comp. X to keep in touch with the preferences of their customers and adjust their products accordingly instead of simply pushing new products into the market. The importance of this dimension should therefore be reflected in the model that is developed.

Business case and plan dimension

As this dimension was poorly addressed in most tests, it was identified as a problem area. Often only single aspects of the business case were measured (such as energy savings) while respondents were not asked what they are willing to pay for the product or what their attitude is towards a potential proposition.

This was the case with the ED as well as both MCHP tests. Many factors in the business cases of these products are therefore still based on estimates, which undermines the strength of a business case as a whole. Only in the OT-test learning goals of this dimension were established and evaluated properly but since the participants in that test were eleven Comp. X employees this validation was not very strong.

The respondents from Comp. X also recognized that this dimension is underexposed and made "a more thorough business case and plan validation" one of the design requirements. The beta phase was selected as most appropriate for this task. In this phase the product is technical completed and actual customers are involved. Waiting till a gamma phase would mean that decisions on investments and commitment from the organization need to be made on assumptions alone, which is not desirable. Just as the customer acceptance dimension, the importance of this dimension should become larger in the model which is being developed, compared to its current status.

Integration and Support

This dimension was also identified as a problem area. In the ED- and OT test the integration and support dimension was almost completely neglected. In both tests the products operated separately from Comp. X's systems and were monitored by a third party. In the MCHP-B test the validation of

this dimension was slightly better. The products were connected to the regular in house heating system but were controlled by a separated unit. Validating the integration and support dimension of the MCHP properly is one of the main goals in the MCHP-G test.
The lack of attention for this dimension in beta tests on the one hand, and the specific focus on it in gamma tests on the other hand can be explained by the vision of Comp. X on these tests.
In their view the customer should be placed central in the beta phase and in the gamma phase the organizational capacity to effectively install, manage and support the new product on a larger scale is validated. These last aspects are all part of the integration and support dimension. Therefore, Comp. X is aware of the importance of this dimension only they do not apply it in earlier phases.

According to theory this does not have to be a problem, as long as the dimension is addressed properly in the testing and development stage as a whole.
From a practical point of view it is understandable that Comp. X waits with testing this dimension to the gamma phase. A complete integration of a test product can be a large effort while at that point it is still unsure if this will actually be necessary in the future. Additionally it largely depends on the complexity of the product whether this dimension is really relevant. Some products work completely by themselves, where others need to be integrated with several other products and infrastructures.
There can be concluded that from a practical point of view, the choice of Comp. X not to focus on this dimension in the beta phase is logic and that this does not lead to problems with the theory as long as the dimension is properly addressed in the gamma phase.
This difference in importance of this dimension in the beta and gamma phase will therefore be used in the model which is being developed.

In the above section several implications of the practice of Comp. X, on the importance of the four test dimensions in the testing and validation stage, were identified:

- The focus on the Product function dimension should better be balanced with the other dimensions
- The Customer acceptance dimension should become more important
- The Business case and plan dimension should already be properly validated in the beta phase
- The Integration and support dimension should be tested and validated in the gamma phase

The theory furthermore prescribed that the importance of the test dimensions can vary between tests or phases as long as they are all addressed in the testing and validation stage as a whole. Therefore a differentiation will be made between their importance for the beta and gamma phase.

In a beta phase it is essential that Comp. X realizes that it is has to test a product on technical issues in a users' actual environment, but that the customer itself needs to be central. Because a beta phase consist of a relative small group of participants it is very well suited for in depth analysis of customer preferences and attitudes towards the product, thereby validating the complete customer acceptance dimension. Testing and validating the complete business case and plan dimension should furthermore become part of this dimension.
While the timing in a beta phase to do so is right, the possibilities are limited. As the beta phase consists of approximately 20 households, this is not enough for a thorough validation. Simply increasing the number of test persons in the beta phase is not an option as this also means a large increase in the costs of the phase. To tackle this problem Comp. X should expand the beta phase with

a concept based market research. This way a large group of customers can be questioned about possible selling prices and propositions, without making a lot of extra costs. An additional advantage is that the opinion of people based on a concept is more likely to match the attitude towards the product of future customers than the opinion of persons who got to extensively try the product for free.

As explained, it is not practical to also address the complete integration and support dimension in a beta test. While the focus on this dimension in the beta phase is therefore limited, it is still advisable to already recognize the potential importance of this dimension. Comp. X should try to validate the aspects that are possible. Figure 10 reflects the beta phase in its new form. The circle symbolizes a complete test and the importance of the four test domains is shown by the red section. The concept based marketing research which should be executed parallel is the purple ring around the test (O).

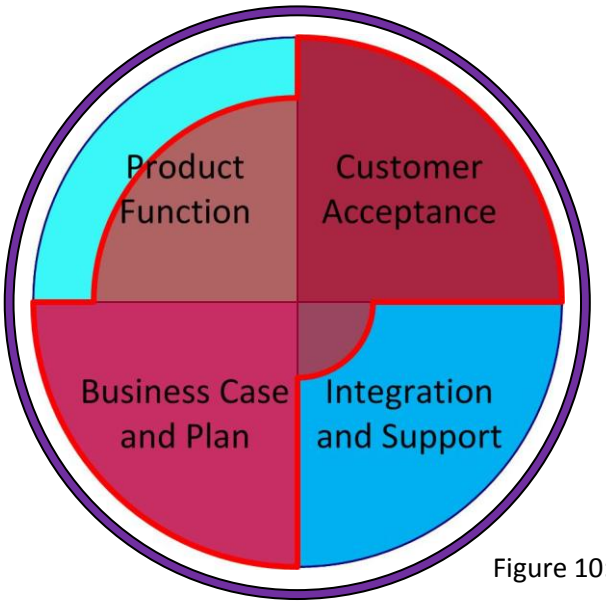


Figure 10: Beta phase

According to the findings in this above section, the gamma phase its main focus should be on the integration and support dimension. The other three dimensions should not be neglected but when the newly developed beta phase is executed properly, most aspects are already tested. The gamma phase can then be used to verify the earlier results on the customer acceptance dimension and business case and plan dimension for a larger test group. This way the assumptions made on these outcomes can be strengthened. Especially for assumptions on potential energy saving a greater amount of test results adds value. Technically the product should be completely finished when starting this phase since otherwise its integration has no use. Ideally there would be no attention needed for the product function dimension at all. However, it can happen that a technical problem comes to light by testing the product on a larger scale and obviously this should be solved.

The red section in figure 11 shows this division of importance for a gamma test.

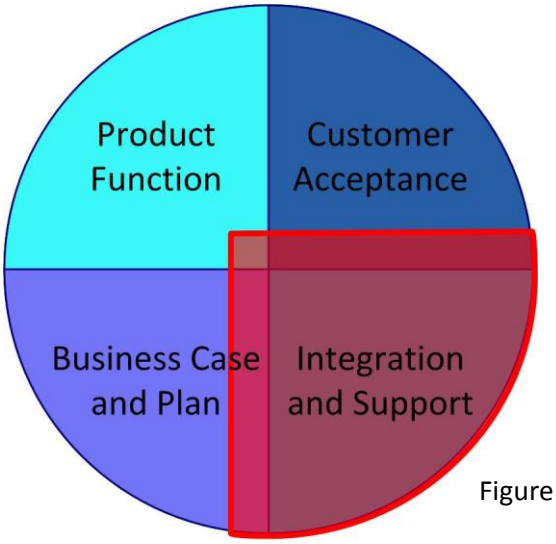


Figure 11: Gamma phase

When the original criterion is recalled "*Effective test programs recognize the full set of benefits of all four test dimensions"* there can be concluded that this guideline is still applicable but interpreted in a different way. Instead of focusing on all four test dimensions in each specific test, as the theoretical model prescribed, a division was made between the beta and gamma phase. In the beta phase the customer acceptance dimension and the business case and plan dimension should receive the most attention and in the gamma phase the integration and support dimension is placed central.
To make sure this guideline is interpreted in the right way it is specified into:

"*Effective test programs recognize the full set of benefits of all four test dimensions and address all four of them in the testing and validation stage as a whole"*

Criterion four:

4) *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience.*

This criterion was not identified as a problem area and does not relate to a design requirement. The practice at Comp. X revealed that the importance of this guideline is recognized as they try to fulfill it in their tests. In the ED test for example, the selection process was specifically designed to match the segmentation of the market. The same holds for the MCHP-G test.
The reason for not always following this guideline is often related to the pace of a project. This was the case with the OT test. As the product can only be used during cold months Comp. X was forced to test in the winter, but the product was not technically completed until February. To save the time it takes to find actual customers willing to participate, Comp. X decided to approach own employees. While this choice might be understandable it is not a good one. Results from such a test are not objective and there is a high chance they differ from a test with actual customers.
Another factor which can lead to ignoring this guideline is a technical requirement. Designers sometimes want to test the product under specific circumstances, thereby excluding possible participants. The MCHP-B test resembled a real world experience but only households with high heating costs were selected. This way the MCHP could be tested extensively, a technical requirement, but the test as a whole became less generalizable. As extensive testing can also be

done in a laboratory, reflecting the segmentation of the market should be more important. Especially since the customer acceptance dimension is so important for Comp. X in a beta test, they should place more emphasis on a good selection process.
The advantages of selecting own employees, one respondent mentioned, are important to recognize but are also captured to a large extent by the next guideline. Decreasing exposure when there are large risks on problems or generating internal support, can also be reached by starting the testing at own employees or friendly customers and then grow the number among actual customers. Also when a product has to remain secret it is necessary to test it in a users' own environment at some point.
There can be concluded that while there are sometimes reasons at first sight for not following this guideline, it is in the best interest of the project when it is followed. Comp. X should select and approach participants as close to the actual market segmentation and experience as possible, independent from project pace and technical requirements. This guideline will therefore be a part of the model that is being developed.

Criterion 5:

5) *Grow the number of testers over time and begin with sophisticated customers with good relations to the firm*

This criterion was followed very well by Comp. X in all analyzed tests. It was therefore not associated with a problem area or related to a design requirement. Comp. X always starts testing at an own employee or customer they have a personal relationship with. This decreases the chance on negative exposure when things go wrong. Comp. X thereby proved the theoretical value in practice. The idea behind this guideline is also embedded in the alpha, beta, gamma division as the amount of participants increases with each phase. Because of its relevance this guideline will be embedded in the model that is being developed and does not have to be adjusted.

Criterion 6:

6) *Monitor and evaluate the product performance on a regular basis in an objective way.*

This criterion is very significant as it was identified as a problem area, Comp. X made a design requirement about it and the findings on criterion 1 also indicated that the evaluation process needs to be improved.
When reviewing the previous tests it became apparent that there were no pre defined evaluation moments in place, furthermore there were large differences in the scores of the described tests. Products that were evaluated by an independent third party such as the ED were evaluated fine (green score) while products that were evaluated by Comp. X itself or an project partner were evaluated a lot worse and scored orange (OT) or even red (MCHP-B).

The lack of a good evaluation process was also identified by Comp. X. They relate it to the absence of a clear planning and structured way of working in tests and formulated the following design requirement: *Better structured and objective evaluation.*

However, this lack of a structured evaluation does not only hold for the testing and validation stage but for the whole new product development process. The gates that form the evaluation moments between stages are taken “on the run” instead of clear and structured on specific moments.
The lack of a proper evaluation throughout the process is furthermore reflected by the very small amount of projects that are being killed by Comp. X’s management. When it does happen, often the responsible employee working on the project itself indicated that it is better to stop with the project instead of the management.
Another indicator for a bad evaluation is the fact that several of the products that were tested in consumer environments, still had fundamental technical problems. In point 2 of this chapter this was already linked to a dysfunction of the gate before the testing and development stage (gate 4).
Proper evaluation at that point is very important since actual tests with consumers are very expensive, both time and money wise, and technical problems negatively influence the benefits.

The reason for this lack of a structured evaluation can be found in the incomplete implementation of the stage gate model for the NPD process Comp. X developed on their own.
In the theoretical model specific evaluation moments were prescribed (between each test) and together with the regular gates of the stage gate NPD process they provide more than enough evaluation moments in a structured way. The challenge is therefore to find a way to actually use them and the questions is: who should do the evaluation.
Because the theoretical model is on a general high level this aspect was not addressed, but it turns out to be very relevant for Comp. X. In the model which is developed it should therefore be clear how the evaluation process should to be executed.

Many respondents indicated that it would be wise to involve persons from other departments within Comp. X in the evaluation process to increase objectivity as well as internal support. However, they also stressed to keep a balance with the decision authority the person working on the project should have.
To solve this issue, the input from the rest of the organization should be increased together with the maturity of the product. When a product is launched into the market the organization needs to be involved anyway, so increasing their influence step by step from an early point in the development process can only help to improve their understanding and support.
Furthermore it is important to avoid that only one person is responsible for a product. This was the case with the OT and it increases the chance that the responsible person will take critics personally. Because it can happen that the market or other developments catch up on a product, without this being someone’s fault, this is not a desirable situation. An objective evaluation should be guaranteed.
The management team of NBD is most suited to evaluate products during the first new product development stages. They should also be the ones guarding the gate between the development stage and the testing and development stage. They have to make sure that only products which fulfill the technical market requirements are allowed to pass. As the alpha test of Comp. X is placed in the development stage this means that the NBD MT should be responsible for its evaluation. This should be no problem because at that point a project is still relative small and easy to manage.

When the evaluation of an alpha product test is completed, and the decision is made to proceed to the testing and development stage, the project becomes more complicated as costs and exposure grow. At this point other departments should be involved.
It is best to start this involvement by forming a steering committee with representatives from relevant departments, who can advise during the design of a beta test. This way they have the chance to embed aspects that are important to measure for their own department in the beta test. By designing a test in which all relevant aspects for Comp. X are measured, the evaluation report will be much more fitted to the criteria of the other departments.
Depending on the importance and complexity of a product, the evaluation of the beta phase can be done by the NBD MT alone or together with the steering committee. In any case the results should be presented to the steering committee so they can decide together with the NBD MT whether to proceed to the gamma phase.
Because a gamma test focuses on the integration and support dimension, the other departments need to be actively involved in this test to make the integration a success. There role in the test design and evaluation should therefore be formalized accordingly. Higher management should also be updated on the project status since they are the ones deciding if the product will actually be launched into the market.
When Comp. X actually follows this evaluation process it will lead to better funded decisions on the development process and will prevent the execution of tests which add no value. Figure 12 reflects the practical fulfillment of the theoretical guideline which will be used in the model that is developed:

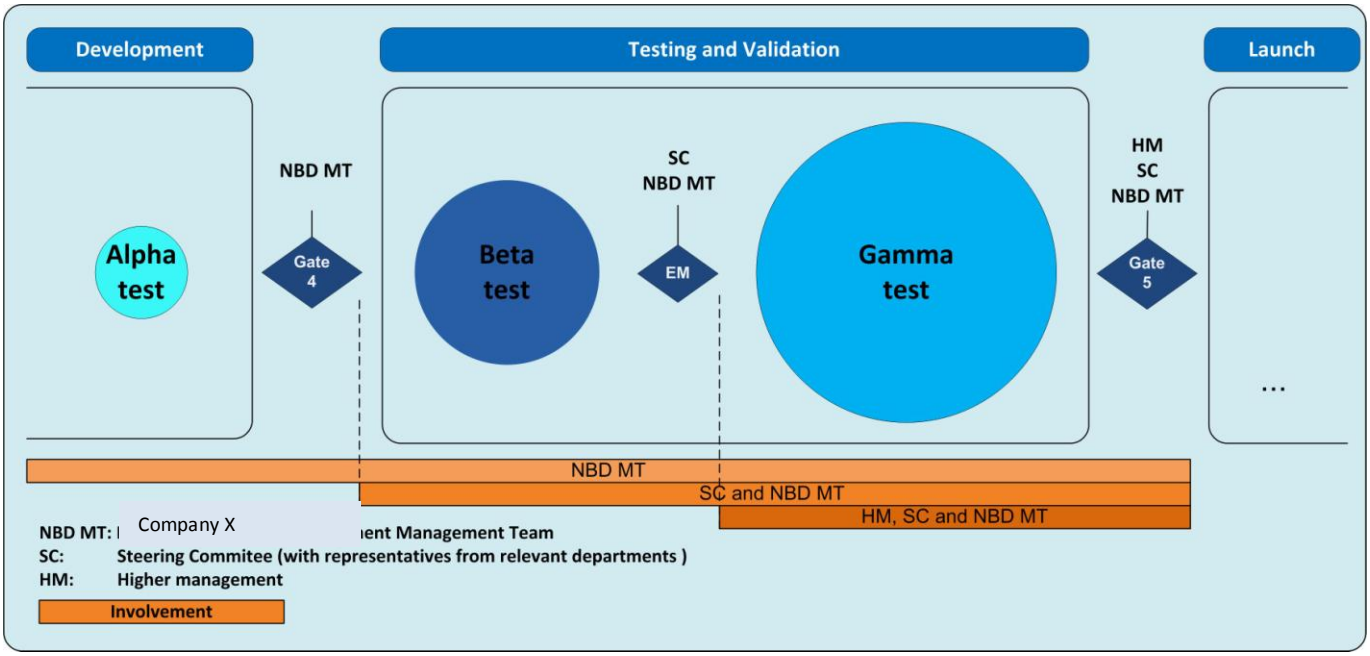


Figure 12: Evaluation process

When interpreting this model one should realize that the appointment of an organizational group to a gate or evaluation moment does not mean that they should actually be the ones analyzing and processing the data from the test. This is the task for the project team or an external research agency, but they should present the results to the appointed organizational groups. The appointed organization groups have to evaluate the results and should decide what will happen with the

project. As mentioned before, it is wise to involve the responsible evaluators also in the design of a phase, so that it can be adjusted to their preferences.
Apart from the hard evaluation moments between stages or tests, the experience Comp. X had in previous tests showed that regular evaluation and monitoring is very valuable during a test. The halfway interviews in the OT test really paid off since the identified problems could be solved, this led to better results for the remaining test period. In the MCHP-B this method was not applied and errors in the datasets were discovered when the test was already finished. The value of this aspect from the theoretical guideline was hereby confirmed by Comp. X in practice and should be embedded in the model which is being developed.

When criteria 6 is recalled, *Monitor and evaluate the product performance on a regular basis in an objective way,* it can be concluded that Comp. X acknowledged that it is important to apply this guideline but that they often neglected or found it difficult to do so. Therefore more substance was given to the guideline by translating it into a workable practice for Comp. X, visualized in figure 12. The guideline and its practical application will both be embedded in the model that is developed.

Although the evaluation process in figure 12 should become the department standard, there is one variable which should be taken in account; the position of Comp. X in a test. While in many NPD tracks Comp. X is a co developer of the product and is actively involved in the route to market of the product, they are sometimes only participating to gain publicity or satisfy an information need. In such a case the attitude of Comp. X towards a proper evaluation is different since the results are of less importance. In such a case it might not be necessary to actively involve other departments. This issue is further addressed at criterion 8.

Criterion 7:

7) *Product design should be frozen prior to a test*

The test scores on this criterion reflect that Comp. X does not always follow this underlying guideline (OT scored orange and both MCHP tests red) and it was therefore identified as problem area 5 (the product design is not frozen during all tests).
The three tests in which this criterion was not followed have in common that in each of the tests the product was technically not completely ready yet. Experiences with tested products in these tests were directly used to build a new type of the model.
While it is possible that technical problems are not noticed until the product is tested in a consumer environment, in these cases it is more likely that the products were tested too soon. The fact that Comp. X should more carefully guard the gate before the testing and validation stage, which was already mentioned before.
Apart from the technical readiness there are complicating factors for Comp. X in freezing the product design of products tested. For example Comp. X often wants to measure behavioral changes or energy saving potential.
Properly testing and validating these kind of effects take a long time in which the product, according to literature, may not be developed further. However the practice at Comp. X showed that it is not realistic to stop a development process for more than half a year as this has a big impact on the time to market. Then again, it has no use to extensively test and validate a product when an improved product is already being developed. The choice was therefore made to make this guideline a part of

the model that is being developed, but only for the gamma phase. In the gamma phase the product is integrated in the actual organization, which has no use when the product is still being altered.
This results in an adjusted guideline to be used in the model that is developed: *Product design should be frozen prior to the gamma test*

Criterion 8:

8) *Be certain of the benefits a test will deliver before starting*

The guideline underlying this criterion was established to make sure that organizations carefully weigh the benefits of a test against the costs and time that needs to be put into it, before the test is started.
In the ED and MCHP-B test Comp. X did match this criterion but the red score on the OT test showed that Comp. X does not apply this guideline strictly. In order to keep pace in the product development process, the OT test was designed and executed while Comp. X could have recognized beforehand that the learning goals in the test could not be met.
The lesson which can be learned here is that Comp. X is limited in their test possibilities by seasonal influence. This was also mentioned as a test restriction by Comp. X. Many products can only be tested and validated during winter months. Delays in a test schedule can therefore have serious consequences for the usefulness of a test. Recognizing this dependence is important and Comp. X should reconsider the benefits of a test with each delay.
Furthermore the MCHP-G test showed that there is sometimes a mismatch between the goals in a test and the role Comp. X plays. Comp. X puts a lot of resources in this test while there are no new benefits for Comp. X compared to the MCHP-B test, other than additional exposure. This should be prevented by making sure that the goals and benefits of a test are clear to everyone before starting and adjusting the role of Comp. X specific to these goals. When other aspects in the current testing and validation practice, such as guarding the gates and improving the evaluation process, are applied it will be much easier and natural to do so.
Because the value of this guideline was proven by Comp. X, and it is beneficial that it will be applied in future tests, it will be embedded in the model that is developed. To clarify the aim the guideline has even better the guideline is somewhat adjusted. Not only should the benefits of a test be clear before starting, they should also be agreed upon by both project management and the evaluators. This way uncertainties and misunderstandings can be prevented. The guideline as applied in the model that is developed will be:

"Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators before starting"

The remark some respondents had on the first design requirement, *Alpha, Beta and Gamma division should be preserved,* also relates to this guideline. They were concerned about the extra time and funds that are related to a gamma test and the impact it has on the time to market. They argued that Comp. X can achieve the same goals in a staged market introduction. A complicating factor for executing a gamma phase is the test restriction of available funding and FTE. Gamma tests are too expensive to be executed from the Comp. X budget which means a project should also be backed by other departments or project partners.

Whether or not the gamma phase is necessary is for a large part a risk management considiration and depends on the complexity and requirements of a product. However, to make a well based decision, proper insight in the results of previous tests and a clear overview of the goals and benefits which will be achieved with the gamma test are necessary. This stresses the importance of a proper evaluation as wel as following the newly formulared guideline. Since it is not possible to give a definite answer to the question if the gamma phase is always necessary as a test, or that its goals can also be fulfilled with a staged market introduction, it will be made optional in the model that is developed. After a beta phase is evaluated there should be decided whether to stop the project, move back to a previous stage, move forward to the gamma phase or move directly to a market introduction.

5.2 Integration of the findings into a testing and development model suitable for Comp. X

In the previous paragraph the testing practice of Comp. X and the design requirements of Comp. X were analyzed, guided by the criteria from theoretical framework. In this paragraph the findings from that analyses, the implications of the practice of Comp. X for the model that is build, will be integrated and used to adjust the theoretical model to suit Comp. X's work methods.

The first issue which was explored in the previous paragraph was the fact that the described alpha test in chapter 4 did match the theoretical criteria at all, but that Comp. X made the preservation of this phase a design requirement.
The issue was solved by placing the alpha phase in the development stage of the stage gate model (Cooper, 2001). As the purpose of the development stage is the development of a prototype which resembles all commercial functions, this stage is also more fitted to the main focus in the alpha phase; a technical prove of concept. Since the criteria on which the alpha test was scored were specifically established for tests within the testing and development stage, they did not apply anymore. Also the relevance of the four identified testing dimensions changed. The main focus in the alpha phase should be on the product function dimension and components of the other three dimensions should not be tested but used as design requirements. Figure 9 reflected this new position of the alpha phase and the division of importance over the test dimensions.

The next step was analyzing the scores on the 8 criteria one by one, together with the corresponding problem areas and design requirements. It was concluded that criterion 1 and 2 remain a part of the model that is build.

1) *Physical prototypes which resemble the commercial model are used*
2) *The testing should be done outside the organization in a users' actual environment*

In the analysis of the third criterion it became apparent that it is still applicable but should be interpreted in a different way when used by Comp. X. Instead of focusing on all four test dimensions in each specific test, as the theoretical model prescribed, the four test dimensions are addressed in the testing and validation stage as a whole. A division of the importance of the four test dimensions was made between the beta and gamma phase, which was reflected in figure 10 (beta phase) and

figure 11(gamma phase). This division was made to accommodate the design requirements of Comp. X to focus more on the customer and validate the business case and plan more thorough.
There was concluded that in the beta phase the customer acceptance dimension and the business case and plan dimension should receive the most attention and in the gamma phase the integration and support dimension is placed central. In order to make a proper validation of the business case and plan dimension possible in the beta phase, a concept based marketing research should be executed parallel with the test. To make sure this guideline is interpreted in the right way it was specified into:

3) *Effective test programs recognize the full set of benefits of all four test dimensions and address all four of them in the testing and validation stage as a whole*

As the relevance of criterion 4 and 5 was confirmed by the practice of Comp. X, they will both be a part of the model that is developed:

4) *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience*
5) *Grow the number of testers over time and begin with sophisticated customers with good relations to the firm*

The analyses of criterion 6 was very important since this guideline was identified as a problem area, Comp. X made a design requirement about it and the findings on criterion 1 also indicated that the evaluation process needs to be improved.
It was concluded that the problems with the evaluation at Comp. X were mainly caused by the absence of a clear structure to execute this process. More substance was given to the guideline by translating it into a workable practice for Comp. X, visualized in figure 12. In this practice the role and importance of other parts of the organization grow along with the maturity of the product. To make sure that all aspects are tested which are relevant for the evaluators, they should be involved from the moment the test is being designed. Based on the results a go/recycle/kill decision of the project can be made. The guideline and its practical application will both be embedded in the model that is developed:

6) *Monitor and evaluate the product performance on a regular basis in an objective way*

Criterion 7 embedded problem area 5; the product design is not frozen during all tests. In the analyses it turned out that Comp. X often tests on aspects like energy saving potential, which requires a very long test period. It is therefore not realistic to stop a development process of a product in each test. The choice was therefore made to make this guideline part of the model that is developed, but only for the gamma phase. In the gamma phase the product is integrated in the actual organization, which has no use when the product is still being altered. The guideline was adjusted into:

7) *Product design should be frozen prior to the gamma phase*

The analysis of criterion 8 revealed that following the guideline underlying this criterion is very important to be used in tests. It can prevent a mismatch between the goals in a test and the role Comp. X plays and it can also put a stop to the executing of tests in which the goals cannot be met

within the test design. Goals and benefits of a test should be clear to everyone before starting and that the role Comp. X plays should be specifically adjusted to match them. To clarify the aim of the guideline, it was slightly adjusted for the model that is developed:

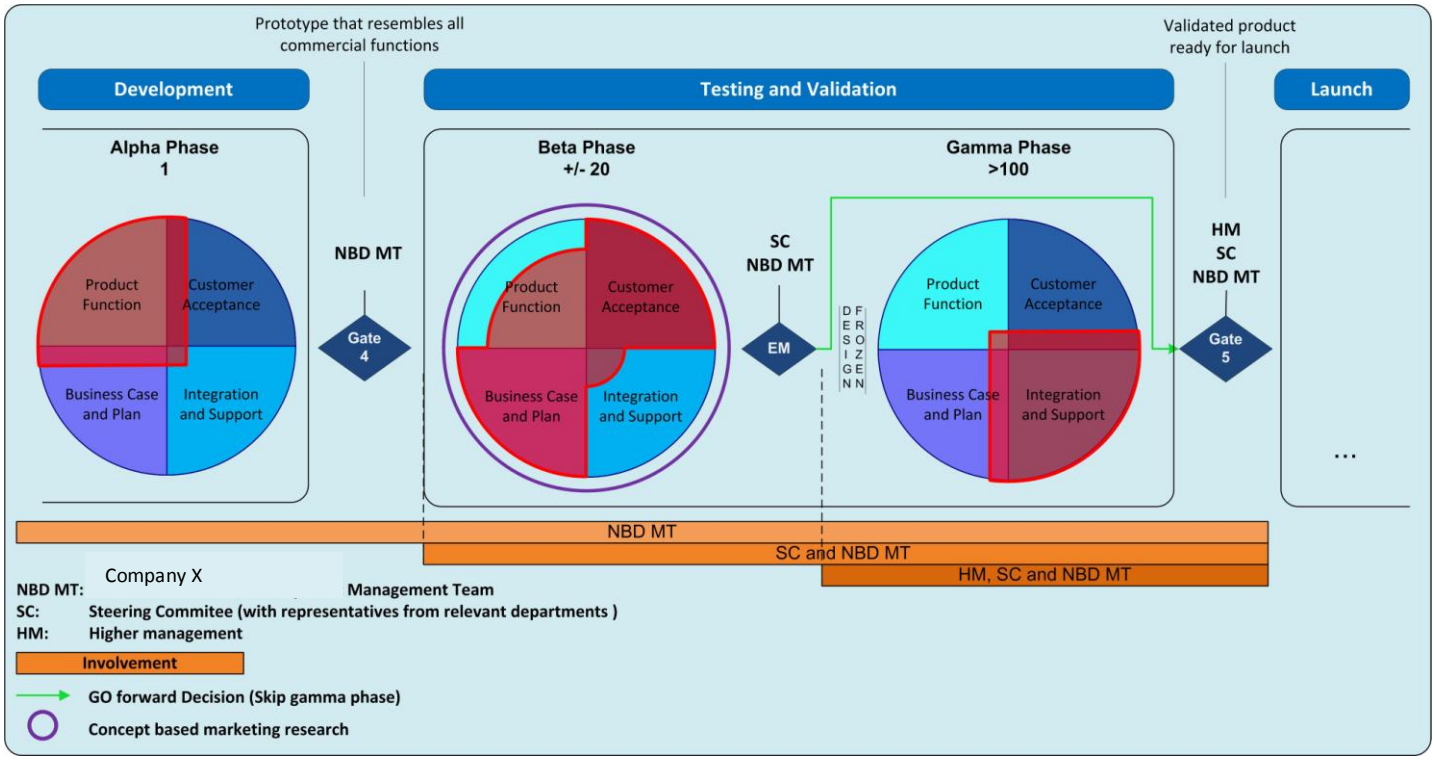
8) *Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators before starting*

In the analysis of this criterion the remarks some respondents had on the necessity of always executing a gamma phase were also addressed. Their main argument was that the benefits of a gamma test, testing and validating the integration and support dimension, can also be achieved in a staged market introduction. Doing so will contribute considerably to the time to market.

Whether or not the gamma phase is necessary, is for a large part a risk management considiration and depends on the complexity and requirements of a product. It was therefore decided to make it optional in the model that is developed. After a beta phase is evaluated there can be decided whether to stop the project, move back to a previous stage, move forward to the gamma phase or move directly to a market introduction.

Adjusted model for Comp. X

Figure 13 reflects the result of combining the adjusted guidelines and implications the practice Comp. X had on the theoretical model. It forms the new testing and validation model for Comp. X.
Just like with the theoretical model there are a few (parts of) guidelines which could not be visualized and who therefore complement the model.
Because the alpha phase was placed in the development stage of the NPD stage gate model figure 13 also encompasses the last part of the development stage. Furthermore the division of importance of the four test dimensions in each phase is made clear by integration figure 9 (alpha phase), figure 10 (beta phase) and figure 11 (gamma phase) into this new model. At gate four the NBD MT is responsible for determining whether the product resembles all commercial functions. Along with the maturity of the product other people from the organization become involved (visualized by the orange “involvement bars”). It can furthermore be seen that the product design only needs to be frozen before the gamma phase and the green arrow in figure 13 reveals that after the evaluation of the beta phase is completed it is also possible to move directly to a staged market introduction.
By combining literature with the practice at Comp. X the model is effective for Comp. X while at the same time based in theory. With establishing this new model the third research question is answered and step three in the research structure (figure 4) is completed.



Complementing Guidelines

- The testing should be done outside the organization in a users' actual environment
- Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience
- Grow the number of testers over time and begin with sophisticated customers with good relations to the firm
- Monitor and evaluate the product performance on a regular basis in an objective way
- Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators before starting

Figure 13: Adjusted testing and validation model for Company X

6. Application of the developed model on the West Orange Project

In paragraph 1.1.2 the West Orange project was identified as an additional motive for this research assignment. The aim of this project is to prove the value of energy management for the consumer market by testing and validating four new products (figure 2). The lack of experience Comp. X has in testing and validating new products, combined with the importance and complexity of this project, stressed the importance of the development of a model to structure this process.
While the complete project comprises 4 products, it was decided to single focus on the energy display at first and introduce the other three products in a later stadium, when the energy displays are up and running. The developed model was therefore only applied on the Product H. Since the product was already successfully beta tested, it was one of the previous executed tests described in chapter 4, it is gamma tested in the West Orange project. The aspects of the developed model that are relevant for a gamma phase are therefore the main point of interest in this chapter.
By analyzing the experiences with applying the developed model, its value and relevance in practice could be determined.

6.1 Research Design, Gamma phase energy display

The first two guidelines of the developed model aim to make sure that a test belongs in the testing and validation stage:

1) *Physical prototypes which resemble the commercial model are used*
2) *The testing should be done outside the organization in a users' actual environment*

This section has not been disclosed due to confidentiality issues

3) *Effective test programs recognize the full set of benefits of all four test dimensions and address all four of them in the testing and validation stage as a whole*

According to the developed model the most important test dimension in a gamma phase is the integration and support dimension. Other test dimensions need to be addressed, but only to verify the results of previous (beta) phases on a larger level and strengthen the assumptions made on these results.

However, for the gamma phase of Product H this line of reasoning does not hold. The previous tests and phases of the product have not been guided by the developed model and not every test dimension has therefore been properly addressed.

This section has not been disclosed due to confidentiality issues

This table has not been disclosed due to confidentiality issues

Table 6: Test categories, gamma phase product H

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

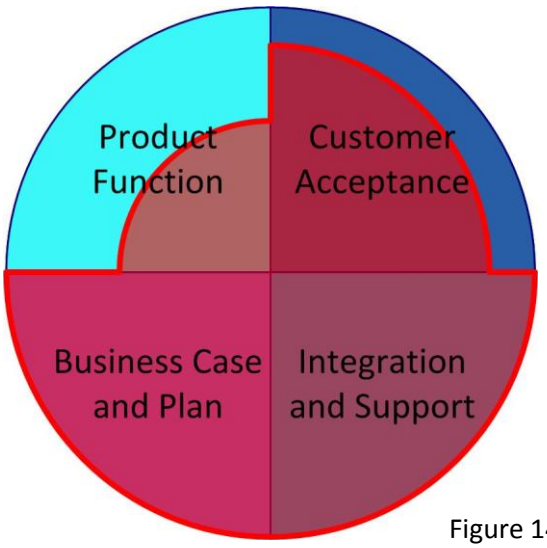


Figure 14: Gamma phase Energy Display

When the learning goals of the test were clear, the next step was to design the test in such a way that the goals of all project partners are going to be met. To do so the other guidelines of the developed model were applied:

4) *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience*

This section has not been disclosed due to confidentiality issues

5) *Grow the number of testers over time and begin with sophisticated customers with good relations to the firm*

This section has not been disclosed due to confidentiality issues

6) Monitor and evaluate the product performance on a regular basis in an objective way

This section has not been disclosed due to confidentiality issues

This table has not been disclosed due to confidentiality issues

Table 7: Data collections gamma phase Energy Display

This section has not been disclosed due to confidentiality issues

This section has not been disclosed due to confidentiality issues

7) Product design should be frozen prior to the gamma phase

This section has not been disclosed due to confidentiality issues

8) *Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators before starting*

This section has not been disclosed due to confidentiality issues

6.2 Learning points from applying the developed model

Although the West Orange project is not yet finished, the use of the developed model to build a research design for an actual project made it possible to determine its value and relevance in practice. Additionally several learning points for the use of the new model were identified.

In general it can be concluded that the application of the developed model was a success. It led to a proper research design which is supported by all organizations part of the project. Furthermore it proved to be a very good tool to develop the research design in a structured way and made sure that all relevant aspects were addressed.
Guideline one and two, on the use of prototypes which resemble the commercial model and testing in a users' actual environment, were applied without any difficulty. This was also the case with guideline 5, a staged roll out plan was designed in which the first participants consist of customers with good relations to the firm.

The relevance of guideline 6, about a regular and objective evaluation was also confirmed. In the developed model more substance was given to this guideline by specifying which organizational groups should be responsible for the evaluation in the three phases. Because only persons from Comp. X were involved in the beta phase of the energy display it turned out to be very difficult to get approval for the execution of the gamma phase. This was a very delicate process as the project was completely new for the majority of the organization. This underlined the importance of involving other organizational groups earlier in the testing and developed process. In the gamma phase the organization is properly involved so the decision process after the test is completed should run much smoother.

While the above guidelines were found both useful and easy to apply there were also a few guidelines which proved to be harder to apply exactly as intended:

Guideline 3: *Effective test programs recognize the full set of benefits of all four test dimensions and address all four of them in the testing and validation stage as a whole*

A part of this guideline is the division of importance, among the four identified test dimensions, over the three phases in the developed model (alpha, beta and gamma). It turned out that the specific focus on the integration and support dimension in the gamma phase is fine, as long as the composition of the organizations participating in the test remain the same throughout the phases. When new organizations join a test in the gamma phase, the division of importance should be reconsidered because it is possible that leaning goals on the other test dimension have not been validated in the alpha or beta phase. This point was proven by the emphasis of the government of Amsterdam on measuring behavioural changes, a customer acceptance aspect which was not yet validated properly. When new organizations become involved in a test their learning goals should be identified without presumptions on the relevance of test dimensions. However, Comp. X should keep the idea behind the division of importance in the alpha, beta and gamma phase in mind and also explain this to new project partners. This way understanding can be created when there is decided not to embed all learning goals of a “new” organization in a research design. This was also done in this test with very specific learning goals on the product function dimension.

Guideline 4: *Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience*

In the previous paragraph it became apparent that it was not possible in the energy display gamma phase, to actively steer towards a segmentation of participants who reflect the market. There were so many practical dependencies that it will be hard enough to actually find 500 households. While this is for a large part related to the complexity of the product and its integration requirements, it can be expected that similar circumstances will accompany future products.
This does not mean that Comp. X should not try to apply a selection process that matches the segmentation of the market, but they should realize that it will not always be possible to follow this guideline completely. In a similar situation Comp. X should estimate what the impact of the selection that is possible is, on the value of the test. Can the intended goals and benefits still be met?
Guideline 7: *Product design should be frozen prior to the gamma phase*

In the developed model the guideline on freezing the product design was already adjusted in its current form so that it only needs to be applied in the gamma phase. Practice with the energy display however showed that Comp. X is not always in a position to determine whether this process is stopped as they do not make the product themselves. Because Comp. X will not be the actual producer in most of the future tests they should interpreted this guideline somewhat different. For Comp. X it is important to make sure that the product that is tested in a gamma phase is the same product that they will bring to the market, when the gamma phase proves successful. When the product is changed fundamentally before going to the market, the results of the gamma phase become irrelevant. The guideline should therefore be interpreted in that sense.

Guideline 8: *Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators before starting*

This guideline was found useful and was fulfilled by discussing the developed research design with the steering committee. However there was an external factor identified which can negatively influence the benefits of the test. Comp. X has already started the preparations for an actual roll out, while the gamma phase is not even up and running. This is an effect of a great emphasis on time to market, but it can have important consequences. When the point is reached that the decision to move towards the market introduction is planned independent from the test results, the learning goals in the test should be reconsidered. In such a scenario the gamma phase is used as a promotion and learning device instead of a testing and validation tool. This means that for example a proper validation of the business case and plan dimension becomes less relevant.
When the decision to go to the market is already made it might be better to steer towards a staged market introduction instead of executing a gamma phase. The developed model accommodates such a decision by making the gamma phase optional.
With the current emphasis on time to market it is possible that a “point of no return” will be reached in the preparations of the actual market introduction in this or a future project. While this has a negative influence on the usefulness of the gamma phase, as indented in the developed model, the importance of the beta phase grows. The outcomes of the beta phase will then form the basis for the decision making on a market introduction. This underlines the value of carrying out a concept based marketing research parallel with the beta phase, as prescribed by the developed model. The higher management has to be convinced based on financial figures and the beta phase on its own is too small to properly validate this dimension.

Main findings

The use of the model in practice showed that even though not all the guidelines could be applied exactly as intended, each one of them proved to be relevant and valuable. It is insurmountable that practical dependencies or pressure from higher management will lead to compromises between an ideal and actual test, but Comp. X should always try to apply the developed model as precisely as possible. When the developed model is used in total, and not just in a gamma phase, the application will also be easier.
There can be concluded that the complete model such as developed in chapter 5 is relevant and valuable for Comp. X. Therefore all aspects are persevered.
However, in order to make sure that the developed model is applied as intended in a project, there is one guideline added to the model:

“Someone should be made specifically responsible for guarding the research requirements in a test design process”

During the development of the research design in the West Orange project it turned out that there is a large chance that research requirements (such as selection criteria) are ignored by a steering committee. This follows from the fact that a steering committee is mainly focused on project speed and handling the practical dependencies. In the West Orange project the choice was almost made to use newly renovated apartments as test households since this would be convenient for the roll out. At that point the steering committee did not oversee the negative complications this would mean for the relevance of the research outcomes. To prevent a similar decision from actually being made in the future, there should always be someone specifically responsible for guarding the research requirements.

7. Conclusions, Contributions, Recommendations, Limitations and Directions for further research.

This final chapter addresses the main conclusions of this study, as well as the contributions to both practitioners and academics involved in the testing and validation of new products. Subsequently recommendations will be given on how to use the findings of this study in practice. This report concludes with the acknowledgement of the main limitations and by providing directions for further research.

7.1 Conclusions

The motive of this research was formed by the implementation of a stage gate model by Comp. X to structure their NPD process. While Comp. X had a clear view on most of the stages embedded in this model, they were uncertain how to execute the testing and validation stage. The need for the development of a model to execute this stage was of extra importance as a large testing and validation project was already scheduled (the West Orange project).
The objective in this research study was therefore to clarify and determine the characteristics of the testing and validation stage and to develop a standard model to execute this stage, suitable for Comp. X. Subsequently the developed model was applied on the West Orange project to build a proper research design and to determine its relevance and value in practice.
To complete this objective, the following main research question was established:

"How should a model to execute the testing and validation stage be structured for Company X?"

Four supporting research questions were derived to answer the main research question:

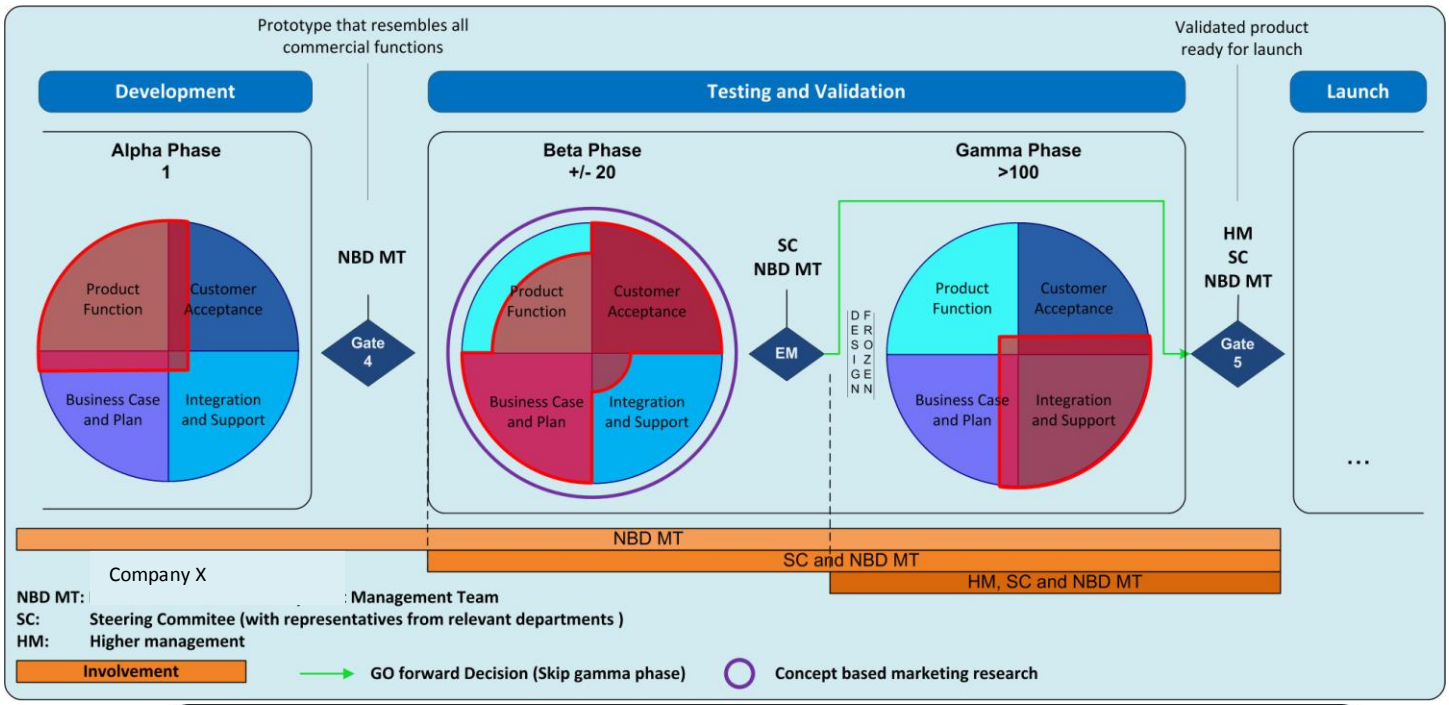
1. How should a model to execute the testing and validation stage be structured according to theory?
2. How does Comp. X currently organize the testing and validation process of new products and what are their design requirements for an execution model of this process?
3. What are the implications of the practice at Comp. X for the testing and validation model which is being developed?
4. How does using the developed testing and validation model in practice affect its value and relevance?

The first question was answered in the theoretical framework. A theoretical execution model was developed by combining the purpose and boundaries of the stage with the overall test dimensions which should be addressed in the stage. Guidelines for designing tests within the stage were further used to specify and complement the model. This model was still relative general because from a theoretical point of view, the testing and validation stage can comprise a variety of testing types and has no unconditional form.
In order to specify and adjust the model for Comp. X the practice of Comp. X was analyzed with the help of; interviews with the employees and management of Comp. X, document analysis and observational research. This analysis had two areas of attention, following from question 2:

- The extent to which the identified key dimensions and guidelines in theory were already applied by Comp. X.
- The design requirements of Comp. X itself

The findings on both points were combined and analyzed together, to adjust the theoretical testing and validation model on. This way an effective model was developed which suited the work methods of Comp. X, while at the same time having its basis in theory. Thereby answering the third question.
The last step in this study was the application of the developed model on the West Orange project to determine its relevance and value in practice, the purpose of the fourth question.
The findings showed that even though not all the aspects could be applied exactly as intended, each one of them proved to be relevant and valuable. It is insurmountable that practical dependencies or pressure from higher management will lead to compromises between an ideal and actual test, but Comp. X should always try to apply the developed model as precisely as possible.
To make sure that the aspects of the developed model are not overlooked or overruled in a project the guideline, "*Someone should be made specifically responsible for guarding the research requirements in a test design process",* was added to the model.

In figure 15 the final model and answer to the main research question is reflected:



Complementing Guidelines

- The testing should be done outside the organization in a users' actual environment
- Maximum utility of a test is derived from recognizing the segmentation of the market and make tests a close approximation to a real world experience
- Grow the number of testers over time and begin with sophisticated customers with good relations to the firm
- Monitor and evaluate the product performance on a regular basis in an objective way
- Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators before starting
- Someone should be made specifically responsible for guarding the research requirements in a test design process

Figure 15: Final testing and validation model for Company X

When looking at figure 15 it can be noted that it does not only consist of the testing and validation stage but also embeds a part of the preceding stage, the development stage. This is a result of the fact that in the analysis of the practice at Comp. X it turned out, that they had several times been testing products with redundant technical problems. By making the single goal of the alpha phase the validation of the product function dimension and placing it before gate 4, this can hopefully be prevented in the future. Aspects from other test dimensions should not be tested in the alpha phase but their importance should be recognized in the form of design requirements. This division of importance over the four test dimensions is reflected by the red section overlying the four test dimensions. Figure 15 furthermore shows that a product should resemble all commercial functions before being allowed to pass gate 4.

The testing and validation stage consists of two phases ascending in size, beta and gamma. Again a division was made between the importance of the four test dimensions. Because the strengths of Comp. X lie in their large customer base and marketing power, this was reflected by the emphasis on the customer acceptance- and business case and plan dimension in the beta phase.
Since the beta phase is too small to properly validate the business case and plan dimension, a concept based marketing research should be carried out parallel (reflected by the purple ring in figure 15).
When the product function dimension, customer acceptance dimension and the business case and plan dimension are validated in the alpha and beta phase, the gamma test can focus on the integration and support dimension. This phase should be used to validate the organizational readiness to effectively install, integrate and support a new product on a large scale with the regular organization. Furthermore the gamma phase can be used to strengthen the findings of previous phases.

Depending on the complexity of a product and its integration requirements, it is also possible to meet the goals of the gamma phase in a staged market introduction. This decision consists for a large part of a risk assessment and should be made after the evaluation of the beta phase is completed. The green arrow in figure 15 reflects this possibility. In general there should be decided after each stage or phase whether to go forward with the project, move back to a previous stage or phase or kill the project.

The complete evaluation process was also made specific in the final model. Relevant departments should be involved from an early point in the testing and validation process and their involvement should grow along with the maturity of a product. By making other people in the organization jointly responsible for the evaluation it will become more objective.
As the practice of Comp. X showed that it is not always realistic to freeze the product design in the beta phase, this should be done in the gamma phase. In this phase the integration of the model in the actual organization is tested which is useless if the product is still being altered.

The guidelines accompanying the model in figure 15 could not be (completely) visualized and therefore complement the model.
When Comp. X follows all elements of the model and the complementing guidelines, the developed products will be adjusted to match the demands of the market and the market introduction will run smoothly.

7.2 Contributions

Contributions for practice

The practical contribution of this research study is evident. As was demonstrated in the West Orange project, the developed model is a good way of structuring a test to validate a new product and leads to a proper research design. The model as a whole furthermore gives substance to the execution of the testing and validation stage and provides guidance for Comp. X to improve their current practice. Recommendations on how Comp. X should implement the developed model, in order to completely utilize its value and make sure it becomes an integrated part of the departments work methods, are provided in the next paragraph.

An additional contribution of the model is that it enables Comp. X to apply an effective portfolio management. Portfolio management is about resource allocation, more specifically about selecting the right set and number of NPD projects. Most businesses have too many development projects underway, and often the wrong ones. That is, they fail to focus, spreading their resources too thinly across too many projects, and their project choices result in the wrong mix and balance of development projects in the portfolio (Cooper *et al.*, 1999).

In the current unstructured way of working at Comp. X it is hard to compare projects to each other. This results in a difficult and subjective portfolio management. The analysis of previous executed test by Comp. X furthermore revealed that the role Comp. X plays in a project is often not adjusted to their goals and potential gains, in other words their resources are not always spread optimal.

When the developed model is applied and the evaluation process is formalized, Comp. X has tools at her disposal to objectively compare products and projects. This enables the NBD MT to actively utilize portfolio management. When doing so they should have four goals in mind (Cooper, 2002):

1. Selecting high-value projects.
2. Achieving the right balance of projects.
3. Selecting the right number of projects.
4. Strategic alignment.

With these goals in mind the NBD MT can allocate their resources in such a way that; poor quality projects are killed, short-term and low-risk projects are balanced with long-term high-risk problems and the project portfolio represents the strategic priorities of the business.

The use of scorecards for projects would improve this additional contribution even more by making projects even more comparable. In which way scorecards should be used, is further explained in the next paragraph.

Contributions for science

An important motive for this research study was the absence at Comp. X of a model to structure the testing and development stage. However also in the literature such a model did not exist. This corresponds with the statement of Ozer (1999) that in the NPD process the testing and validation stage is still a field of research which is less investigated compared to the other stages. Also during this research study it became clear that it was hard to find relevant information on the testing and validation process of new products in literature.

While the absence of a commonly used testing structure complicated the development of the theoretical model to execute the testing and validation stage, it did underline the relevance of this research.

The theoretical model was eventually developed by combining the stage gate methodology of Cooper with literature on test dimensions relevant for the testing and validation stage (Ullrich and Eppinger 2008, Yang and El-Haik 2009, Dolan Matthews 1993, Ozer 1999 and Carbonell-Fouliquie et al. 2004) and literature on guidelines to organize tests within the testing and validation stage (Dolan Matthews 1993, Cole 2002, Crawford and Benedetto 2000, Bonner 1997, Vellandi 2008).
This resulted in a relative general model as from a theoretical point of view, the testing and validation stage can comprise a variety of testing types and has no unconditional form.
The positive side of this generality is that the theoretical model is also of value for other organizations. Such a model did not exist yet in the literature and can therefore be a great help for organizations, who are experiencing problems with the testing and validation of new products. They can choose to use the theoretical model to provide guidance or adjust it to their own work methods, the same way as was done for Comp. X in this study, and actually implement it. However before an organization makes this decision they should realize that the model is based on the stage gate methodology of Cooper (2001) and they should determine whether this methodology suits them.
Well known remarks on this methodology are that the model's emphasis on structure and processes may limit out-of-the box thinking and that the linearity focus on direct market success, neglects indirect effects such as knowledge creation, cyclic innovation processes and the impact on the wider business ecosystem. The model is furthermore based on a rational perspective and it can be argued that it therefore overlooks that organisations are political arenas as well, where less than optimal deals are struck from a financial/business perspective.
While these remarks by far do not overshadow the usefulness of the model for Comp. X, the consideration should be made by each organization independently.
The fact that the theoretical model was build upon the stage gate methodology also has implications for the view academic scholars should have on the model. While the model combined several theoretical findings, it is possible that relevant elements in the testing and validation process are not addressed by the model. Furthermore it remains unknown what the value of the model is for organizations who do not use a stage gate methodology. As these issues lead to directions for further research they are addressed in paragraph 7.4.

The final model developed in this research study is also scientifically important but in a different way than the theoretical model. Because this model was specifically adjusted for Comp. X it is more difficult to apply at other organizations. Some elements are relevant, such as increasing the involvement from organizational groups along with the maturity of the product, but many elements do not suit other organizations. For example the position of Comp. X in the NPD process differs from organizations who actually build the product themselves. Comp. X's main goal is to serve the needs of their large customer base and is therefore much more focused on the customer then an organization who produces a product and uses other organizations as market channels. This difference in context will also result in a different division of importance over the four test dimensions. Furthermore the practice at Comp. X showed that it was not realistic to freeze the product design in each test, because their focus on energy savings and behavioral change requires a long test period. However, exactly this guideline could be very important for other organizations who have no need to test a product for such a long time . It is also questionable whether the alpha, beta and gamma phase division is appropriate for every organization. Therefore it can be concluded that the final testing and validation model as developed for Comp. X, is only useful in its complete form, at organizations with the same characteristics of Comp. X.

While the application possibilities of the final model are fewer then the theoretical model, its scientific relevance in terms of understanding of the actual testing and validation process is higher. The final model was developed to solve the identified problems Comp. X encountered in previous tests and as little is known about the execution of the testing and validation process, insight in these problems is valuable. The developed model can form a starting point in an academic discussion on the reasons for specific problems to emerge and possible ways to solve them. Again this is further discussed in paragraph 7.4.

In general there can be concluded that the research study made a start in filling in a blank spot in literature by establishing a theoretical execution model for the testing and development stage and providing insight in how a theoretical model can be adjusted to fit and improve the testing and validation practice of an organization.

7.3 Recommendations for implementing the model in practice

While the model developed in this research study is a great contribution for Comp. X, as explained in the previous paragraph, it has to be implemented in such a way that it becomes an integrated part of the departments work methods and its value it completely utilized. This section will give substance on how the developed model should be implemented to achieve this.

- *Implement the proposed testing and validation model completely and explicitly hold to the different stages and evaluation moments in the model.*

An important motive for this research assignment was the lack of structure and a “one way of work” in the testing process. Evaluation was often performed on the run which resulted in many redundant problems and defects further on in the process.
If Comp. X truly wishes to improve their way of working they should clearly distinguish and hold on to all of the identified steps.
When a stage or evaluation moment is skipped this will lead to a mismatch between the purpose of a test and the possible outcomes. This is best illustrated by the OT test. The goal of Comp. X in this test was to validate the customer acceptance dimension, but as they did not complete and evaluate the development stage properly many technical problems occurred and their goal could not be met.
Furthermore Comp. X should realize that the proposed model is a part of a larger model, the stage gate model for the entire new product development process they developed themselves.
For the workability of the developed model in this research study, Comp. X should also hold on to each identified step of their own stage gate model.
By doing so the developed product will better fit the market requirements and the time to market will be shortened as less rework is necessary.

- *Formalize the evaluation process*

The current evaluation process turned out to be an important improvement point for Comp. X and it is advisable that Comp. X further formalizes the proposed structure.
In the developed model responsible management teams or committees were identified for each evaluation moment and gate. Assigned groups were:

Gate 4: NBD management team
Evaluation moment between beta and gamma phase: NBD MT and steering committee
Gate 5: NBD MT, steering committee and higher management
While it is obvious who form the NBD management team the other two groups remain a bit vague. Comp. X should specifically identify the persons who will take place in a steering committee beforehand in order to involve them in the design of a test. Since the steering committee consists of representatives from relevant departments the committee will not have the same composition in each project. Relevant departments in any project are Marketing and Product and Service Management (PSM).
Higher management needs to be involved in gamma tests or when a decision has to be made whether or not a product is ready for a market introduction. Depending on the scale of a roll out and the impact is has on the organization this can be the Management team of Sales Netherlands, Sales Benelux or even the management team of Comp. X N.V.
Furthermore it is important to formalize the roll of the evaluators beforehand so that they can already be involved in the design process of a test.
Apart from the evaluators, the gates and evaluation points itself should also be formalized. There should be clear and visible criteria so a go/recycle/kill decision can be made objectively. A possible way to do so, is to make use of scorecards. They are rated highly in terms of effectiveness (make the right decision) and efficiency (in a timely fashion) and fit management's style (Cooper *et al.*, 1999).
Users of the scorecard also indicate that, although the overall project score is useful to prioritize projects, the real value in the scorecard method is the behavioral aspect; the fact that a group of managers meet, discuss the project, walk through a set of key questions, debate the questions, reach closure and then make a decision.
In the developed model the importance of each testing dimension was determined for each phase. By translating the goals underlying these dimensions into requirements a product should meet, a scorecard can be developed. Table 8 reflects an example of such a scorecard for the beta phase on a general level.

Scorecard Beta phase	
"Must Meet" Criteria	**Rating scale**
Product functions properly in a customer environment	Yes / No
The customer "accepts" the product	Yes / No
Approved and positive business case	Yes / No
Identified problems concerning the (future) integration and support	Yes / No
"Should Meet" Criteria	
Product Function:	
- Occurrence of technical problems	1/2/3/4/5
- Easiness of installation	1/2/3/4/5
Customer Acceptance:	
- Customer needs are fulfilled	1/2/3/4/5
- Customer uses the product often	1/2/3/4/5
-Customer behaviour (e.g. energy consumption) is positive influenced by the product	1/2/3/4/5
- Customer would recommend the product to others	1/2/3/4/5

Business case and plan validation:	
- Positive customer case (e.g. savings potential)	1/2/3/4/5
- Willingness to purchase the product	1/2/3/4/5
- Positive cost versus return	1/2/3/4/5
- Proposition matches the customers preference	1/2/3/4/5
Integration and support:	
- Estimated feasibility of integrating the product in the actual organization	1/2/3/4/5
Overall project score	**1/2/3/4/5**

Table 8: Scorecard example beta phase

This scorecard could be used as a starting template, but product requirements should be formulated more specifically in a final scorecard. Furthermore the needed score for a product, to move on to the following phase or stage, should be determined. A scorecard for a project must be finalized and agreed upon, by both the project team and the evaluators, before a test is designed. By making the goals of a test clear to everyone a mismatch with the outcomes can be prevented. Doing so is also a way to fulfill the guideline "Goals and benefits of a test should be clear and agreed upon by project management and responsible evaluators, before starting".

- *Realize organizational commitment in order to successfully implement the model*

In order to successfully implement the model and exploiting a more structural and formal testing and validation process, the department NBD should be clearly informed about the working method and the obtainable benefits. Everybody must be equally informed when using the model and should be convinced about the usefulness. A possible way to obtain this commitment, is the organization of a one day workshop. Employees should be challenged to plot their projects in the model, scorecards drawn up and go/no go criteria discussed. Furthermore the management should demonstrate the added value in terms of time to market, less rework and portfolio management. This will increase the overall employee commitment.

- *Evaluate the new model on a predetermined date with the whole department*

Predetermined evaluation moments of the new model should be a part of the implementation. In any change project it is important to define and celebrate the wins after some months to prevent that the new working method slides to the background and old working habits emerge again (Kotter, 1996). Management needs to prove to the employees that the new model is paying off and discuss with them, which aspects run properly and which aspects still need to be improved.

While this approach to the evaluation aims at making sure that the model is embraced by the organization it is also important to objectively reflect on it. However, this can only be done when at least a couple of projects have moved from development to an actual market introduction guided by the model. The application of the developed model on the West Orange project already proved its value but also showed that it is often hard to exactly apply all elements. Because the model was introduced when the testing and validation process of the energy display was already in the gamma phase, the impact it can have on a complete product testing process could not be determined.

Changing the way people work takes time and an evaluation of the complete model after only a couple of months is therefore not advisable.

When the moment is right to evaluate the complete model, points of analysis should be; fit with market requirements of the new products, rework which was necessary, internal support for the new products, the product portfolio and special attention should be paid to the time to market of products. This last point is important because, while the model is an effective process to control risks and detects redundant problems, it can also slow down the new product development process. The developed model is based on the stage-gate methodology and from a queuing theory perspective, this is a typical batch queue as the many gate reviews can hold a product development team at a standstill for a long time (Yang & El-Haik, 2009). Evaluating the impact of the model on this point should make clear if the model led to standstills of the project team.

7.4 Limitations and Directions for further research

This paragraph reflects on the limitations of the findings in this research study, the issues that were part of and provides suggestions for follow-up research studies.

The first step in this research study consisted of the development of a theoretical model to execute the testing and validation process. By building this model on the stage gate methodology of Cooper (2001) it is possible that relevant elements in the testing and validation process where not identified and or included because they did not fit within this methodology. It would therefore be interesting when further research would analyze the testing and validation process of several successful organizations and explore how their processes relate to the theoretical model. Especially organizations who do not use a stage gate methodology would be valuable to include in this study in order to explore whether there were aspects overlooked.
It would also be interesting to adjust the theoretical model to fit the practice of a completely different organization than Comp. X. The adjustment of the theoretical model to the practice of Comp. X only led to specifications and minor remarks but maybe the practice of another organization will require fundamental changes.

The second step in this research study was the analysis of the current practice at Comp. X. This was done by scoring previous tests Comp. X executed on the aspects of the theoretical model. As the scores were based on interviews with the Comp. X personnel, desk research and (for a minor part) observational research, not all points of view were taken in account. Comp. X often collaborated with other organizations in the previous tests and also embedding their point of view would have given the analysis more substance. The same holds for the experience of actual participants in the previous tests. Neglecting to include these information sources did not have a negative impact on the developed model, as the aim of the scoring was not to criticize or judge the previous tests but to provide insight in how Comp. X is currently working as input for the developed model. However it would have been useful to provide a deeper understanding in the identified problem areas and the reasons for problems to emerge. While the developed model is designed to solve these problem areas, it might be possible that they occur for other reasons than assumed in this study. An example is the occurrence of redundant technical problems which was related to the absence of a proper evaluation between stages. While this is a logic conclusion, it is also possible that Comp. X completely lacks the technical expertise to assess a products functioning. In such a scenario the proposed

solution of an improved and structured evaluation process will not be sufficient to solve the problem. The choice not to include participants of tests and respondents from organizations Comp. X collaborated with, was made purposely based on time restrains. However this is something which would be valuable to change when the research is repeated.
For Comp. X it is therefore important to evaluate the developed model as prescribed in the previous paragraph and make sure that the problems are solved.
For academic scholars the findings can form a starting point in an academic discussion on the reasons for specific problems to emerge and possible ways to solve them. It would be interesting to perform further research on these problem areas both at Comp. X and at other organizations. Are these problems areas common in testing and validation processes, and are they indeed solved by following the developed model, or are they typical for Comp. X?

Specifically for Comp. X an interesting follow-up research would be an analysis of the other stages of their NPD stage-gate model. In the analysis of the current practice of Comp. X it became clear that they did not properly separated the development stage from the testing and validation stage. While Comp. X claimed that the NPD stage-gate model was clear to everyone and only the testing and validation stage remained vague, the analysis showed that this was not the case.
The fact that the NPD stage-gate model as a whole is relative general contributes to the confusion about it. A follow up research with the goal to specify all stages and adjust them for Comp. X could be a great help for Comp. X to improve their work methods.

References

Aldrich, F.K. (2003), Smart Homes: Past, Present and Future.

Babbie, E.R. (2007), The practice *of Social* Research (11th edition*),* Wadsworth, Thomson

Carbonell-Foulquie, P., Munuera-Aleman, J.L. and Rodriquez-Escudero, A.I. (2004), Criteria employed for go/no-go decisions when developing successful highly innovative products, Industrial marketing management, 33, 307-316

Clark, K. and Fujimori, T. (1991), Product Development performance: strategy, organization and management in the world auto industry, Harvard business school press, Boston, Mass

Cole, R.E., (2002) From continuous improvement to continuous innovation. Total Quality Management, vol 13, NO. 8, 1051-1056

Coombs, R., McMeekin, A.and Pybus, R.(1998) Toward the development of benchmarking tools for R&D project management. R&D Management, 28, 3, 175–186.

Cooper, R.G., Edgett, S.J., Kleinschmidt, E.J. (1999). "New Product Portfolio Management; Practices and Performance", *Journal of Product Innovation Management*, 16: 333-351.

Cooper, R.G., (2001) Winning at new products: accelerating the process from idea to launch. Publisher: Basic Books

Cooper, R.G., Edgett, S.J., Kleinschmidt, E.J. (2002), Optimizing the Stage-Gate Process: What best-practice companies do-II, *Industrial research institute, nov/dec pp. 43-49.*

Cooper, R.G. (2008), Perspective: The Stage-Gate Idea-to-Launch Process, Update, What's New, and NextGenSystems, *The journal of Product Innovation Management*

Crawford, M. (1989), Evaluating new products: A system not an act, *Business Horizons, 48-55.*

Crawford, C.M. and Di Benedetto, C.A. (2000), *New Products Management*. 6th edn., Irwin/McGraw-Hill: Boston

Dolan, R.J. and Matthews, J.M. (1993) Maximizing the Utility of Customer Product Testing: Beta Test Design and Management, Journal of product innovation management, 10 318-330

Eppinger, S.D. (2001) Innovation at the speed of information, Harvard business review January 2001, 149 – 158

Elias, A.E., Cavana1, R.Y. and Jackson, L.S. (2002) Stakeholder analysis for R&D project Management, R&D management 32, 4, 301-310

Freeman, E. (1984), Strategic Management: A Stakeholder Approach (Pitman, Boston).

Freeman, E. (2003), Lecture – Stakeholder Management Revisited: What's the State of the Art? Leuven, 20 November.

Gardner, J.R., Rachlin, R. and Sweeny, H.W.A. (1986) - Handbook of Strategic Planning

Griffin, A. (1997) PDMA research on new product development practices: updating trends and benchmarking best practices. *Journal of product innovation management, 14 429-458.*

Gramsbergen-Hoogland, Y.H. and H.T. Van Der Molen (1992). Gesprekken in organisaties. Wolters-Noordhoff: Groningen

Ilmer, V.A. (2009), Consumer-oriented New Product Development; Home automation Technologies, Master Thesis University of Twente

Keizer, J., Halman, J., & Song, M. (2002) From experience: Applying the risk diagnosis methodology. Journal of product innovation management, 19(3) 213-232

Kotter, J.P. (1996) Leading change, MA: Harvard Business School Press, 187p

Miller, M.B., Greenwood, D., Maguire, P. (2003), Why action research?, Action Research vol. 1 No. 9

Minot, M.J. and Wood, L. (2003) New product development in the lean six sigma environment. MIT presentation.

Ozer, M. (1999) A Survey of New Product Evaluation Models, Journal of Product Innovation Management, Vol 16 pp. 77-94.

Sosa, M.E. (2008) A structured approach to predicting and managing technical interactions in software development. Res Eng Design 19, 47–70

Ullrich, K.T. and Eppinger, S.D. (2000), Product design and development, Publisher: Mc Graw-Hill

Yang, K. and El-Haik, B.S. (2009) Design for six sigma, a roadmap for new product development second edition, Publisher: McGraw-Hill

Yin, R.K. (1989) Case study Research: Design and methods (Applied Social Research Methods). London: Sage Publications

Appendixes

Interview Protocol on the Testing and validation process of Comp. X

In het interview wordt onderscheid gemaakt tussen uitgevoerde testen in het verleden en de opzet van toekomstige testen.

Verleden:

1. Bij welke product testen ben je de afgelopen 2 jaar betrokken geweest?
   -Wat hield de test precies in?
   -Hoe groot was de test (hoeveel deelnemers en producten)?
   -Wat was de tijdsduur van de test?

2. Wat waren de belangrijkste onderzoeks/leer doelen in de testen?
   -Waar werd specifiek op getest en onderzoek naar gedaan?
   -Kun je de leerdoelen groeperen in categorieën?

3. Wie waren de deelnemers in de test?
   -Wat was de samenstelling van de deelnemers in de testen?
   -Hoe werden deze geselecteerd?

4. Wie waren verantwoordelijk voor de opzet, uitvoering en evaluatie van de test(en)?

5. Werd er gebruik gemaakt van scorings of evaluatiemodellen in de testen?
   -Welke modellen?
   -Op welke variabelen werd er gescoord?
   -Hoe vaak en wanneer werd er geëvalueerd?

6. Waren er moeilijkheden of problemen tijdens de test?

7. Op basis van welke factoren werd er besloten om al dan niet tot een test over te gaan?
   -Wie maakt deze beslissing?

8. Waren er binnen het totale test proces momenten waarop een go/no go beslissing gemaakt wordt?
   -Zo ja op basis van welke gegevens wordt dit gedaan?
   -Wie doet dit?

9. In welk stadium van het algehele product ontwikkeling proces vielen de test?

10. Hoe verhield het ontwikkelingsproces van het product zich tot de tijdsplanning van de test?

11. Pasten deze testen binnen een bepaalde test methodologie en zo ja welke?

12. Zijn er zaken die je achteraf wel graag had willen weten maar niet op getest hebt?

Toekomst

13. Hoe zou je een nieuwe test opzetten?
    -Zijn er zaken die je anders zou aanpakken?
    -Zijn er valkuilen die je moet vermijden?

14. Op welke testdoelen of dimensies zou je focussen?

15. Hoe groot (deelnemers en producten) zou een nieuwe test moeten zijn?
    -Is dit vaststaand?

16. Hoe moet de evaluatie en het beslissingsproces geregeld worden?
    -Wie doet dit?
    -Op welke momenten?
    -Go/No-Go beslissingen?

17. Zijn er randvoorwaarden waar testen mee te maken hebben?
    -Tijdsduur testen
    -Beschikbare mensen
    -etc

18. Hoe zouden de testen geïncorporeerd moeten worden is het algehele NPD proces?

19. Vind je het een meerwaarde een vaste test methodiek te hebben?

20. Hoe kijk jij aan tegen de alpha, beta, gamma methodiek?
    -Wat is het doel van respectievelijk een alpha, bèta of gamma test?
    -Is het doorlopen van al deze testen bij elk product noodzakelijk?