NTNU

Norwegian University of
Science and Technology

An exploration of mixed methods, researching business model phenomena

Investigating the business model phenomena
through literature, in-depth case study,
statistical testing and user experiments

Henrikke Dybvik

Master of Science in Mechanical Engineering
Submission date: October 2018
Supervisor: Martin Steinert, MTP

Norwegian University of Science and Technology
Department of Mechanical and Industrial Engineering

Henrikke Dybvik

An exploration of mixed methods, researching business model phenomena

Investigating the business model phenomena through literature, in-depth case study, statistical testing and user experiments

Master's Thesis in Mechanical Engineering

Trondheim, October, 18

Supervisor: Ph.D. Martin Steinert

Norwegian University of Science and Technology

Faculty of Engineering Science and Technology

Department of Engineering Design and Materials

Abstract

This thesis explores the use of a mixed method research approach to investigate a complex phenomena – the business model. Qualitative work is more common in business model research, however quantitative work is called upon as qualitative methods alone cannot capture all levels of the business model phenomena. A mixed method approach utilizes both qualitative and quantitative approaches, it can strengthen confidence in results and it is recommended for researching complex phenomena.

The methods selected for exploring and mixing are literature, in-depth case study, statistical testing and user experiments. The methods were selected with a training intention ion in mind, notably to create the skillset of a researcher and prepare for a PhD. Usage, comparison and discussion of the different methodological approaches is the focus of this thesis. The results are four papers.

Paper 1 encompass a literature investigation and could aid in understanding the business model phenomena on an overarching process level and how these processes could be improved. Paper 2 encompass an in-depth case study and could aid in understanding the business model phenomena on an organizational and network level. Paper 3 encompass statistical testing of company data and could aid in understanding the business model phenomena on a company level and in-between companies. Paper 4 encompass a user experiment and could aid in understanding the business model phenomena at a product level.

Conducting multiple methods poses great requirement on a researcher. It may be next to impossible for one single researcher to conduct the methods properly. Academic rigor should not be compromised. Therefore, the handover between the methodological approaches and the specific methods must be appropriate. Complex, uninvestigated phenomena and context require a genuine mixed method approach, which research designs with a solitary purpose fall short on. A deliberate selection of methods combined with applying the flexibility principle to research design results in a research design that is not preset, but that adapts to the research process as it progresses. A flexible research design can further enhance the mixed method research approach's ability to investigate broad and complex phenomena such as the context in this thesis.

Sammendrag

Denne masteroppgaven utforsker bruken av kombinerte forskningsmetoder (også kalt mixed methods) for å undersøke et komplekst fenomen - forretningsmodellen. Kvalitative metoder er mer vanlig i forskning på forretningsmodellen, men kvantitative metoder etterspørres, da kvalitative metoder alene ikke fanger alle nivåer av forretningsmodellfenomenet. En kombinert forskningsmetode benytter både kvalitativt og kvantitativt forskningsarbeid, det styrker tilliten til resultater, og er anbefalt til å undersøke komplekse fenomener.

Metodene som er valgt for å utforske og mikse er litteraturstudie, grundig casestudie, statistisk testing og brukereksperimenter. Metodene ble valgt med en opplæringsintensjon, spesielt for å skape ferdighetene til en forsker i forberedelser på en doktorgrad. Bruken, sammenligning og diskusjon av de ulike metodologiske tilnærmingene er fokuset i denne masteroppgaven. Resultatene er fire artikler.

Artikkel 1 omfatter en litteraturstudie og kan bidra til å forstå forretningsmodellfenomenet på et overordnet prosessnivå og hvordan disse prosessene kan forbedres. Artikkel 2 omfatter en grundig casestudie og kan bidra til å forstå forretningsmodellfenomenet på organisasjons- og nettverksnivå. Artikkel 3 omfatter statistisk testing av bedriftsdata og kan bidra til å forstå forretningsmodellfenomenet på bedriftsnivå og mellom bedrifter. Artikkel 4 omfatter et brukereksperiment og kan bidra til å forstå forretningsmodellfenomenet på et produktnivå.

Å gjennomføre flere metoder stiller store krav til forskeren. Det kan være nesten umulig for en enkelt forsker å gjennomføre metodene riktig. Akademisk korrekthet bør ikke kompromitteres. Derfor må overlevering mellom metodikker og spesifikke metoder være hensiktsmessig. Komplekse, uutforskede fenomener og kontekst krever ekte kombinert forskningsmetode, og her kommer forskningsdesign med smale formål til kort. Et bevisst utvalg av metoder kombinert med å anvende et fleksibilitetsprinsipp i forskningsdesign resulterer i et forskningsdesign som ikke er forhåndsbestemt, men som tilpasser seg forskningsprosessens utvikling. Et fleksibelt forskningsdesign kan ytterligere forbedre kombinerte forskningsmetoders evne til å undersøke brede og komplekse fenomener, som konteksten i denne masteroppgaven.

Acknowledgements

For this template, thank you Jørgen.

For support throughout a labor-intensive adventure. Thank you TrollLABS and trolls. Thank you friends and family. Thank you Martin.

Preface

The project has been conducted at the Norwegian University of Science and Technology (NTNU), Department of Mechanical and Industrial Engineering (MTP) in 2018. It is a part of the research conducted at TrollLABS.

The master thesis is written in a cumulative style, encompassing four papers.

Although stated in the title of the thesis, the author would like to emphasize that this is a method exploration. The methods were selected to create the skillset of a researcher and prepare for a PhD. Therefore, the specific contextual background is not consistent throughout the articles, and thus the overall thesis composition. Usage, comparison and discussion of the different methodological approaches is the focus.

The research context for the individual papers are related to the course TMM4280 - Advanced Product Development conducted at NTNU, relations and interests in the hardware startup environment, the PhD of Yngve Dahle and the PhD of Andreas Simskar Wulvik.

Trondheim, October 2018.

Table of contents

List of Figures

Figures for each paper are indicated with relative page numbers.

Paper 1:

No figures.

Paper 2:

Paper 3:

No figures.

Paper 4:

List of Tables

Tables for each paper are indicated with relative page number.

Paper 1:

Paper 2:

Paper 3:

Paper 4:

Abbreviations

BM = Business Model

KPI = Key Performance Indicators

PD = Product Development

NPD = New Product Development

TPDS = Toyota Product Development System

TPS = Toyota Production System

BMC = Business Model Canvas

GSR = Galvanic Skin Response

ECG = Electrocardiography

PSS = Product-Service System

1 Introduction

1.1 Problem Statement (Thesis Scope)

This thesis explores the use of a mixed method research approach to investigate a complex phenomena – the business model. Qualitative work is more common in business model research, however quantitative work is called upon as qualitative methods alone cannot capture all levels. A mixed method approach utilizes both qualitative and quantitative approaches, it is claimed to strengthen confidence in results and it is recommended for researching complex phenomena.

The methods selected for mixed method exploration are literature, in-depth case study, statistical testing and user experiments. The methods were selected to create the skillset of a researcher and prepare for a PhD. The task is to train and gain firsthand experience with the methods. Usage, comparison and discussion of the different methodological approaches is the focus. Therefore, the specific contextual background is not consistent throughout the overall thesis composition.

1.2 Readers guide (Thesis Organization)

The thesis is written in a cumulative manner, encompassing four papers. Each paper explores the business model phenomena using different research approaches to investigate the different levels of the business model. The papers are listed in order of appearance, which corresponds to the level of investigation, starting broadly on a top level, before going into more depth and detail.

Paper 1:

The first paper is "Dybvik, H., Erichsen, J. A. B., Steinert, M., & Welo, T. (2018, June). Evaluating Continuous Improvement Efforts in New Product Development. In ISPIM Innovation Symposium (pp. 1-14). The International Society for Professional Innovation Management (ISPIM).". It was presented at ISPIM Innovation conference, "Innovation – The Name of the Game", in June 2018.

It encompasses a literature investigation of means to measure continuous improvement of processes and it proposes possible metrics for making such assessments with lean principles as the foundation. Furthermore, the paper includes a discussion regarding the effectiveness of such methods and the proposed metrics. To improve an entity, one must first be able to

measure the current state of the entity. This is also true for a business model, the business model development process, or any subset of it. This could aid in understanding the business model phenomena on an overarching process (or company) level and how these processes could be improved.

Paper 2:

The second paper is; "Makerspace sustainability enabled by crucial partnerships and a gentrification strategy". It is a draft to be submitted to Journal of a Cleaner Production.

It encompasses a qualitative in-depth case study investigating the business model of a hardware centered startup incubator. It describes a set of crucial partnerships in a larger, complex ecosystem consisting of many distinct stakeholders necessary for business model sustainability. This could aid in understanding the business model phenomena from an organizational level by looking at the incubator's internal organization and from an network or ecosystem level by looking at the network surrounding the incubator.

Paper 3:

The third paper is; "A Dynamic and a Static Approach to the Business Model - Investigating the potential difference in business model focus". It is submitted to ICE-Conference 2019.

It encompasses a quantitative research approach where quantitative data from two independent samples (in total 34 companies) were statistically tested for differences. The context is lean business model development, where a dynamic approach have been compared to a static approach to using the business model canvas tool. This could aid in understanding the business model and business model development on a company level or in-between companies.

Paper 4:

The forth paper is "Dybvik, H., Wulvik, A., & Steinert, M. (2018). STEERING A SHIP-INVESTIGATING AFFECTIVE STATE AND WORKLOAD IN SHIP SIMULATIONS. In DS92: Proceedings of the DESIGN 2018 15th International Design Conference (pp. 2003-2014).". It was presented at the 15th International Design Conference, "DESIGN 2018" in May 2018.

It encompasses a description of how an experiment can be set up, conducted and used to investigate how one could start to design and develop a successful product. The context is maritime, where the working conditions for ship captains are tested to provide foundation

for subsequent design and development of ship bridges. This could aid in understanding the business model phenomena on a product level, how the company could design and develop offerings delivering superior value to the customer enabled by true understanding of user needs and requirements.

2 Background of mixed methods

A general introduction to the mixed methods is given here. It includes a definition of a mixed methods research approach, known advantages and limitations, its use in terms of domain, research design and reporting, how to establish validity and some issues.

2.1 Mixed method research approach

Multiple terms refer to the mixed methods research approach (Creswell, 1999; Creswell and Clark, 2007), for example multimethod research, integrating qualitative and quantitative approaches, methodological triangulation and multimethodological research. The author sees these terms as appropriate as their central idea is the same, combining or integrating different methods in either a series of - or a single research study.

Creswell (1999) defines a mixed-method study as one incorporating both qualitative and quantitative methods of data collection and analysis in a single study. Here, at least one qualitative and one quantitative method must be used to collect, analyze and report results. Creswell and Clark (2007) define mixed methods research as a research design with philosophical assumptions and methods of inquiry. Philosophical assumptions guide the direction of data collection and analysis from, and the mixing of, both qualitative and quantitative approaches. Mixed methods can therefore be thought of and treated as a methodology and a method. These notions are employed as an appropriate definition for the purpose of this thesis.

Equally important as the definition of mixed methods is its core assumption, which is that a combination of methods rather than a single one will provide a better understanding of a topic (Creswell and Clark, 2007). By building upon the inherent strengths and limitations of the methods, mixing methods may uncover unique variance which may have been neglected by applying one single method (Abowitz and Toole, 2010; Creswell, 1999; Creswell and Clark, 2007; Morgan, 1998). The purpose with mixing methods is corroboration; searching for consistencies in findings, elaboration; enhancing the result from one method to the other, and development; efforts to inform one methods from the other (Creswell, 1999). If applied with success it would enhance validity and reliability of results and it could produce more than a monomethod applied to the same issue (Morgan, 1998).

2.2 Qualitative and quantitative research exhibit distinct characteristics

Qualitative research approaches are designed to explore the human elements of a given topic, typically addressing new phenomena or analyzing a phenomena in retrospection (Given, 2008). Explanations and meaning are sought after by capturing information regarding individuals' thoughts, feelings, or interpretations of a topic, usually in a narrative or descriptive form. Data collection methods associated with qualitative work typically includes field work such as; observation and participant observation, interviews, key informants, multimedia documentation, site mapping and more. Notes and documents, audio, social media, recordings, documents (historical record, financial record, news, internal reports, financial data etc.) are used to store information. The instrument of data collection here is the researcher, aided by their selection of modern technology (Creswell, 1999).

Quantitative research *"refers to approaches to empirical inquiry that collect, analyze, and display data in numerical rather than narrative form"* (Given, 2008). The scientific method is generally closely associated with quantitative work, more so than with qualitative. Quantitative research also gathers information, though the key distinction from qualitative is that quantifiable or numerical data is collected through an actual instrument or tool different from the researchers. Here, data collection methods includes the following; survey data, information collected through an experiment or quasi-experiment, historical or financial data, data collected through sensors and other means of capturing (Creswell, 1999). Analytical techniques includes statistical testing, mathematical and computational modeling.

To conduct research using a mixed method approach a researcher or research group must hold competencies in both qualitative and quantitative methods (Creswell, 1999; Morgan, 1998). The knowledge base providing the research's theoretical grounding must consist of multidisciplinary domains, as must the skillset required for conducting both qualitative and quantitative research. Though qualitative and quantitative research is not mutually exclusive, they do exhibit significantly different characteristics (Given, 2008). Naturally, this bring challenges in integrating multidisciplinary, but also opportunities (Abowitz and Toole, 2010; Creswell, 1999; Morgan, 1998; Östlund et al., 2011).

2.3 Advantages in using a mixed methods approach

The main argument for selecting and using a mixed methods research approach is complementary of methods: to utilize and take advantage of the strengths of both quantitative and qualitative methods, and cover the limitations each method holds by the other, using the strengths of one method to enhance the performance of another (Abowitz and Toole, 2010; Morgan, 1998; Östlund et al., 2011). Results regarding the same research question are cross-validated (or converged) by using multiple methods, thereby demonstrating that the result are not simply due to an artifact or invalidity inherent with the particular method (Morgan, 1998; Östlund et al., 2011). This could produce more than a monomethod (Morgan, 1998) applied to the same research question. Benefits with a successful application of mixed methods include increased reliability and validity of data, and greater confidence in hypotheses testing. This allows a greater confidence in results or conclusions, whether those display data convergence or divergence (Abowitz and Toole, 2010).

2.4 What is mixed method used for?

Mixed methods has been suggested for investigating complex issues (Morgan, 1998; Östlund et al., 2011) of for example social phenomena (Creswell, 1999) or topics in a scientific field that involves human behavior or actions (Abowitz and Toole, 2010). This is because it allows for understanding of a complex phenomena on a qualitative level as well as quantitative, through numbers, metrics and measurements (Creswell, 1999). It holds potential for better understanding of and response to multiple stakeholders as it can see the world trough (or with) multiple lenses (Creswell, 1999) when compared to a monomethod approach.

Mixed methods is essentially seen as a research method for the social sciences (Abowitz and Toole, 2010; Creswell, 1999; Creswell and Clark, 2007; Östlund et al., 2011), but it is also being used in other disciplines. It's use in health research (Morgan, 1998), due to the complexity of the many different factors that influence health and the magnitude of such factors. It is also used in decision making processes, where it provides research information assisting policy making due to mixed methods addressing the need for high technical quality and comprehensiveness (Creswell, 1999). Examples of mixed methods use in scientific fields where human behavior or actions are involved (Abowitz and Toole, 2010),

include construction research (Abowitz and Toole, 2010) and engineering design research (Jensen et al., 2016).

2.5 Limitations of a mixed method research approach

Mixing methods typically costs more than either qualitative or quantitative, it requires more time and energy, - both for the researchers and the research subjects (Creswell and Clark, 2007; Onwuegbuzie and Teddlie, 2003). The main cost related to a methods is typically not related to the same stage across methods (Abowitz and Toole, 2010). It may also differ within methods depending on the context and framing of research questions. The researcher or group of researchers need to have a broader skillset. One of the main arguments against using mixed methods is that it is not practical in terms of research design (Morgan, 1998), due to the technical challenge in creating an effective combination of qualitative and quantitative methods and potential conflicts in paradigm (quantitative and qualitative methods have different assumptions about the nature of knowledge and the appropriate means of generating knowledge). Some details of mixed method research remain to be fully worked out by research methodologists. This could cause an additional cause-and-effect chain since methodological purists contend to one approach (Onwuegbuzie and Teddlie, 2003) and may therefore not want to do this work. An additional challenge brought by mixing methods is to ensure validity across the methods.

2.6 Research design in mixed methods

For an effective application of mixed methods research, much emphasis is placed on research design and several strategies and procedures have been outlined (Creswell, 1999; Morgan, 1998) as well as classifications (Creswell and Clark, 2007) and models describing where method integration occurs. Some of these are described below.

Creswell (1999) suggest basic procedures for designing a mixed method study. Firstly, one must determine if mixed methods is needed and feasible. Having selected mixed methods, qualitative and quantitative research questions should be formed, followed by selecting qualitative and quantitative data collection types. Relative weight and implementation strategy for each method should be assessed and one may present a visual model of this. Further, data analysis and quality assessment of the study must be decided. A plan guiding the researcher through the mixed method study can then be developed based on this information.

Strategies underpinning model design are often easier to identify or decide on by using simple visual models, as advocated by mixed methods researchers (Creswell, 1999; Morgan, 1998). Creswell (1999) describes three models. In *the convergence model*; quantitative and qualitative data are collected first, then both data sets are examined and analyzed to determine findings. In *the sequential model*, the qualitative and quantitative approaches are used in a sequence where the second builds on or extends what was found in the first method. *The instrument-building model* begins with an explorative qualitative method, collecting and analyzing data, then uses this information to form quantitative questions. This instrument development should appropriately reflect views of the people who will use the instrument later on.

The priority decision assume priority must be given to either the qualitative or the quantitative approach based on the overarching motivations for the study. A complementary method should be selected based on it offering a set of strengths adding to the research designs ability to meet the overarching goals of the study.

The sequence decision mainly concerns itself with the order or sequence of the qualitative and the quantitative method. A sequence should connect the different types of information in a way that maximizes their contributions to the overall research purpose. The key consideration to make is regarding whether a complementary method should come first as a preliminary input to the principal method, or second as a follow-up to the principal method. This naturally leads to four basic design: qual → QUANT, quant → QUAL, QUANT → qual, and QUAL → quant, where the principal method is in capital and complementary method in lowercase letters (Morgan, 1998).

Both the priority and sequence decision model deliberately omits the option of giving the qualitative and quantitative method equal weight and applying them simultaneously. This is due to introducing a problem of analyzing the combination of data, which may require a third method to connect insights, and results might be contradictory. Furthermore, there is the question of who will do the work, and which research will hold the necessary skills required for qualitative and quantitative methods. Additional arguments against *simultaneous use* include the logistics of coordinating insights and the timeline the methods operate on – which is inherently different. Therefore, simultaneous research design are more often not chosen, due to priority-sequence models being more practical in that they are easier to implement and it is argued that they lead to more productive combinations of

qualitative and quantitative data (Morgan, 1998). It may be easier to anticipate a research outcome from using *priority* or *sequence* models. Despite challenges related to this approach, Morgan (1998) argues for attempting to find a more practical and effective research design for this approach as this could enable 'true triangulation'.

Analytical strategies specifically for analyzing the data in mixed methods research have been identified by Onwuegbuzie and Teddlie (2003), and they describe three approaches. *Concurrent data analysis* integrates each data set during the analysis stage to provide a complete picture developed based on both data sets after them being qualitised or quantitised. *Parallel data analysis* consolidates and compare findings at the interpretation stage, after separate collection of data and data analysis. *Sequential data analysis* analyzes data in a particular sequence with the purpose of informing the other method, rather than a direct integration.

The various models outlined above all advocates a carefully and tailored plan and research model. The original plan may not be the same as the actual model described in the final writeup of results, which is the case in some mixed methods studies (Creswell, 1999). Mixed method researchers are aware of that one might have to conduct such changes (Creswell, 1999; Morgan, 1998), though it is no placed much emphasis on this in the literature reviewed here.

2.7 Establishing validity in mixed methods

Establishing validity in mixed methods research requires at least two distinct procedures for validity due to the distinct nature of the qualitative and quantitative method. The researchers viewpoint for establishing validity in a study differs (Creswell and Miller, 2000).

Quantitative researchers' viewpoint is based of instruments, scores or measurements, criteria, validity of research design such as survey, experimental or quasi-experimental designs and more. Validity procedures in quantitative research include assessing internal validity and external validity, whereas reliability is seen as an instrument property. Internal validity can be assessed by considering the measurement instruments' concurrent validity, predictive validity and construct validity. External validity or generalizability is established by statistical procedures (Given, 2008).

Qualitative researchers' viewpoint relies on the scientific and personal views of the humans conducting, participating in, and review research. Validity procedures in qualitative research include rigorous methods and systematic forms for inquiry (some actively look for quantitative equivalence and use those procedures for validity), trustworthiness and authenticity, questioning assumptions and researchers reflexivity (Creswell and Miller, 2000). Here, validity depends on the researcher perspectives to a higher degree in that the researcher decides how much time to use, whether data is saturated and how data analysis evolve. Researchers have to return to their data in an iterative manner, to compare and corroborate constructs, explanations and interpretations (Given, 2008). Combining qualitative viewpoints and paradigms results in numerous validity procedures. Creswell and Miller (2000) have arranged these and a few are mentioned here; a) triangulation across data sources, theories, methods, and among different investigators, b) searching for disconfirming evidence and c) member checking; taking data and interpretations back to the study participants to confirm information and narrative credibility.

Suggestions for mixed methods validity are based on combining validity procedures from qualitative and quantitative methods. Particular emphasis is placed on various forms of triangulation by many (Creswell, 1999; Creswell and Clark, 2007; Johnson and Onwuegbuzie, 2004; Östlund et al., 2011). Östlund et al. (2011) emphasizes triangulation and illustrate it on an overarching level, that is, between theoretical proportions, quantitative empirical findings and qualitative empirical findings. Examples of different use of this triangulation form are; triangulating complementary, convergent and divergent findings. Abrowitz and Toole (2010) on the other hand emphasize the procedure to ensure validity in mixed methods. This should include proper research planning and design, explicit definition and operationalization of theoretical concepts, explicit statement of hypothesized causal relationships and appropriate statistical analysis.

2.8 Reporting mixed methods

Reporting mixed methods studies have been described as a challenge (Östlund et al., 2011), much due to the same issues and challenges described above. Communication issues between qualitative and quantitative researchers are also common (Morgan, 1998). Since mixed methods require effective communication between the qualitative and quantitative field, emphasis should be put on clear communication of protocol and results. Clarity, transparency and a consistent language with proper definition could aid here. Clarity and

transparency are crucial in reporting (Östlund et al., 2011), and should include disclosure of the researchers own bias and interpretation, including paradigm and point of view (Creswell and Miller, 2000).

2.9 Issues in using a mixed methods research approach that will be addressed in this thesis

It is not given that the four sequence models described are the only or even the best ways to combine methods – this depends on the goal and research question of a project (Morgan, 1998) as well as the context to which it is applied. The mixed methods field is moving towards practical applications of mixed methods (Östlund et al., 2011), developing research designs (Morgan, 1998) that better analysis and integration (Östlund et al., 2011) when mixing methods. Despite challenges related to this, Morgan (1998) argues for attempting to find a more practical and effective research design for *priority* and *sequence* models as this could enable 'true triangulation'. The author believes it is important to be aware and attentive of the possibility for models and research designs to change or adapt, the discussion of this thesis further addresses this.

3 Contextual background - Business Model Phenomena

A general introduction to the context, namely the business model phenomena is given here. Theoretical grounding of the research context will be provided before describing research gaps and some related current trends.

3.1 Providing background to business model research

The business model phenomena as a research topic is relatively new, in that the term gained traction and was used in research articles from 2005 and out, having received limited attention up until then (Morris et al., 2005). Research on the 'business model', was further fueled by a special issue on the concept in Long Range Planning in 2010. Investigation of the phenomena can be conducted on multiple levels.

3.2 A general introduction to the business model phenomena

On a fundamental level, business model research is set to answer two broad questions: "What are business models? How are they used?" (Bocken et al., 2014; George and Bock, 2011; Morris et al., 2005; Zott and Amit, 2010). These broad questions seem to have straightforward answers, - at the very least on a superficial level. A business model seeks to capture, that is understand and describe, the fundamentals of how a company does business. It is used for initial creation and subsequent development of the company's core business. However, there are many and different definitions, research perspectives and focus areas, and problems displayed by gaps in research. This ambiguity and inconsistency suggest that the answers may not be so simple, the answers may even be insufficient at their current state.

The broad statement of the two questions posed initially display the two different uses of the concept business model, - the static approach and the transformational approach (Demil and Lecocq, 2010). The static approach emphasizes the noun 'model', a picture or a recipe describing and classifying the activities used by the company to create value, and the functioning and logic behind the value creation. The transformational or dynamic approach emphasize the verb 'model', the act of modeling by using the business model as a tool for change and development in the organization or in the business model itself. The latter use, developing new business models have been celebrated as a means to innovate and shift industry perspectives.

3.3 The inconsistent definition of the business model

There is a common understanding of the business model as a whole. Namely a systemic and holistic description of 'how a firm does business' (Bocken et al., 2014; Zott and Amit, 2010). Despite this, there is no consensus on one single standing definition (Bocken et al., 2014; Morris et al., 2005). Descriptions of business model definitions and perspectives are fragmented and inconsistent (Bocken et al., 2014; George and Bock, 2011). Several researchers have pointed towards the need for a common terminology, definition and understanding of the business model to facilitate future business model research (George and Bock, 2011; Morris et al., 2005; Ritter and Lettl, 2018).

3.4 The business model perspectives find theoretical grounding in several research domains

Theoretical grounding of the concept 'business model' can be found in multiple research domains. These focus for example on the organizational side, activities and strategy within the business itself, while others focus on the social aspect, including the human resources and capabilities. A few are highlighted here.

In a resource-based view, the resources and capabilities of a company are exploited to achieve sustainable competitive advantages. This is if the resources and/or capabilities are particularly valuable, rare, hard to imitate, and not substitutable. They could be bundles of tangible and intangible assets, such as management skills, technological skills organizational processes, skilled employees, product or technology or other information and knowledge (Barney et al., 2001; Wernerfelt, 1984). Thus, by focusing on the unique resources a company with resource-based view would alter other elements of its business to maintain or achieve a competitive advantage. An example is altering market, customer or industry to achieve a product-market fit.

The opposing view is thought of to be the demand-side perspective, also labeled as market-based-view, positioning or market-positioning-view. Here, the focus lies outside the company. Orientation towards customer and market, viewing their preferences and needs as dynamic and sometimes latent guides management. Managerial decisions can and should be made pertinently in line with that, as these decisions cause crucial value generation for the company. This is providing the company competitive advantages (Priem et al., 2012). Thus, by focusing on the customer needs or market gap, a company would alter other

elements of its business to fulfill customer need and sustaining a position in the market. An example is adhering to the customer needs by altering product or technological solution.

Effectuation (Sarasvathy, 2001), deliberate practice (Baron and Henry, 2010) and the socio-cognitive capabilities points toward the behavior of those developing a company. Traces of a resource-based-view can be found, however the focus of these works lies with the cognitive capabilities and the entrepreneurs ability to develop these. As a consequence the goal of the entrepreneur, the company and the business model adapts according to a range of factors as the business development process progresses.

Dynamic capabilities, as Sarasvati (2001) describe, is also studied on a company level. For example, Ausrød et al. (Ausrød et al., 2017) conducts a case study on business model development. Here, the shift from one initial business model to a second and third, increasingly advanced business models displays how a company developed their dynamic capabilities.

Ritter and Lettl (2018) present a range of theoretical frameworks and perspectives in a structured way to facilitate a broad discussion. Five perspectives are presented, namely activities, logics, archetypes, elements and alignment. A focus on business activities and their systematic composition and interactions is latent in *activities*. *Logic* focus on why and which activities to conduct based on the value-creation logic. *Archetypes* as models of value creation focus on where to locate and how to design the revenue streams for the economic gains. *Elements* may be the most known perspective, where the structural elements of the business model and their interactions are displayed. Alignment of the strategy and the business elements, particularly resources and management is the focus of the fifth perspective (Ritter and Lettl, 2018; Zott and Amit, 2010). These perspectives are complementary and overlapping. Therefore, the business model has been suggested to act as a semipermeable membrane between theories, providing a valuable connection between them.

3.5 Business model research streams

As a field of research progresses new streams are introduced. Here, two research streams that are considered to be relevant today are presented.

3.5.1 Sustainable business models and their development is as current at sustainability

The notion of sustainability is gaining traction on a global scale, in industries, individual companies and academic research. The UN sustainable development goals are increasingly being used as research challenges for industry and academia and environmental goals are set by nations.

A reflection of the static and dynamic approach described initially naturally appear as sustainability is introduced to the business model phenomena. Sustainable business models are not straightforward. Neither is sustainable business model development and they have both been described as a challenge for industry and academia.

Sustainable business models are obviously an advantage assuming a company want to exist over a longer time period. The expansion of a service economy can be observed on a global scale. There is a need for new jobs exhibiting distinct characteristics and new technological solutions must be developed. These changes in market and customer, including the shift towards collaborative (Hamari et al., 2016) and experience economies (Tukker, 2015, 2004) causes market requirements and customer needs to change. Capturing intangible value poses greater challenges for value creation, challenging companies to innovate by developing business models, products and services. Business model development deliberately changes the core elements and activities of a company (Ritter and Lettl, 2018). The changes mentioned calls for a non-traditional value creation logic, which could be facilitated by changing company activities and core elements. Therefore, the *development* of the business model could an important advantage for a company. Further, a strong focus is advantageous when developing novel solutions and since Morris (2005) states that the business model can serve as a focusing device it can assist in this endeavor.

3.5.2 Lean perspectives are current

The term lean originated to describe the manufacturing process and the product development process at Toyota, namely the Toyota Production System and the Toyota Product Development System respectively (Morgan and Liker, 2006). Lean is adopted because it is highly recognized to increase the ability to rapidly introduce innovative products to the market. Lean principles can aid companies in an increasingly competitive marketplace in that it provides certain competitive advancements. These are related to understanding and maximizing customer value while reducing activities and inventory seen

as waste (Browning, 2003; Welo, 2011). Lean is seen as a socio-technical system that integrates people, process, tools and technology. To improve customer value and better this integration a company should continuously improve by learning, changing and developing the company processes (Morgan and Liker, 2006; Welo, 2011).

Lean principles have been extended from only manufacturing to the actual product offering by applying it to the product development phase. Furthermore, lean principles have been extended to the business model phenomenon in several ways. Several large corporations and consultancy companies have started to train and certify employees in lean principles to be able to offer such knowledge in the work they conduct for a customer. The lean mindset could provide the foundation for an entire business model, or it could be the strategic underpinning of one part of the business model. Entrepreneurs and startups have adopted the principles through the lean startup movement, where the lean mindset is utilized as a tool for business model development and product offerings development. Lean development of the business model was initially described by Ries (2011) who is said to originate the lean startup movement. Here, an agile behavior is advocated, there is a focus on rapid product and business model development by seeking early feedback from the customer and iterating upon that making sure one is providing the customer with value. The chances of developing a successful business is said to increase by engaging in this behavior. Startup companies typically have limited resources and need to maintain a strong focus on serving the customer value and not spend time on activities not adding value, and can use the business model as a focusing device (Morris et al., 2005). Simple and flexible tools are said to facilitate business model development. Ostewalder and Pigneur (2005; 2010) and Ash Maurya (2012) have both developed a visual one-page tool that describes a business model as a series of elements and interrelations. This approach to the business model is grounded in the business-model elements research perspective (Ritter and Lettl, 2018) as previously described. The canvases are said to accommodate the need for a simple, yet dynamic tool for business model planning and development and have been widely adopted.

3.6 Research gaps

Research gaps occur on multiple levels of the business model phenomena. This is natural, as it is a complex phenomena that one began to explore only recently.

Morris (2005) highlights the need for systematic approaches for assessing business model viability. Especially, business model performance in accommodating changing environmental conditions. The dynamics of business model emergence and evolution calls for new insights and deserves empirical attention.

One common challenge with developing sustainable business models is to maintain or increase firm revenue while delivering social and environmental benefits, and continue to add value to the customer (Bocken et al., 2014). Product-service-systems (PSS) are emerging in industry and promoted in research as having a sustainable business model with multiple benefits in the economic, environmental and social dimension compared to traditional product or service-centered business models (Bocken et al., 2014). However, PSS is not shaped practical applications, (Mont, 2002), key barriers for implementation are present (Tukker, 2004) and sustainable PSS theory with explanatory and predictive capabilities lack (Tukker and Tischner, 2006).

Demil and Lecocq (2010) state that the business model is and should be in a 'permanent state of evolution', where minor and incremental changes (or iterations) are conducted continuously. This results in the business model improving over time, which corresponds well with the principle of continuous improvement inherent to the lean mindset. This notion, *applying lean principles to an entity will improve the entity*, have led to wide adoption and implementation of lean principles and a lean mindset. This is the case for leans original use, manufacturing processes (Morgan and Liker, 2006; Welo, 2011). However, other cases of successful implementation of lean are rare (Tortorella et al., 2016). For example, the performance capabilities in product development teams remain unchanged despite implementing lean principles (Welo and Ringen, 2017). The question becomes; are lean principles effective? Do they work at all?

Despite the wide use and adoption of the Lean Canvas (Maurya, 2012) and the Business Model Canvas (Osterwalder and Pigneur, 2010), there is some confusion related to the use of and effectiveness of using such tools (Borseman et al., 2016). Which canvas should one use? Do we know if these tools result in a leaner behavior? And; does lean development result in a amore successful startup?

3.7 Summarizing the business model, providing foundation for the scope

By synthesizing the notions above, one could say the essence of the business model is to aid company success by discovering, developing and sustaining competitive advancements

of the company. Such competitive advancements and developments can be found on multiple levels of the business model. The interrelations and interactions of these levels is a part of the core business functioning and the activities in the company. How these levels function and how they function together affects the success of the company and this should be reflected in business model research. Therefore, to begin to fully understand the business model phenomena, it is interesting and important to investigate all levels. This ranges from a top level, investigating the concept business model and business model development, to the network and company level, further down on a level addressing the company's product offerings to the customer.

3.8 Research methods for business model research

Qualitative work is more common in business model research, which serves the purpose to an extent (Gemmell et al., 2012; George and Bock, 2011; Yin, 2017). While there are strong advantages with qualitative research, there are aspects of the business model that qualitative methods cannot capture, to which we should turn to other methods. Empirical quantitative work in business model research is called upon (Barney et al., 2001; George and Bock, 2011) The author believes that a comprehensive set of methods could be utilized to explore and investigate a complex phenomena, such as the business model with a multitude of levels. A mixed method methodology allows for the method to adapt according to the research process' progress and initial insights. This allows different directions and does not constrain the research in a potentially disruptive manner, as a monomethod may do if it is not suited to the theoretical domain or the explorative mode.

4 A literature investigation (paper 1)

4.1 Introduction to the literature investigation

This section introduces the first paper, which conduct a literature investigation on means and measurements to be used for continuous improvement in new product development. It aims at placing the paper context in the overall business model context and provide details on the literature approach.

4.1.1 The business model phenomena on a process level

As argued in the introduction; a company should have one or more competitive advantages over competitors to gain and sustain a market share. Lean receives increased attention in product development and manufacturing companies, as well as also increasingly being applied for business planning and development purposes. Lean principles aims at providing a business or a process with competitive advantages by maximizing customer value (Browning, 2003; Morgan and Liker, 2006), while reducing unnecessary activities. Lean principles could be used as a strategy for the whole or parts of a business model, or even be the mindset driving the initial planning and further development of the business model. For example; in a company providing a physical product to the customer by using a lean product development process, lean principles are a core part of their business model. Adhering to lean principles enables the business to deliver an increased customer value at a decreased time-to-market. To do so, one should strive to continuously improve and learn (Morgan and Liker, 2006).

To improve an entity, one must first be able to measure the current state of the entity. This is also true for a business model, the BM development process, or any subset of it. The paper encompasses means to measure continuous improvement of processes with lean principles as the foundation. This could aid in understanding the business model phenomena on an overarching process (or company) level and how these processes could be improved.

4.1.2 The literature approach

A literature review, study or investigation can be conducted to provide an overview of what currently exist in research. It involves identifying and determining existing knowledge, enabling deep investigation and understanding of a topic. It could point towards research

gaps, towards unfinished theories, collect and compare theories, etc. From there one can propose new research directions or propose a theory from collected and synthesized knowledge (Fink, 2005; Given, 2008).

The process in essence includes a search, identification and selection of relevant articles before going in depth of the selected articles. A search can be conducted in a structured or unstructured manner. It is emphasized that one must review a number of different aspects around the research topic of interest (Fink, 2005; Given, 2008), as other research domains holds the potential to inform the topic from a different perspective.

A classic systematic literature review is a rigorous and standardized methodology, it should be comprehensive and the procedure reported in an explicit manner for it to be reproducible. This process begins with selecting bibliographic databases and websites from the research questions, followed by selection of research terms. Research terms are important and should cover multiple terms or labels of the same construct to be comprehensive, and consulting experts is recommended here. A practical process should be developed, covering data collection metrics (such as sample size and language) and data analysis, before piloting the process. Results should be synthesized, keeping the quality of the selected studies and the results in mind, before writing the final study. The final study could be descriptive (describe how a construct works or process takes place), prescriptive (literature used to form an opinion on and state how a construct should work or a process be conducted) or analytical (combing the resulting articles by statistical tests) (Fink, 2005).

A descriptive study (Blessing and Chakrabarti, 2009; Fink, 2005), can be used to inform empirical studies. To correctly inform an empirical study it is particularly important to distil statements from remarks in the selected literature and be critical of their quality.

4.1.3 Introducing the paper

The first paper is “Dybvik, H., Erichsen, J. A. B., Steinert, M., & Welo, T. (2018, June). Evaluating Continuous Improvement Efforts in New Product Development. In ISPIM Innovation Symposium (pp. 1-14). The International Society for Professional Innovation Management (ISPIM).“. It was presented at ISPIM Innovation conference, “Innovation – The Name of the Game”, in June 2018.

It encompasses an overall literature investigation of means to measure continuous improvement of processes and it proposes possible metrics for making such assessments.

The context here is the new product development process with lean principles providing the foundation for the process and the mindset. Here, an unstructured literature investigation was conducted, and to distinguish it from the classic systematic literature review, the term *mapping* appears in the paper to describe the process. Furthermore, the paper includes a discussion regarding the effectiveness of such methods and the proposed metrics.

4.2 The first paper: “Evaluating Continuous Improvement Efforts in New Product Development”

The paper is enclosed in the following pages.

This paper was presented at The ISPIM Innovation Conference – Innovation, The Name of The Game, Stockholm, Sweden on 17-20 June 2018. The publication is available to ISPIM members at www.ispim.org.

Evaluating Continuous Improvement Efforts in New Product Development

Henrikke Dybvik*

Norwegian University of Science and Technology, Richard Birkelands vei 2B, 7491 Trondheim, Norway.
E-mail: henrikkd@stud.ntnu.no

Jørgen Andreas Bogen Erichsen

Norwegian University of Science and Technology, Richard Birkelands vei 2B, 7491 Trondheim, Norway.
E-mail: jorgen.erichsen@ntnu.no

Martin Steinert

Norwegian University of Science and Technology, Richard Birkelands vei 2B, 7491 Trondheim, Norway.
E-mail: martin.steinert@ntnu.no

Torgeir Welo

Norwegian University of Science and Technology, Richard Birkelands vei 2B, 7491 Trondheim, Norway.
E-mail: torgeir.welo@ntnu.no
* Corresponding author

Abstract: This paper addresses the need for methods for evaluating continuous improvement efforts in new-product development (NPD), and gives an indication of various possibilities. In an increasingly competitive marketplace, the notion of lean principles in a NPD context is emerging. Essentially those principles are concerned with increasing customer value while eliminating waste, however many companies fail to continuously improve despite implementing lean principles. Methods for evaluating continuous improvement efforts within lean principles have been mapped along with suggested indicators. Establishing performance indicators with this in mind is not a straightforward process, as evaluating performance in the NPD phase is not straightforward, neither within a team or across contexts.

Keywords: continuous improvement; new-product development; lean principles; methods; metrics.

1 Introduction

This paper presents suggestions for indicating performance and evaluating continuous improvement efforts in New Product Development (NPD). This includes addressing NPD

performance, both in understanding what makes a product manufacturing company successful and how to compare PD companies across various contexts.

Gaining or maintaining competitive advantages are, among other, dependent on the ability to continuously improve, a core part of lean principles. Lean is becoming an embedded part of the western manufacturing facilities (Morgan and Liker, 2006). After implementing Kanban, Just-in-Time and other lean principles as a part of their production process they turn to lean principles also for the PD process. Thus, the adaptation and implementation of lean principles and practices in PD is growing. However, the reported cases of successful implementation of lean principles in the NPD is rare (Tortorella et al., 2016), in fact recent findings reveal no change in terms of capability for the PD team, both current and desired (Welo and Ringen, 2017). This includes performance. Hence, it appears that the lean PD transformation initiatives initiated by many companies are not adding sufficient value.

In a NPD setting, where the concepts are genuinely novel, there is little previous knowledge and large amounts of uncertainty (Gerstenberg et al., 2015; Reinertsen, 1999; Steinert and Leifer, 2012), systematic PD process optimization is somewhat inapplicable. Therefore, investigating the impact of various improvement efforts in NPD is of high interest. As lean in a NPD context is relatively new, there is a need for methods and metrics for measuring PD with all its dimensions. It is also important to distinguish between 1) having implemented a lean PD process, 2) ensuring lean work effectively, 3) having implemented continuous improvement efforts, and 4) have these efforts work effectively. Therefore, the objective of this paper is to spark a discussion around methods and metrics for assessing continuous improvement efforts in new product development. This aim is to contribute to this discussion by proposing some dimensions of interest for assessments of NPD performance.

In this work, the various dimensions of distinct lean models in a NPD context have been identified along with the key principles from Toyota Product Development System. These were compared and combined based on resemblance and from that research on methods and suitable metrics for continuous improvement efforts conducted. Principles or dimensions not seen as directly applicable to a general NPD context have been omitted. General improvement practices and metrics have also been considered.

This paper consists of an introduction to the key principles of lean NPD, a section summarizing the lean principles with proposed methods for measuring PD efforts. Lastly, a discussion on assessing PD performance is presented.

2 Lean PD in a New Product Development Context

There are various definitions of lean in a NPD context. Essentially, lean in a PD context concerns itself largely with understanding of customer value. Maximizing customer value (Browning, 2003), by conducting the correct activities at the correct time while eliminating activities that do not add any value.

There are various ways to describe a lean PD model in more depth. Most of them have common principles, a total of 13 (the first column of all tables) have been identified from Toyota Product Development System (TPDS) (Morgan and Liker, 2006) and summarized with various models by Welo (2011). Further, it is worth noting that since the TPDS principles have been developed for high functionality in the Japanese automotive industry over decades, not all 13 principles are directly applicable in a general

NPD context. A knowledge-based lean PD framework proposed consist of six dimensions (Welo et al., 2013):

- Customer value
- Knowledge and learning
- Culture
- Stabilization
- Standardization
- Continuous improvement (applied to all other dimensions)

Of the six dimensions listed above, continuous improvement is the one dimension that directly relates to measuring PD performance. Continuous improvement in every part or principle will lead to continuous improvement of the overall NPD performance.

In the following sections, methods for initiating continuous improvement discussed and an indication as to how these can be used evaluating performance are discussed. In this attempt, three focus areas have been chosen, those being process, people and lastly, tools and technologies according to the sociotechnical system that lean is (Morgan and Liker, 2006).

2.1 Process

2.1.1 Customer value

Capturing customer defined value is by most considered to be the single most important effort to create a differentiated product. There exist many methods for doing so, ranging from so-called immersing to enhanced requirement specifications. However, the notion of customer value in PD is far from uniform, and the semantics within the lean PD community is not clear on how to quantify this value. Results from Overvik Olsen and Welo (2011) suggest that methods revealing emotionally-related customer information (e.g. workshops and observation) are more suited for product innovations, in terms of increasingly differentiated products. Methods providing more functionally-related information (e.g. web based survey and interview) are suited for product improvements. Hence, the earlier in the PD process, the harder it is to quantify the customer value, as there are inherently more uncertainties present earlier on in a project (Sutcliffe and Sawyer, 2013). This also shows that early customer interaction is important in order to assess customer value. As such, the PD team need to make sure they are cautious of chosing the right methods, adapted to their development context. To measure if this knowledge-creation process is effective, the quality of the information the method provides in terms of type and uniqueness is a relevant metric (Overvik Olsen and Welo, 2011). The level of qualitative information can be assessed by comparing them to previous values, aiming at continuous improvement of those or conduct goalsetting.

Table 1 Key principles in lean PD models, matched with the six dimensions

Process

Key principle	*Description*	*Proposed methods for evaluating efforts*
Establish customer defined value to separate	Waste is non-value added as defined from customer value. The traditional	Interviews

added value from waste. (Customer value)	definition of waste in manufacturing cannot be used in PD; focus must be placed on information and knowledge.	Observations Workshops Qualitative information can be assessed and compared to previous values.
Frontload the PD process to thoroughly explore solutions while there is maximum design space. (Knowledge and learning)	Defining the wrong problem or selecting a premature solution will have large cost implications throughout the product life cycle. Problems must be solved at the root cause, and all solutions must be carefully evaluated using set-based design methods.	Project-to-project knowledge transfer. Rapid problem-solving. Preparation of postmortems. Content comparing for correct problem-solving focus. Suitable document type. Strategic use of CAD/CAE. Updates knowledge and software.
Create a levelled Product Development process flow. (Stabilization)	Stabilize the PD process so that workflow can be predicted and planned. Resource capacity should be planned at a level that maximizes efficiency. Manage workload in a project and between projects, using process and resource planning and flexible labor pools.	Scrum and careful reflection upon the process flow. Feedback log, reflection log, velocity and effort estimates conducted by team.
Utilize standardization to reduce variation, and create flexibility and predictable outcomes. (Standardization)	Continuous improvement requires standardization, which represents the foundation for all process principles in Toyota's model. Follow the implementation sequence of stabilizing, standardizing and continuously improving.	
Other characteristics in the area of process. (Stabilization and standardization)		

2.1.2. Frontloading

Frontloading the PD process is a strategy to thoroughly explore solutions while there is maximum design space. Knowledge of technology and from previous project conducted in the company is important. Defining the wrong problem or selecting a solution too early will have large negative implications throughout the PD process and the product life cycle, thus approaches such as 'decide as late as possible' have emerged. Time is also wasted on solving the same or similar problems over again. Problems must be solved at the root cause, and all solutions must be carefully evaluated (Sobek et al., 1999). Frontloading is highly dependent on the quality of information flow from one project to another, that is project-to-project knowledge transfer, which is leveraging previous projects by transferring problem and solution-specific information to new projects.

Methods for improving project-to-project knowledge transfer include the use of "postmortems", and utilizing advanced technologies. Postmortems are records of post-project learning and thus can bring forward the knowledge, especially the problems

encountered and solved, from ongoing and past projects (Thomke and Fujimoto, 2000). To measure the effect, one could first ask if postmortems are being prepared. Then comparing content from older and newer postmortems to track if the same problems have been encountered and to ensure that the problem-solving efforts are focused on new problems rather than the ones already solved. It's crucial to have the purpose of the postmortems thoroughly explained and suitable document type according to the problem chosen.

Quantification of frontloading efforts could include the mapping of people assigned to a particular project, to then assess the amount of frontloading present in that project. Moreover, mapping of postmortems might reveal a culture for project-to-project knowledge transfer, and would help in assessing who-knows-what when starting new projects.

2.1.3. Utilizing advanced technologies

The ability to utilize advanced technologies could potentially have great impact on the effectiveness of the team. This is also known as rapid problem-solving; using advanced technologies and methods to increase the overall rate at which development problems are identified and solved (Thomke and Fujimoto, 2000). As Thomke and Fujimoto suggests, rapid problem-solving can be achieved by combining careful use new technologies (such as computer simulations, CAD and CAE), which enables faster problem-solving cycles. This technology must be used efficiently, for instance through strategic use of CAD (Bhavnani et al., 1999), keeping both software and knowledge updated. It has become increasingly important for companies to utilize Product Lifecycle Management software to manage and keep track of large, complex projects (Alavi and Leidner, 2001). This kind of software provides a multitude of useful information, both on design activities (such as CAD), as well as information (e.g. document workflows).

2.1.4. Stabilization

Stabilization of the PD process so that workflow can be predicted and planned is one of the key principles, hence a system for this need to be in place. Resource capacity should be planned at a level that maximizes efficiency. Flexible, yet effective process and resource planning is desired and could be achieved by agile methods such as Scrum (described under Tools and Technologies). One way of assessing stabilization of both PD process and work-environment would be an assessment of new project initiatives, and map this against the amount of administrative changes in the PD organization.

2.2. People

2.2.1. Goal Setting

According to Locke et al. (1981) goals affect performance by directing attention, mobilizing effort, increasing persistence, and motivating strategy development. Hence, we argue that setting specific and challenging goals will be sufficient for driving continuous improvement in each of the principles. Goal setting will most likely elevate performance when the goals are sufficiently specific and challenging. Engineers must have the ability (in terms of academic knowledge or experience, which can be mapped and controlled through for instance a survey) to complete the task. Feedback is provided continuously to show progress in relation to the goal, rewards are given when the goal is reached, the supervisor is supportive, and assigned goals are accepted by the individual,

or the goals are co-created. Thus, it is crucial to make sure there is given continuous or periodical feedback, from the PD team in addition to supervisor. This could for instance be solved by an assigned timeslot for the sole purpose of feedback, the correct usage of this timeslot must be monitored and the people involved held accountable. This will also promote transparency and high-quality knowledge transfer in the team. It could also organize the team around a common goal, help resource and task planning by having sub-goals targeted to the team members field and functional expertise. To assess if the correct goals have been set the success of the product in terms of added customer value can be quantified as described above.

2.2.2. Technical and personal competence

High technical competence in all engineers in addition to superior specialized knowledge should be present where actions take place, that is important stages in the PD process such as decision making. To achieve this, simply hire the best people and have the continuous development of their skills in place. That stated, hiring is a complex process and could be tiring despite being crucial. Industry and start-ups have found it more efficient to use extensive resources and time during this process to be more effective (rather than hiring the wrong person) and a series of technical and personal tests have proven to be a success factor. Technical capabilities, depth and ability to think and communicate can be evaluated by a series of graded questions. If a job applicant makes a good impression during such an initial screening there will be several stages with questions in greater depth, eventually passing the test. Similarly a candidate's business judgment, psychological and emotional fit must be assessed in a similar test as described by Nanda and Mahmood (1997). It is problematic to continuously assess employee performance and at the same time leverage a trustful and inspiring work-environment, and there are challenges to measuring continuous improvement (or change whatsoever) on this topic, but one way around this might be the use of qualitative assessments from team and project leaders in addition to goalsetting.

2.2.3. Organizational learning

Organizational learning is identified as a key component in a knowledge-creating company (Nonaka and Takeuchi, 1995), and as output from formalization in the SECI spiral of knowledge transfer (Nonaka et al., 2000). One can argue that since learning is a key output of any PD activity, organizational learning will therefore be a vital part of a continuously improving company. Leifer and Steinert (2011) have identified three learning loops in PD activity, each describing a different level of abstraction in a learning organization. Building on the relation between the SECI-model and learning, this is further investigated by Erichsen et al., (2016), who discuss some aspects of knowledge transfer in a PD setting. Erichsen et al. (2016) put special emphasis on learning through tacit knowledge (the learning that is done by the individuals) and the formalization of tacit knowledge (i.e. creating organizational learning). In these three learning loops by Leifer and Steinert (2011), Learning Loop One is based on formalization of knowledge which can be collected, managed and combined into formal processes after a formalization has taken place. Explicit knowledge is often described as information, and can be embodied in quantitative technical data such as business processes, CAD files and workflows, data warehouses etc. Learning Loop Two occurs in the informal space of the PD team, activities are design practices resulting in faster learning and better output, and consists of informal process content, such as concepts, semantics, and architecture and

during which questions like when and why arise. Learning Loop Three involves accumulating tacit knowledge embedded in the team and their established practices. Team members learn from each other and prior team's experiences by applying, reflecting upon and improving informal practices.

Metrics for measuring activity in these three learning loops could be amount of documented knowledge, quantified communication and knowledge transfer (both tacit and explicit). Logging design questions proposed by Steinert and Leifer (2011) may act as metrics, where Deep Reasoning questions (DRQs) (which reflects convergent thinking) and Generative Design Questions (GDQs) (which reflects divergent thinking) can be used for assessing the activity done in the three learning loops. Other metrics for determining effective capturing and use of knowledge should be further investigated, as outlined by Erichsen et al. (2016).

2.2.4. Team culture and performance

Morgan and Liker (2006) stress that building a culture through a common mind-set supports excellence and relentless improvement. To support this effective communication, transparency, team building, confrontation (in the case of conflicts) and various other tools (e.g. the use of prototypes) are suggested. Charged teams should have an arrangement where members are accountable to the team and where their evaluations and rewards are also linked to the performance of the team. According to Leifer (1998) the combination of an explicit feedback assessment models and advanced technology for measuring and facilitating team activity gives performance metrics that can be used by the team to monitor and improve their productivity. A collaborative team environment is facilitated by; 1) A subjective, behavior-based, index for controlling the uncertainty of learning preferences which acts as an input to information and team performance. 2) An objective, content-based, measurement of the content of work-in-progress to predict the quality of a team's final product. Data from work-in-progress assessments should be used by teams for self-assessment, thus enabling comparison to earlier performance.

There have been various efforts to evaluate team performance using different forms of information sharing (Edelman, 2011; Jung, 2011; Wulvik et al., 2016). Here, forms of information sharing are utilized as indicators for well performing or under performing groups. Attempts have been made to reveal communication patterns through the use of Temporal Static Visualization (Wulvik et al., 2017a), where different modes of information sharing (e.g. monologue, discussion) can be visualized. The use of computational tools for analyzing these conversation dynamics (Wulvik et al., 2017b) gives a metric for the contributions of each team member as well as for the team overall. This could be used to evaluate the team dynamics, for instance during the decision-making process ensuring equal contributions from the team members.

Table 2 Key principles in lean PD models, matched with the six dimensions

People

Key principle	*Description*	*Proposed methods for evaluating efforts*
Develop a Chief Engineer (CE) system to integrate development	The CE 'owns' the product with final authority and responsibility for the entire PD process. He is the	

from start to finish.	customer representative, managing integration and decisions.	
Organize to balance functional expertise and cross functional integration. (Culture, knowledge and stabilization)	Functional expertise combined with project goals and CE system provide the balance of the matrix org. Functional Managers (FM) owns functional knowledge, and are in charge of resource planning /allocation to serve the CEs	Goalsetting.
Develop high technical competence in all engineers. (Knowledge and learning and culture)	High competence and superior specialized knowledge represents the basis; and these have to be established at the places where actions take place.	Hire top people. Develop current employees. Technical and personal tests.
Fully integrate suppliers into the PD system. (Standardization)	Suppliers must be integrated into the PD process and their competence, capabilities and culture must be compatible. Define long-term supplier relationships.	
Build in learning and continuous improvement. (Culture, knowledge and learning)	Organizational learning represents the basis for cont. improvements, and build on all the other principles.	Engage in Learning Loop One, Two and Three. Suitable documentation, effective knowledge transfer and communication. Deep Reasoning questions and Generative Design Questions.
Build a culture to support excellence and relentless improvement. (Culture)	Excellence and Kaizen in the final analysis reflect the organizational culture.	Communication, team building, etc. Subjective, behavior-based, index as input to team performance. Objective, content-based, measurement of work in progress. Self-assessment.

2.3. Tools and technology

2.3.1. Agility

Agile methods are used to handle the challenges of managing complex projects during the development phase by exerting product development flexibility (Smith, 2007), and are a collection of incremental and iterative methods that are more effective than traditional project management tools. Kanban and Scrum are two powerful agile project management approaches (Lei et al., 2017) because it leverages the development process by identifying the tasks, managing time more effectively, and setting up teams.

Firstly, making sure the team have reached a consensus as to what method to use is crucial, measured by a simple yes or no. Secondly, for continuous improvement of Scrum and Kanban, Scrum provides a feedback loop, in which the feedback given needs to be taken into account and adjusted for. Keeping a log of feedback and suggested actions to be used for comparison when it is time to re-reflect upon the process (during the retrospect meeting (Lei et al., 2017) could enforce change and improvement. The team should assess if the process as a whole has improved, a subjective measure established to

suit the process. For instance, counting positive and negative feedback from customers and stakeholders resulting from each sprint or iteration and continuously comparing the increase in positive feedback (or decrease in negative feedback) to the team's own reflection upon the process. High influence metrics used by agile teams have been analyzed by Kupianen et al., (2015) based on occurrences and perceived importance factor. The results include many metrics, the most influential one being Velocity and Effort estimates. The reasons for and the effects of using metrics are focused on the following areas: sprint planning, progress tracking, software quality measurement, fixing software process problems and motivating people.

2.3.2. Prototyping and aligning the organization through visual communication

Prototyping as a mean to uncover uncertainty in early stage PD projects is becoming a reoccurring topic in early stage development research (Erichsen et al., 2016; Jensen et al., 2015; Sutcliffe and Sawyer, 2013). It is effective for communication of concepts, establishing a consensus in the team and research establishes it as an absolute essential part of a successful NPD process (Jensen et al., 2015; Leifer and Steinert, 2011). To measure prototypes Jensen et al., (2015) defines 51 closed, quantifying questions, which quantifies prototypes, their characteristics and their generated output in a standardized way, and they are intended as standard parameters. The answers to these questions will serve as documentation and reflection, and in addition when answering the 51 questions the researchers or engineers will be 'forced' to consider and 'count' how the prototype performed. I propose that if this is conducted as a team, the team will reach a consensus regarding the quality of the prototype, the learnings and the knowledge transfer from project to project, - in turn aligning the organization, and facilitate and improve organizational learning.

Even though prototypes are often physical and documenting in themselves, it is hard to include them in written deliverables or presentations. Quantifying physical concepts during the development process in order to track which actions in the process gives the most improvement could provide a useful technique, possibly through the capture of prototypes and other project output (Sjöman et al., 2017).

Other suggested tools for aligning the organization through simple, visual communication are visual boards, both manual and automatic to illustrate for instance the backlog (Lei et al., 2017) highlighting problems or assigning tasks, in addition to the well-known use of A3s in Toyota (Sobek II and Smalley, 2011).

2.3.3. Knowledge of state of the art

It can be argued that identifying and understanding emerging technologies within a field is becoming increasingly important as more and more technologies emerge and mature. Having knowledge of the newest technologies available is therefore a competitive edge over competing PD companies. Nowadays, there is a rapid development of technologies and the use of them, and it is argued that updated knowledge will contribute to the use of these technologies leading to more and better prototypes, as well as faster iterations. The team and organization could collectively keep each other updated. A suggested method for this could be a weekly state of the art meeting where current advancements are discussed and the metric being if this meeting is conducted or not, and if news and relevant research communities have been researched.

Table 3 Key principles in lean PD models, matched with the six dimensions

Tools and technology

Key principle	*Description*	*Proposed methods for evaluating efforts*
Adapt technology to fit people and process.	Technology must be customized to fit people and process, and is always subordinated to the people and the process.	
Align organization through simple, visual communication. (Stabilization, knowledge and learning)	Aligned goals must be flown down in the organization, and problem solving is enabled by visual communication.	Scrum. Kanban adaptations and other agile methods. Reflection upon prototypes. Assessing reflection upon correct execution by logging positive or negative change. Velocity and Effort estimates. Framework with 51 quantifying question. To be performed periodically.
Use powerful tools for standardization and organizational learning. (Standardization)	Powerful tools can be simple. Their power comes from enabling standardization, which is necessary for organizational learning.	Reflection upon prototypes. Framework with 51 quantifying question. To be performed periodically.
Other characteristics in the area of tools and tech. (Knowledge and learning)		Having updated knowledge of state of the art. Periodically update on research communities and news.

3. Concluding remarks

One suggested method for measuring performance in PD is through the use of Key Performance Indicators (KPI), generally used because of their effectiveness as a performance metric in the automotive industry (Haque and James-Moore, 2004; Ringen, 2010). KPI in a NPD context supports and united the team, as it promotes common grounds for decision-making, motivation, clarifying goals and priorities, facilitate communication, etc. KPIs can be made especially useful for evaluating efforts in a NPD context (Godener and Söderquist, 2004) in that they serves a purpose of documenting the value of development efforts (Hauser and Zettelmeyer, 1996), by using them to measure the effect of an organizational change activity. Trend monitoring used for aligning the team, through easy visual communication, and focus can additionally be obtained through KPIs.

However, it is worth noting that when developing KPIs from the suggested metrics, lean thinking should be present (Haque and James-Moore, 2004), and one must be careful as establishing KPIs is not a straightforward process. KPIs can be formulated using a variety of dimensions, which need to be adapted to method of process evaluation and the metric seen as the most crucial ones. Dimensions range from qualitative, quantitative, positive, negative, leading, lagging etc., and one should strive to incorporate a suitable

variety of these. For instance, recording both positive and negative as opposed to solely negative measures, and include both qualitative and quantitative measures, as for instance metrics related to "People"-principles in the NPD phase are highly qualitative measures.

The heuristic on creating KPIs is that each PD organization or team should establish their KPIs for their improvement efforts, in order to fit their specific context. This raises the questions of whether a performance indicator can be used to indicate performance in two different PD contexts. Another key question to ask is whether a good performance indicator is those that only indicate good PD performance. It can be argued that a good performance indicator will help reveal PD performance on a nuanced scale, with both good and poor, and can be applied to a multitude of PD contexts - also NPD.

In this paper, we attempt to raise some questions on how to assess NPD performance, both in understanding what makes a NPD company successful and how to compare NPD companies across various contexts.

4. Conclusion

This paper attempts to address this issue by proposing some dimensions of interest for assessments of NPD performance, based the proposed evaluation methods in Tables 1, 2 and 3. How to identify good performance on various abstraction levels in a NPD organization, and how to identify good performance over different contexts are among the questions arise when discussing NPD performance. To get to grips with PD performance, we have proposed to first look to continuous improvement efforts and the evaluation of such efforts. Since the notion of continuous improvement comes from deliberate and positive change over time, one can argue that it is impossible to measure continuous improvement efforts without studying the same PD context over time. By starting to monitor and quantify input to and output from various PD activities (e.g. chosen design methods, prototypes, prototyping methods, communication, people, cost, time, etc.), we can start to understand what metrics to use for assessing (both good and poor) PD performance.

Acknowledgements

This research is supported by the Research Council of Norway through its user-driven research (BIA) funding scheme, project number 236739/O30

References and Notes

Alavi, M., Leidner, D.E., 2001. Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Q. 25, 107–136. https://doi.org/10.2307/3250961

Bhavnani, S.K., John, B.E., Flemming, U., 1999. The strategic use of CAD: An empirically inspired, theory-based course. Presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM, pp. 183–190.

Browning, T.R., 2003. On customer value and improvement in product development processes. Syst. Eng. 6, 49–61.

Edelman, J.A., 2011. Understanding radical breaks: media and behavior in small teams engaged in redesign scenarios. Stanford University.

Erichsen, J.A.B., Pedersen, A.L., Steinert, M., Welo, T., 2016. Learning in Product Development: Proposed Industry Experiment Using Reflective Prototyping. Procedia

CIRP, 26th CIRP Design Conference 50, 454–459. https://doi.org/10.1016/j.procir.2016.04.142
Erichsen, J.A.B., Pedersen, A.L., Steinert, M., Welo, T., 2016. Using prototypes to leverage knowledge in product development: Examples from the automotive industry, in: 2016 Annual IEEE Systems Conference (SysCon). Presented at the 2016 Annual IEEE Systems Conference (SysCon), pp. 1–6. https://doi.org/10.1109/SYSCON.2016.7490586
Gerstenberg, A., Sjöman, H., Reime, T., Abrahamsson, P., Steinert, M., 2015. A Simultaneous, Multidisciplinary Development and Design Journey – Reflections on Prototyping, in: Entertainment Computing - ICEC 2015, Lecture Notes in Computer Science. Presented at the International Conference on Entertainment Computing, Springer, Cham, pp. 409–416. https://doi.org/10.1007/978-3-319-24589-8_33
Godener, A., Söderquist, K.E., 2004. Use and impact of performance measurement results in R&D and NPD: an exploratory study. RD Manag. 34, 191–219.
Haque, B., James-Moore, M., 2004. Applying lean thinking to new product introduction. J. Eng. Des. 15, 1–31.
Hauser, J.R., Zettelmeyer, F., 1996. Evaluating and Managing the Tiers of R&D.
Jensen, M.B., Balters, S., Steinert, M., 2015. Measuring Prototypes-A Standardized Quantitative Description Of Prototypes And Their Outcome For Data Collection And Analysis. Presented at the DS 80-2 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 2: Design Theory and Research Methodology Design Processes, Milan, Italy, 27-30.07. 15.
Jung, M.F., 2011. Engineering team performance and emotion: Affective interaction dynamics as indicators of design team performance.
Kupiainen, E., Mäntylä, M.V., Itkonen, J., 2015. Using metrics in Agile and Lean Software Development – A systematic literature review of industrial studies. Inf. Softw. Technol. 62, 143–163. https://doi.org/10.1016/j.infsof.2015.02.005
Lei, H., Ganjeizadeh, F., Jayachandran, P.K., Ozcan, P., 2017. A statistical analysis of the effects of Scrum and Kanban on software development projects. Robot. Comput.-Integr. Manuf., Special Issue: Extended Papers Selected from FAIM 2014 43, 59–67. https://doi.org/10.1016/j.rcim.2015.12.001
Leifer, L., 1998. Design-Team Performance: Metrics and the Impact of Technology, in: Evaluating Corporate Training: Models and Issues, Evaluation in Education and Human Services. Springer, Dordrecht, pp. 297–319. https://doi.org/10.1007/978-94-011-4850-4_14
Leifer, L.J., Steinert, M., 2011. Dancing with ambiguity: Causality behavior, design thinking, and triple-loop-learning. Inf. Knowl. Syst. Manag. 10, 151–173. https://doi.org/10.3233/IKS-2012-0191
Locke, E.A., Shaw, K.N., Saari, L.M., Latham, G.P., 1981. Goal setting and task performance: 1969–1980. Psychol. Bull. 90, 125.
Morgan, J.M., Liker, J.K., 2006. The Toyota product development system : integrating people, process, and technology. Productivity Press, New York.
Nanda, A., Mahmood, T., 1997. Vermeer Technologies (A): A Company Is Born.
Nonaka, I., Takeuchi, H., 1995. The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford university press.
Nonaka, I., Toyama, R., Konno, N., 2000. SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. Long Range Plann. 33, 5–34. https://doi.org/10.1016/S0024-6301(99)00115-6
Overvik Olsen, T., Welo, T., 2011. Maximizing product innovation through adaptive application of user-centered methods for defining customer value. J. Technol. Manag. Innov. 6, 172–192.
Reinertsen, D.G., 1999. Taking the fuzziness out of the fuzzy front end. Res.-Technol. Manag. 42, 25–31.

Ringen, G., 2010. Organizational learning and knowledge in the Norwegian automotive supplier industry. Unpubl. PhD Thesis Nor. Univ. Sci. Technol.
Sjöman, H., Erichsen, J.A.B., Welo, T., Steinert, M., 2017. Effortless Capture of Design Output A Prerequisite for Building a Design Repository with Quantified Design Output. Presented at the Presented at the IEEE International Conference on Engineering, Technology and Innovation 2017, IEEE, Madeira.
Sobek, D.K., Ward, A.C., Liker, J.K., 1999. Toyota's Principles of Set-Based Concurrent Engineering. Sloan Manag. Rev. Camb. Mass 40, 67–83.
Sobek II, D.K., Smalley, A., 2011. Understanding A3 thinking: a critical component of Toyota's PDCA management system. CRC Press.
Steinert, M., Leifer, L.J., 2012. "Finding One's Way": Re-Discovering a Hunter-Gatherer Model based on Wayfaring. Int. J. Eng. Educ. 28, 251.
Sutcliffe, A., Sawyer, P., 2013. Requirements elicitation: Towards the unknown unknowns, in: 2013 21st IEEE International Requirements Engineering Conference (RE). Presented at the 2013 21st IEEE International Requirements Engineering Conference (RE), pp. 92–104. https://doi.org/10.1109/RE.2013.6636709
Thomke, S., Fujimoto, T., 2000. The Effect of "Front-Loading" Problem-Solving on Product Development Performance. J. Prod. Innov. Manag. 17, 128–142. https://doi.org/10.1111/1540-5885.1720128
Tortorella, G.L., Marodin, G.A., Fettermann, D. de C., Fogliatto, F.S., 2016. Relationships between lean product development enablers and problems. Int. J. Prod. Res. 54, 2837–2855. https://doi.org/10.1080/00207543.2015.1106020
Welo, T., 2011. On the application of lean principles in Product Development: a commentary on models and practices. Int. J. Prod. Dev. 13, 316–343.
Welo, T., Ringen, G., 2017. Lean assessment and transformation strategies in product development: a longitudinal study. 87-2 Proc. 21st Int. Conf. Eng. Des. ICED 17 Vol 2 Des. Process. Des. Organ. Manag. Vanc. Can. 21-25082017.
Welo, T., Tonning, O.R.B., Rølvåg, T., 2013. Lean Systems Engineering (LSE): Hands-on Experiences in Applying LSE to a Student Eco-Car Build Project. Procedia Comput. Sci. Complete, 492–501. https://doi.org/10.1016/j.procs.2013.01.052
Wulvik, A. (1), Menning, A. (2), Steinert, M. (1), 2017a. A computational approach to expose conversation dynamics in engineering design activities. 87-2 Proc. 21st Int. Conf. Eng. Des. ICED 17 Vol 2 Des. Process. Des. Organ. Manag. Vanc. Can. 21-25082017.
Wulvik, A., Erichsen, J., Steinert, M., 2016. Capturing Body Language in Engineering Design–Tools and Technologies. 85-1 Proc. Nord. 2016 Vol. 1 Trondheim Nor. 10th-12th August 2016.
Wulvik, A., Jensen, M.B., Steinert, M., 2017b. Temporal Static Visualisation of Transcripts for Pre-Analysis of Video Material: Identifying Modes of Information Sharing. Anal. Des. Think. Stud. Cross-Cult. Co-Creat. Leiden CRC Press. Francis.

4.3 Discussing the literature investigation approach

This discussion concerns the methodological approach used throughout the paper "Evaluating Continuous Improvement Efforts in New Product Development", a literature investigation. Experiences and reflections from using the method will be presented and discussed, including known advantages and limitations. A few key outtakes are emphasized.

4.3.1 A literature investigation in general

A highly detailed and thorough investigation is possible by conducting a literature review, study or investigation. There is a magnitude of information to synthesize and make sense of (Fink, 2005; Given, 2008; Hart, 2001). Conducting such an in-depth literature investigation on one particular topic provided a broad and relatively full picture of existing knowledge in the field, an ability inherent to the method which is also a key takeaway for the author.

Usually there are multiple perspectives to explore (Given, 2008), which combined with the share amount of knowledge brings both advantages and challenges. Being exposed to multiple and different perspectives can be an advantage as it provide a multitude of different input information and by considering these, mindsets can be shifted to include new perspectives. In turn, this could lead to new (and hopefully better) topics or constructs to be investigated from a larger range of perspectives, adding details to the study. The various perspectives displayed in various literature can better represent the complexity of the humans element and human relations (Given, 2008). As it can be a source of sensory details and contain experiences with human consciousness (Given, 2008) incentive is added for using literature when investigating any topic involving a human element. Having distinct and potentially disconfirming perspectives can also be confusing, especially if one is exploring a new research topic without a frame of reference to benchmark concepts with. Since there is an 'abundance' of information in each perspective it is next to impossible to cover everything, which can limit thoroughness to time.

There is always a question of when to end a literature investigation due to the share amount of accessible information and that one could extend a review to an incredible level of detail by deeply investigating a single article. Common reasons for ending a literature review are time and funding, constraints created by sample selection or other rationale particular to

the topic or study. Here, the end of investigation was determined by a time constraint and one other important consideration. In the topic selected there were a set of thirteen principles to be investigated, and the author made sure to provide a somewhat equal coverage to the principles in terms of both time spent and presence in paper writeup.

The audience of a literature review is often large as they can be more readable than other types of studies and have received credibility from a research perspective (Given, 2008). Conducting a literature review is seen as a safe move in research, provided it is conducted properly. This could be a great benefit as it could increase the audience of a topic and be leveraged in seeking sources of funding (Johnson and Onwuegbuzie, 2004), enabling future research.

4.3.2 The structured and unstructured literature investigation

One distinction between structured and unstructured literature investigations is how search terms are developed. Search terms could be actively and strategically expanded by including synonyms of the topic, other labels describing a construct and consulting other domains. A search can also be conducted utilizing only existing knowledge without expansion of search terms. Not actively seeking other search terms (due to passiveness or other reasons) could lead the research in a narrower direction, providing only conforming evidence which verifies existing knowledge. This make it more challenging to uncover research gaps. In general it is hard to discover novelties only by consulting literature, and this is typically not the intended purpose of a literature study. Overall, depending on to what degree the researcher has conducted a thorough search with search terms of high quality, the *deep investigation* of the research topic may be put in to question. This limitation may be more prominent in unstructured literature investigations as described above.

Here, an active expansion of search terms was done by the author by looking up synonyms, other labels and domains, though not to an extreme extent. Seeking experts opination on search terms is recommended (Hart, 2001), which was briefly done, though the author only realized that this was the case when looking at the paper in hindsight. An aggregation process, both of search terms and of identified articles forced the author to identify the uttermost important parts of the topic.

4.3.3 Skills for a literature investigation;

The ability to quickly read academic articles and distil crucial information benefit the research process as well as the outcome of it. If one is able to collect, review and synthesize knowledge quickly, ideas, constructs and theories could evolve further and more rapidly. This could in turn lead to either an earlier completion of a study (enabling more time for subsequent studies), a more detailed or thorough study, or both.

The ability to jointly consider different perspectives could aid in synthesizing information and forming constructs for further investigation, while a purist attitude could inhibit the researcher since it limits this process of forming a construct from diverse domains. Furthermore, a purist approach (Given, 2008; Johnson and Onwuegbuzie, 2004) to reviewing and selecting literature would result in a literature sample from one domain only and include only certain terms describing the construct. Both would limit the research outcome, since diversity remains un-represented.

In beginning the literature investigation for the paper, the author had relatively little experience in reading and extracting information from academic articles. An ability to learn at a rapid pace lead the author to experience an increased momentum of sorts, as the search progressed. Initially, reading and evaluating one single article consumed much more time and cognitive capacity than those read more toward the end of investigation.

4.3.4 Selecting sample literature

In a classical literature review a sample selection provide representativeness of a population (Fink, 2005) and the sample selection process is established beforehand. By doing so the sample is limited to the initial boundary conditions or assumptions[1]. It is important to be aware of how literature (that is, the data) is selected, gathered, what it contains, to make correct interpretations and comparisons. This is also necessary to report for study reproducibility.

4.3.5 Outcome from conducting a literature investigation

In general, the various forms of a literature approach could provide the input to several types of studies. A descriptive literature study can be conducted to inform empirical studies

[1] An exception is choosing a random sample, though in the authors' experience this rarely seems to be the case. This may be due to leaving out known or important studies, and/or including studies of poor quality.

(Blessing and Chakrabarti, 2009). A study containing multiple perspectives having synthesized existing knowledge could form a construct from applying a logical technique, which could be tested in other studies as a step in developing a theory from the construct. It is worth noting that developing constructs are less common. More common is identification of research gaps (Fink, 2005), which provides an uncovered topic to study further, while leaving the methodological approach open for following studies. Overall, informing following studies may be the most common contribution from any literature search. It identifies literature relevant for your work, could avoid duplication work, and aid research design by collecting previous methodologies including their advantages, limitations as well as actual performance in investigating the topic (Fink, 2005). In essence a literature investigating excel at collecting a multitude of information, providing a comprehensive source for this information and point towards the gaps in this information

The research contributing resulting from conducting a literature investigation was in this case a number of methods and a few metrics to assess efforts to continuously improves a process (in the context of product development). Furthermore, a discussion on assessment of performance provides some nuance to establishing key performance indicators as there still are uncertainties to what makes a good (and bad) performance indicator.

5 The qualitative approach (paper 2)

5.1 Introduction to the qualitative approach

This section introduces the second paper "Makerspace sustainability enabled by crucial partnerships and a gentrification strategy". It takes on a qualitative approach by conducting an in-depth case study of a hardware centered startup incubator. It aims at placing the paper context in the overall business model context and provide additional details to the qualitative research approach and the case study in particular.

5.1.1 The business model phenomena on an organizational and network level

Business models not adhering to tradition emerge as global tendencies shift towards a sharing (Hamari et al., 2016) and experience (Hamari et al., 2016; Tukker, 2004) economy. The explanation to their success may not lie with the same level as with traditional business models. More so with new and untested business model is an issue with sustainability in combination with higher demands (Hamari et al., 2016). Efficient or suitable monetization of an asset or a service does not serve current societal demands, today a sustainable business model must capture economic, social and environmental value for a wide range of stakeholders extending beyond company boundaries (Bocken et al., 2014). Such a value proposition can be enabled by new technological and social innovations, which if coordinated with overall system level sustainability can result in a sustainable business model (Bocken et al., 2014). To truly understand the workings of untraditional business models, their use and development, is makes sense to investigate such an emerging phenomena from a system level point of view such as the organizational and network level.

The paper examines the case of a hardware centered startup incubator, the research question seeks to explain how and why the business model of the mentioned incubator works, - if at all. The incubator falls into a category of organizational constructs that usually experiences great challenges in establishing - let alone sustain a traditional business model. Despite this very fact, incubators and similar concepts continue to appear globally. Therefore, it is of interest to use the case of a hypothesized successful concept to develop a hypotheses that may aid other in the search for a sustainable business model. As described in the paper, a traditional business model may not always serve an organization the best, especially when the purpose of the organization is untraditional. This in-depth investigation could aid in understanding the business model phenomena on an organizational and network level.

5.1.2 The qualitative approach

A qualitative research approach in general is empirical investigation of research questions seeking explanation and meaning to a real-world phenomena, primarily in some narrative form. The *how, why, who* and *what* inherent to qualitative inquiry is designed to explore and seek understanding of the human or the social elements in a given topic (Given, 2008). Empirical evidence is captured by human individuals interpretations and retellings, thoughts and feelings through a broad range of available methods. For example; in-person interviews, various forms for observation including direct and participatory observations, collection of documentation such as diaries and journals, internal reports, financial and historical data, news, social media, etc, and other forms of multimedia documentation.

The case study is one means to the qualitative inquiry and allow for an in-depth investigation of an entity in its real-world context and it can be particularly useful when the boundaries between the entity and context are somewhat fuzzy (Yin, 2017). This entity, the selected "case", can range from being the life of one or more individuals to organizations, processes, programs, neighborhoods, institutions, events or other similar constructs. Having defined the case at the beginning of inquiry is of vital importance because this forms and guide the research design. Since the boundaries between the case and the context may be unclear, incorrect case definitions or uncovering a more interesting case to investigate could occur as the study progresses. This is inherent to and not an uncommon in case study research, in which one properly redefines the case before adjusting and continuing the study. Analytical techniques include pattern search, explanation building, logic models, time-series analysis, cross-case comparisons and synthesis, etc. (Given, 2008; Yin, 2017).

Theory built from case studies are likely to be novel, testable and empirically valid (Eisenhardt, 1989). To do so, grounded theory (Given, 2008) have been utilized in combination with triangulation, a priori specification of constructs and giving existing literature a larger role (Eisenhardt, 1989).

The qualitative approach does not exclude the use of quantitative methods, it is for example common to conduct a survey based of initial findings and conduct a statistical test to corroborate and triangulate results.

5.1.3 Introducing the paper

The second paper is; “Makerspace sustainability enabled by crucial partnerships and a gentrification strategy”. It is a draft to be submitted to Journal of a Cleaner Production.

The qualitative approach taken in this paper encompasses a single in-depth case study of a business model on an ecosystem level. The context is the business model of a hardware centered startup incubator, which includes establishment of multiple partnerships. Semi-structured interviews with stakeholders, dialogue with a key informant, documents such as reports and news, were among the methods used in this investigation. In combination with a Customer-Value-Chain Analysis (Donaldson et al., 2006), results include description of a set of crucial partnerships in a much larger and complex ecosystem consisting of many distinct stakeholders necessary for business model sustainability.

5.2 The second paper “Makerspace sustainability enabled by crucial partnerships and a gentrification strategy”

The paper is enclosed in the following pages.

Makerspace sustainability enabled by crucial partnerships and a gentrification strategy

Henrikke Dybvik[a], Kristine B. Fredriksen[b], Martin Steinert[a].

[a] Norwegian University of Science and Technology, Richard Birkelands vei 2B, 7491 Trondheim, Norway.

[b] R. Kjeldsberg AS, Sluppenvegen 6, 7037 Trondheim, Norway.

Abstract

The phenomena of makerspaces continue to emerge as they are seen as particularly important platforms for facilitating hardware centered startups developing technological solutions to societal issues. Unfortunately, such spaces struggle to find sustainable business models. This article encompasses an in-depth explorative case study of a hardware incubator. Findings relate to product-service system (PSS) theory, specifically the advanced interaction types and value creation logic behind emerging value networks. An in-depth case study shows the applicability of PSS and suggest one possible sustainable business model for makerspaces.

Keywords: makerspace, incubator, value network, product-service system, sustainable business model, sustainability, hardware

1 Introduction

The phenomena of incubators and makerspaces emerges on a national and global scale, though it is not novel (Burke, 2015; Jensen, 2017; Slatter and Howard, 2013). A makerspace (Jensen, 2017; Slatter and Howard, 2013) is particularly important for hardware centered startups because they offer the infrastructure and equipment required for developing physical products. Hardware startups are seen as instrumental in developing new technological solutions to for instance complex industry and societal problems. Bankruptcy of TechShop-pioneer and the short lifetime of FabLabs displays makerspaces continued struggle to find a sustainable business model. A central issue is to maintain revenue while continuing to deliver sufficient value to the customer (Bocken et al., 2014). Interestingly, Bocken (2014) list incubator support models under an organizational business model development archetype.

Product-service-systems (PSS) are emerging in industry and promoted in research as a sustainable business model proving multiple benefits in the economic, environmental and social dimension when compared to traditional product or service-centered business models (Bocken et al., 2014). PSS is not fully shaped for practical applications and companies lacking a system approach (Mont, 2002). Research call for exploring the design side of PSS to develop methodological approaches for practical implementation. Academic rigor in new case studies of PSS in practice is called upon (Lindahl et al., 2014; Tukker and Tischner, 2006). Mougaard et al. (2013), highlight the creation of network-based

development models as an interesting research direction, as it relates to the emerging network management theory and current economical shifts in society. The service, sharing and experience economy are recognized to be relevant emerging economies (Hamari et al., 2016; Tukker, 2015, 2004).

This article encompasses an in-depth explorative case study of a hardware incubator. It displays one practical application for value networks, aiming at showing vast applicability and extending PSS theory. In doing so we suggest makerspaces look towards establishing partnerships exhibiting certain strong characteristics, crucial to a creating a positive net value sustaining a position in the value network. In other words, enabling a sustainable business model.

The remainder of the article includes the following. The second section presents the theoretical foundation; sustainable business models related to PSSs, value networks and urban planning. The third section further described the goal. Fourth section describes the case study research design, and data analysis. Findings are presented in the fifth section, including case and value network description. Discussion and conclusion follow in section six and seven.

2 Foundation

Sustainable business models are important since they can be used to coordinate technological and social innovations with system-level sustainability (Bocken et al., 2014). A PSS have a strong focus on the environmental and social aspect of sustainability, while maintaining a focus on business by offering revenue models to serve the firms economic interest. A PSS business model is one promising potential for business model sustainability (Bocken et al., 2014; Tukker, 2004; Tukker and Tischner, 2006). PSS have been researched (Bocken et al., 2014; Lindahl et al., 2014; Mont, 2002; Mougaard et al., 2013; Tukker, 2015, 2004; Tukker and Tischner, 2006) for the potential increase in economic, social and environmental advantages which overall can bring a company to a strengthened position in the value chain or on the global market.

A PSS is "a system of products, services, supporting networks and infrastructure that is designed to be: competitive, satisfy customer needs and have a lower environmental impact than traditional business models" (Mont, 2002). Other definitions emphasize the need for a distinct network and infrastructure (Tukker and Tischner, 2006), the utility of products and services (Tan et al., 2007), better differentiation from the competition, and better fulfillment of the customer demand and sustainability (Schenkl et al., 2014). Mont (2002) found that a successful PSS requires different societal infrastructure, human structures and organizational layouts to function in a sustainable manner. A value network can offer such characteristics.

In a value network *"(…) organisations support each other, based on operational needs. Instead of purely transactional interactions, where value is added in one direction and capital flows in the other,*

the network is characterised by mutually beneficial relationships or partnerships" (Andersen et al., 2013). The network type is strongly related to the business strategy and the underlying value creation logic. From product to result orientation there are three generic types; current, traditional, renewal and emerging networks (Andersen et al., 2013; Möller and Rajala, 2007).

The network management perspective (Möller and Rajala, 2007) assumes that stakeholders are embedded within networks of interconnected relationships. In emergent networks; companies, and networks of companies are seen as complex adaptive systems, comprising interacting sets of organizational and social relationships where each stakeholder can pursue their own goal, – while the interorganizational network pursue a larger set of goals.

Different combinations of interaction types constitute the different types of organizational and social relationships required for successful value networks. Each interaction type bring value to the activity. Therefore, there must be a specific combination of interaction types between the stakeholders for the network to properly support the business strategies and the value creating activities. As the purpose of the network moves from a traditional one towards renewing, emerging and novel business strategies more advanced interaction types and combinations are needed. Such crucial relationships are therefore enabled by multiple advanced interaction types. Advanced interaction types are illustrated by Anderson et al. (Andersen et al., 2013, pp. 18–19). For instance, if the network is to be capable of supporting an emerging and novel business, coordinated development of offerings and risk-sharing are among other advanced interactions needed (Möller and Rajala, 2007). Informal or personal relations are important, among other to build trust.

Furthermore, overcoming barriers and common drivers are also important for successful, strategic value networks. Common barriers are complexity, conflicts of interest, non-formalized trust, distribution of rewards and cultural differences (Andersen et al., 2013). Drivers have been thoroughly described in literature, for further reference we refer to Mont (2002), Tukker (2004) and Lindahl et al. (2014).

The business of real estate can include projects planning for sustainable urban development. This requires an orientation towards long-term goals, an environmental perspective, facilitate alliance-building, integrate social theory and focus on potential conflict resolution (Campbell, 1996; Næss, 2001). Gentrification is recognized as a global strategy (Smith, 2002) for urban planning, emphasizing the central role of real estate developers.

3 Goal

This article encompasses an exploration of the business model and the stakeholder ecosystem of a local hardware centered incubator. It aims at displaying the vast applicability of PSS by showing one practical application of value networks. Furthermore it suggests the makerspace community look

towards establishing crucial partnerships with certain characteristics in their pursuit for a sustainable business model.

To do so an in-depth explorative case study was conducted as it can explain and describe a contemporary phenomenon (Eisenhardt, 1989; Yin, 2017), generate novel theory which is empirically valid and with testable and falsifiable hypotheses (Eisenhardt, 1989), and because case studies are needed (Lindahl et al., 2014; Tukker and Tischner, 2006). The case was selected by theoretical sampling as it provided unusual research access (Yin, 2017) and as it was the only one with a hardware positioning, representing an extreme situation likely to extend emergent theory (Eisenhardt, 1989).

4 Method

As mentioned, this particular single case was chosen because it provided an opportunity for unusual research access (Yin, 2017). The incubator was the only one with a hardware positioning, therefore representing an extreme situation likely to extend emergent theory (Eisenhardt, 1989). An ongoing relation with the hardware community provided the researchers with full access to the physical infrastructure and a close collaboration with a key informant. An occurrence demonstrating the unique opportunity and mutual trust (Creswell, 1999), is that the researcher were unsolicited provided a keycard, gaining full access to the premises 24/7. Furthermore, this gave access to vital contact information of all stakeholders and by utilizing the key informant added incentive for their participation. An additional advantage were the proximity of the premises.

A multitude of data sources were collected and multiple data collection methods used to increase the validity of the study, support and triangulate the findings (Creswell, 1999; Eisenhardt, 1989; Yin, 2017). Early in the research design it was decided that the case study would employ both qualitative and quantitative evidence if possible. The combination of data types can be highly synergic (Eisenhardt, 1989) and a multimethod approach holds the potential for understanding complex social phenomena (Creswell, 1999; Yin, 2017). Data sources include a series of semi-structured interviews, a key informant, a survey, documents (project application, reports, startup contract, presentations, news, websites etc.), artifacts captured by photography, observations made on-site, and observing participation.

Explorative, in-depth semi-structured interviews were developed and conducted with 15 stakeholders. Interview subjects were thought of and treated to be ‘knowledgeable agents’ (Gioia et al., 2013). An interview procedure and questions template was developed as a part of the case study protocol (Yin, 2017). The verbal line of inquiry questioned the participant about their role, potential company, and how they became involved with the incubator, how they understood the ecosystem and the business model. Collaborating partners and interactions with other ecosystem stakeholders were investigated

along with related incentives, opportunities and challenges. The interview would usually end with an open question. The interview questions naturally changed as the research process progressed, both within-interview-setting and between each interview. This flexibility enabled a shift in focus and allowed probing on emergent themes and taking advantage of arising opportunities (Eisenhardt, 1989; Gioia et al., 2013; Yin, 2017). Interviews were recorded and transcribed for further analysis. One interview was conducted over the phone and notes taken during and after the interview.

Selection of stakeholders to interview was based on an initial mapping aided by key informant and corroborated with interview participant statements. Table 1 contains an overview over the interview subjects, their role and descriptive data. The relation to the incubator and network is illustrated in Figure 1. A total of 15 interviews were conducted. Multiple semi-structured interviews, conversations and informal interactions were conducted with the key informant.

Table 1: Overview of interview respondents. (Key informant not listed)

Title	Role	Date	Location	City	Length of interview
CEO	Private real estate developer	Aug 2018	Their office	Trondheim	65 min
2 Co-founders of technology company	Angel investors and experience entrepreneurs	Aug 2018	Their office	Trondheim	60 min
Investor	Investor	Aug 2018	Their office	Trondheim	53 min
CEO	Mature technology startup	Aug 2018	Phone call	Trondheim	45 min
CEO & Head of labs at Incubator	Partnering company	Aug 2018	Their office (the incubator)	Trondheim	90 min
CEO	Partnering company 2	Aug 2018	Their main office	Regional Trondheim	60 min
CEO	Brewery	Aug 2018	Their office and production facility	Trondheim	50 min
VP	Teleservice provider	Sep 2018	Their office in reception building	Oslo	85 min
Adviser at Rectors Staff Innovation	University	Aug 2018	Their office	Trondheim	45 min
CEO	University Technology Transfer Office	Aug 2018	Their office	Trondheim	50 min
Investment director	Industry Technology Transfer Office	Aug 2018	Their office	Trondheim	45 min

Former CEO, currently chairman of the board	Former CEO, currently chairman of the board at coworking-space	Aug 2018	Coworking-space	Trondheim	62 min
Advisor City Manager Staff	Municipality	Aug 2018	Café	Trondheim	40 min
CEO	Hardware startup at incubator	Sep 2018	Incubator 2.0	Trondheim	40 min
CEO	Hardware startup at incubator	Sep 2018	Incubator	Trondheim	30 min

Data collection and data analysis overlapped as expected in case study research attempting to build theory (Eisenhardt, 1989; Yin, 2017). The data analysis process was dynamic, iterative and continuously adapted to collected data and emerging theoretical concepts (Gioia et al., 2013). The flexibility principle fundamental to case study data collection (Eisenhardt, 1989; Yin, 2017) was taken advantage of also in data analysis. As the research progressed new insights emerged, and new theoretical constructs were introduced. Data analysis methodology was adjusted accordingly, and new methodology or tools especially targeted towards deep investigation of these constructs added.

The analytic techniques used were initial pattern-matching to facilitate explanation building (Yin, 2017). As the data evidence was revisited, the importance of stakeholder interaction became more prominent, leading to use of the logic model on an organizational level. Using this analytic technique, the cause-effect-cause-effect-patterns were scrutinized as interaction-effect-interaction-effect-patterns, which aided in identifying the value created for each stakeholder in a relationship.

NVivo (NVivo qualitative data analysis Software, 2018) was used for coding interview transcripts. 1st order nodes were coded using informant terms. These were themes emphasized as important by the informant, stakeholder interactions and relationships, concepts related to urban planning, and other frequent catchwords. Nodes were continuously added, similarities and differences identified to aggregate the number of 1st order nodes in an iterative manner. The aggregated nodes naturally stood out as the ones with most references and file references. These they were used for initial pattern recognition and carried over in the continued explanation building (Yin, 2017).

For retrospective or current analysis of stakeholder networks and business ecosystems a Customer Value Chain Analysis (CVCA) and actor network mapping can be used (Donaldson et al., 2006; Morelli, 2006). CVCA identify pertinent stakeholders or customers, their relationships, and role in the product's life cycle. It recognizes critical stakeholders and provides value proposition clarification. A CVCA was conducted in collaboration with the key informant according to steps one through five as described by Donaldson et al., (2006). The actor network (Morelli, 2006) emphasize the interaction levels as separate and important aspect of the socio-technical process of the development and

operation of a PSS (Morelli, 2006). It focuses on interactions between groups of stakeholders and direct or indirect individual stakeholder relationships.

Findings were triangulated using the other data sources, such as legal agreements and project applications, survey results, occurrences in news, online statements, artifacts, and conversations.

5 Findings

FAKTRY define themselves as *"a community for ambitious startups in hardware-centric disciplines and a playground for students, researchers and corporates. (…) Through our partners and friends we provide unique support to knowledge-based startups. Our members get access to capital, mentors, colleagues and customers." (Incubator webpage)*. They exhibit strong branding through infrastructure and visual style, and positioning. The hardware startups purchase a membership including offerings (e.g., access to dedicated space, IoT-lab, and workshop) and responsibilities (active community participation).

The incubator is located in an old industrial area two kilometers south of the university campus. The industrial area is to a large extent owned and developed by one single private real estate developer, and the area is a long-term regeneration project with a perspective towards 2050. The overall strategy for the area is a compact mixed-use area with homes, retail and office space close to a transportation hub. The company operating the incubator is owned by the real estate developer. The incubator CEO is employed by the real estate developer and contracted to the incubator as a consultant.

The real restate CEO and the key informant describe how from 2016 they investigated a hypothesis that a community developed through an incubator would add attractiveness to the area as a location for technology-based industry. The real estate developer strategically visited incubators and makerspaces with similar focus and setup (NewLab in Brooklyn, New York and Central Research Laboratory in Hayes, London) to learn and to be able to apply the same principles in their situation. Stakeholders from the university and potential other partners participated. The real estate developer had positive experiences from doing so in their previous regeneration project from the late 90's and possibly generate larger tenants as the startups from the incubator grow. A real estate developer depends on committed tenants in place to be able to start new development projects (Næss, 2001).Both the key informant and the CEO were very clear on their attempt to use positive gentrification as a strategy.

The overall value network is very large and complex, containing numerous stakeholders and a larger number of or relationships. The value network is illustrated in the form of an ecosystem map in Figure 1, displaying the value created from each relationships. All interaction types illustrated by Anderson et al., (Andersen et al., 2013; Möller and Rajala, 2007), were found in some form for all relationships. Here, the main value requirements for the network is fulfilled by having such a large set

of stakeholders and, social and organizational relationships enabling joint value creation. We also found multiple examples of advanced interaction types that added a sense of detail or finesse, transforming an ordinary business partnership to a highly synergetic relationship.

Interestingly, the interactions that were among the most mentioned were those we consider to be highly advanced and strongly related to the sharing economy. The interview participants, naturally, did not use terms such as platform sharing/pooling, co-development of offerings and risk sharing. Rather, collaboration, network and community, dialogue and personal relations reoccurred. Expressions such as “we know each other from before”, “we have discussed”, “the interest is personal” and “they’re old friends”, were frequently used to explain their diverging and congruent storylines, the occurrence of a partnership, how the partnership developed and current functioned.

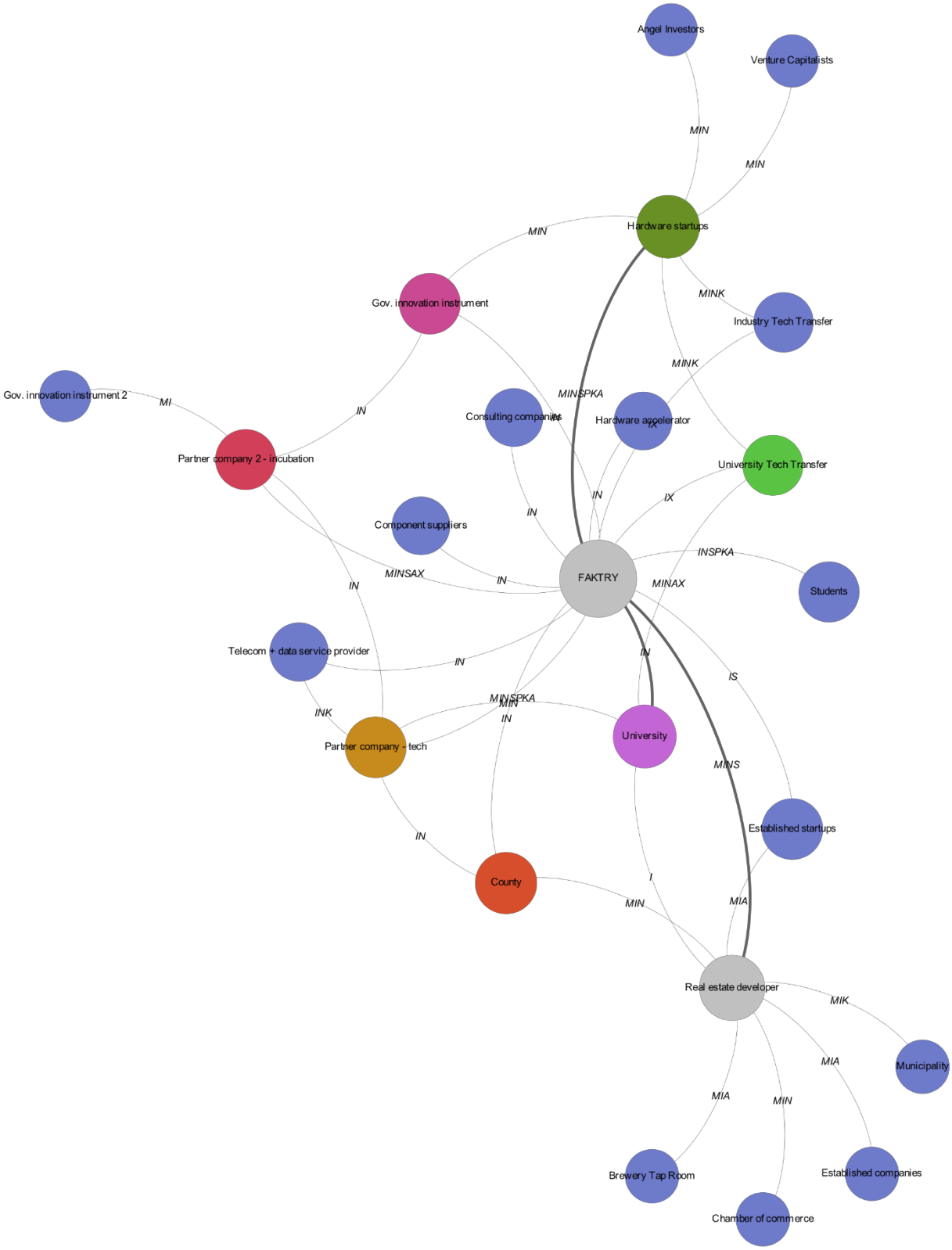
During the interviews there was more often than not, ambiguity to if the relationship with which the stakeholder spoke of was formally established or not, even if the term partner was used (which is defined in business literature). A range of formal, informal and personal interaction terms were mixed and interchanged of used, with no clear separation. We believe that this ambiguity displays how the combination of multiple and many advanced interaction types enabled the stakeholder relationships, which create the interacting sets of organizational and social relationships from value network theory. Moreover, it displays exactly how important they were for enabling multiple benefits that jointly created unique value - crucial for enabling and sustaining that particular relationship. There were three relationships seen as critical for the incubators sustainability in the successful setup of the value network.

5.1.1.1 The real estate developer relationship enable a long term perspective

Advanced interaction types are exemplified by platform sharing where access to infrastructure was simple and inexpensive for the incubator, while the real estate developer were able to utilize an otherwise unattractive and old building. Furthermore, the incubator CEO is employed by the real estate developer and contracted to the incubator as a consultant, given much autonomy from the real estate developer. Co-development of incubator infrastructure and branding, driven by joint interest in supporting technology development causes positive awareness. Young talent, knowledge and technology creates a significant ‘coolness’-factor aiding the gentrification strategy. Value is also created by joint desire for close, long-term and loyal customer relationships which they genuinely want to occur by providing genuine societal benefits to the customer.

Figure 1 The value network. Value from stakeholder relationship is indicated by the legend. Crucial relationships are indicated by thick lines.

5.1.1.2 The hardware startups and incubators joint positioning and relationship



The incubators strong hardware branding, combined with recognized hardware startups creates a unique positioning with specific characteristics. The resulting value is an impression of this-is-the-

best-place-for hardware startups and for stakeholders with complex challenges they wish to solve. Standing out in interviews was the value created for the startups by having a compact and extensive bundle of offering, including mentoring, network and specialized equipment, to mention a few. Startups co-develop offerings with the incubator and with each other by knowledge sharing and assistance in product and business development. Startups advocating these benefits to externals strengthens the position. *"There are some synergies (..), between what we do and the problems the other have (…). So, we can consult them regarding what needs they have, what their challenges are. And so on. An then there is, since it's hardware startups, what we do is… mechanical engineering or electrical powerengineering. You can have engineers there [at the incubator] that are experts in those fields, that can provide advice to what our engineers are doing."* (CEO of hardware startup). This relationship would not have been possible without the advanced interaction types enabling sharing and collaborative consumption (Mougaard et al., 2013), such as personal interaction, co-developments of product/service offerings and platform sharing.

5.1.1.3 Incubator closeness to and relation with university

Findings include an example of an Innovation Network (Möller and Rajala, 2007), guided by research objectives, open innovation and informal interactions. The Norwegian University of Science and Technology (NTNU), is the largest university in Norway with the highest technical education level. This, proximity, talented students, researchers and tech transfer offices, contributes with keeping the technical competencies of the hardware startups at the highest levels. There is a common vision for innovation, development and increase in knowledge. A coincidental hallway encounter at university campus exemplifies the importance of the informal encounters facilitated by personal relations The key informant simply being present caused an IoT researcher to suggest a collaboration, since the incubator contain an IoT lab. *"These random encounters occur on a frequent basis, - they are incredible! In that they often prove to be of value for research and business development. And us. This is great, this environment being so close and small enables a rapid understanding of which stakeholder can do what"* (Citation: Key informant after the mentioned encounter) Here, a common goal, co-development of offerings and platform sharing for research activities is displayed, and it is a great example of advanced interaction enabling an emerging innovation network.

6 Discussion

The value network differs from a 'normal' value chain in that multiple stakeholders support the same core activities, though they do not sit in the same value chain. Some are part of a completely different industry (Andersen et al., 2013) and multiple benefits still occur. One example is a brewery setting up production and a taproom in the area. The resulting value is a tenant for the real estate developer providing attractiveness to the area and beer for incubator events, while the brewery receive flexibility

in production set-up and are enabled to conduct a more customer-oriented product development through direct user feedback.

The common value network barrier trust (Andersen et al., 2013), was not found in any . The only non-Norwegian interview participant pointed to the Norwegian culture to answer stakeholder mutual trust, stating that the egalitarian view is strong, even embedded in the Norwegian culture. As she elegantly put it: *"Trust is never an issue in Norway"* (VP Teleservice provider). Network complexity could cause confusion for stakeholders' understanding of the overall network strategy. Here, the individual stakeholder had their own business strategy and goals defined, and since the relationship produced value in this regard this was not an issue for any of the stakeholders. Stakeholders need not concern themselves with, or even understand overall network goal, strategy nor complexity.

The case study is strengthened by number of interviews (n=15) and access to a key informant. It was valuable as it enabled rapid clarification of minor details. Recording and transcription ensures accurate representation of information. A limitation when building theory from single-case studies in an overly complex or narrow resultant theory (Eisenhardt, 1989).

7 Conclusions

The phenomena of makerspaces continue to emerge as they are seen as particularly important platforms for facilitating hardware centered startups developing technological solutions to societal issues. Unfortunately, such spaces struggle to find sustainable business models. This article encompasses an in-depth explorative case study of a theoretically sampled hardware incubator. Multiple data sources were collected, including semi-structured interviews (n=15) and a key informant, and data analysis methods used to increase validity. Findings display a value network with strong PSS characteristics. Multiple advanced interaction types creates highly synergetic relationships on an ecosystem level, and on an individual level. Three relationships were found to be crucial for the incubator, enabling a situation where the incubator is able to jointly produce and receive value which can sustain the incubators existence. We suggest that the makerspace community look towards establishing such crucial partnerships to find a sustainable business model. Based on these findings we formulate the following hypotheses:

H1: Relationships should be created by multiple advanced interaction types on an ecosystem level and on an individual level.

H2: Crucial individual level relationships should exhibit certain characteristics, a) being an integrated part of a long-term perspective, b) ensuring high technical competence and c) causing a strong and seemingly unique positioning.

Together, a positive net value is produced. Through this in-depth case study we have found the practical applicability of PSS and value networks to be vast and we encourage the PSS community to explore the findings, for instance by finding similar cases and testing the hypothesis.

8 Acknowledgements

This research is supported by strategic funds from the department for Mechanical and Industrial Engineering (MTP) at NTNU. We would also like to express our gratitude towards all interview participants for giving of their time and expertise.

9 References

Andersen, J.A.B., Mougaard, K., McAloone, T.C., i Mateu, A.G., Neugebauer, L.M., Hsuan, J., Ahm, T., 2013. PSS Partnerships: A workbook in the PROTEUS series.

Bocken, N.M.P., Short, S.W., Rana, P., Evans, S., 2014. A literature and practice review to develop sustainable business model archetypes. J. Clean. Prod. 65, 42–56. https://doi.org/10.1016/j.jclepro.2013.11.039

Burke, J., 2015. Making Sense: Can Makerspaces Work in Academic Libraries?

Campbell, S., 1996. Green Cities, Growing Cities, Just Cities?: Urban Planning and the Contradictions of Sustainable Development. J. Am. Plann. Assoc. 62, 296–312. https://doi.org/10.1080/01944369608975696

Creswell, J.W., 1999. Mixed-method research: Introduction and application. Handb. Educ. Policy 455–472.

Donaldson, K.M., Ishii, K., Sheppard, S.D., 2006. Customer Value Chain Analysis. Res. Eng. Des. 16, 174–183. https://doi.org/10.1007/s00163-006-0012-8

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. Acad. Manage. Rev. 14, 532–550. https://doi.org/10.2307/258557

Gioia, D.A., Corley, K.G., Hamilton, A.L., 2013. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organ. Res. Methods 16, 15–31. https://doi.org/10.1177/1094428112452151

Hamari, J., Sjöklint, M., Ukkonen, A., 2016. The sharing economy: Why people participate in collaborative consumption. J. Assoc. Inf. Sci. Technol. 67, 2047–2059. https://doi.org/10.1002/asi.23552

Jensen, M.B., 2017. Opportunities of Industry-Based Makerspaces: New Ways of Prototyping in the Fuzzy Front End.

Lindahl, M., Sundin, E., Sakao, T., 2014. Environmental and economic benefits of Integrated Product Service Offerings quantified with real business cases. J. Clean. Prod. 64, 288–296. https://doi.org/10.1016/j.jclepro.2013.07.047

Möller, K., Rajala, A., 2007. Rise of strategic nets — New modes of value creation. Ind. Mark. Manag., Opening the network - Bridging the IMP tradition and other research perspectives 36, 895–908. https://doi.org/10.1016/j.indmarman.2007.05.016

Mont, O.K., 2002. Clarifying the concept of product–service system. J. Clean. Prod. 10, 237–245. https://doi.org/10.1016/S0959-6526(01)00039-7

Morelli, N., 2006. Developing new product service systems (PSS): methodologies and operational tools. J. Clean. Prod., Product Service Systems: reviewing achievements and refining the research agenda 14, 1495–1501. https://doi.org/10.1016/j.jclepro.2006.01.023

Mougaard, K., Neugebauer, L., McAloone, T.C., Bey, N., Andersen, J.B., 2013. Collaborative Product/Service-Systems – On Conceptualisation of PSS Offerings and Business Nets, in: Shimomura, Y., Kimita, K. (Eds.), The Philosopher's Stone for Sustainability. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 227–232. https://doi.org/10.1007/978-3-642-32847-3_38

Næss, P., 2001. Urban Planning and Sustainable Development. Eur. Plan. Stud. 9, 503–524. https://doi.org/10.1080/713666490

NVivo qualitative data analysis Software, 2018. . QSR International Pty Ltd.

Schenkl, S.A., Rösch, C., Mörtl, M., 2014. Literature Study on Factors Influencing the Market Acceptance of PSS. Procedia CIRP, Product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems 16, 98–103. https://doi.org/10.1016/j.procir.2014.01.013

Slatter, D., Howard, Z., 2013. A place to make, hack, and learn: makerspaces in Australian public libraries. Aust. Libr. J. 62, 272–284. https://doi.org/10.1080/00049670.2013.853335

Tan, A., McAloone, T.C., Gall, C., 2007. Product/Service-System development: an explorative case study in a manufacturing company. Presented at the 16th International Conference on Engineering Design, Design Society.

Tukker, A., 2015. Product services for a resource-efficient and circular economy – a review. J. Clean. Prod., Special Volume: Why have 'Sustainable Product-Service Systems' not been widely implemented? 97, 76–91. https://doi.org/10.1016/j.jclepro.2013.11.049

Tukker, A., 2004. Eight types of product–service system: eight ways to sustainability? Experiences from SusProNet. Bus. Strategy Environ. 13, 246–260. https://doi.org/10.1002/bse.414

Tukker, A., Tischner, U., 2006. Product-services as a research field: past, present and future. Reflections from a decade of research. J. Clean. Prod., Product Service Systems: reviewing achievements and refining the research agenda 14, 1552–1556. https://doi.org/10.1016/j.jclepro.2006.01.022

Yin, R.K., 2017. Case study research and applications: Design and methods. Sage publications.

5.3 Discussing the qualitative approach – specifically the case study research method

This discussion concerns the methodological approaches used throughout the paper "Makerspace sustainability enabled by crucial partnerships and a gentrification strategy", namely a qualitative approach enabled by an in-depth case study. Experiences and reflections from using the method will be presented and discussed along with advantages and limitations. Key outtakes are emphasized.

5.3.1 The case study in general

The qualitative in-depth-case study provides an excellent opportunity to explore a broad research question, or even interest in a topic in general. In this case, there were a range of perspectives that were possible to take and theoretical concepts and research domains to be considered, which is common in qualitative research (Yin, 2017). This renders possible a variety of research questions, which initially made it challenging to focus on one particular framing and trail of thought.

A case study protocol was developed as is recommended (Yin, 2017), which aided in foreseeing potential problems, training and preparing the researcher (the author) to conduct the study. As the study progressed it was helpful to look back at this document to retain focus on the topic and renew skills.

The attention to detail is important in qualitative work and mastery of this ability constitutes some of the finesse inherent to good qualitative work. Details are what provide a rich and full picture since a situation would not have been the exact same without these. Selecting and discarding minor details to create the better picture is an interesting challenge for the researcher, as seemingly unimportant details can be crucial while other details are non-vital to explaining a phenomena and therefore should not have much attention devoted to them. The attention to correct details is an ability under continuous development.

5.3.2 Data collection

Data collection is time consuming in qualitative research (Johnson and Onwuegbuzie, 2004; Yin, 2017), due to many data sources and data collection methods. Moreover, it can be challenging to assess when to stop data collection. Data collection freeze is a difficult assessment to make, as each source can contain new information, strengthening the study and contributing to creating a more complete database. The author experienced this as a

duality. It was a pleasure of doing qualitative work, since the constant evolvement of the research process created a dynamic research arena where *not everyday is the same*. On the contrary, when it comes to writeup there in an enormous amount of information and detail that is both possible and interesting to illustrate, which there unfortunately is not space for in a paper with word limitations. It quickly became crystal clear to the author why many in-depth case studies are communicated through several volumes of many hundred pages (Yin, 2017).

5.3.3 Bias - limit to ensure validity - essential to qualitative inquiry

Bias is an inherent part of qualitative work. Some degree of empathy is required from the researcher to be able to understand the interview participant or other human subjects' point of view, which in turn affects the researcher. Since a researcher should investigate a phenomena objectively[2], the author considers this to be a one of the greater disadvantages to qualitative research. If the researcher is not cautious a potential bias can be large and there is a risk of presenting incorrect or plain wrong results. The author has strived to the best of her ability to limit both the case and the subjects' effect on herself and not affect or alter the case or the human subjects in the study. This is an ongoing challenge, since the author is positive and positivistic by nature. The author's performance in attempting at being somewhat neutral is regarded to have been good, though by no means exceptional.

5.3.4 Conducting semi-structured interviews

Semi-structured interviews are inherently flexible and adaptive in nature. This adaptation to concepts and human subjects is what makes interpretive research good at uncovering new concepts and theories (Eisenhardt, 1989; Gioia et al., 2013; Yin, 2017). The interview questions naturally changed as the research process progressed and theories developed, both within-interview-setting and between each interview. This flexibility principle enabled shifts in focus and allowed the author to zoom in on interesting themes that appeared during the interview (Gioia et al., 2013; Yin, 2017). Arising opportunities were taken advantage of by probing in on emergent themes (Eisenhardt, 1989). This is reckoned to be a strong advantage of qualitative research (Eisenhardt, 1989; Johnson and Onwuegbuzie, 2004; Yin, 2017), to which the author also agrees. To do so, the interviewer

[2] This is a contradictory statement in itself, since one cannot observe an entity without interfering with it. Therefore, one strives to maintain an objective a position as possible.

must adapt to the interview participant to be able to shift the line of inquiry, in other words the interviewer must allow the interview subject to influence them. This is a variety of bias, to which it is important to keep the abovementioned discussion in mind. Depending on the attitude of the interview participant and to which they answered the questions in the mental line of inquiry, the author had very different experiences. For example, an experience of becoming energized occurred when speaking to participants providing good and interesting answers, who seemed knowledgeable in their field, and to some extent mirrored the authors' behavior. Interview participants who had a completely different attitude or behavior, or provided confusing answers by bringing in other elements caused a feeling of drained physical and mental capacity with the author. Yin (2017) stresses that in-depth interviews are both mentally and physically exhaustive in nature, that one must be aware and adapt to it and that mastering this process requires practice and perseverance.

5.3.5 The knowledgeable agent assumption

Interview subjects are treated as knowledgeable agents (Gioia et al., 2013). This is necessary for parts of analysis in developing theory from case studies. An additional advantage easing the analysis process is that the researcher does not have to question the *truth* of what one is being told by the interview participants, - at least to an extreme extent.

In a case study it is important to meet situations and interview participants with an open mind ready for their interpretations. In an attempt at remaining open to participants interpretations the author found it challenging at times to distill those who knew a topic well from those who do not. Moreover, deciding what amount of attention and emphasis should be devoted to each of these situations. In some cases, a novice eyes can provide a distinct point of view which turns out to be vital to interpretations later on, while in other cases it causes confusion, - especially if the topic is one the researcher does not hold highly detailed knowledge on. Similarly, an expert in a field may very well provide rich details on a topic, but here it can be difficult to distill their own interpretations from textbook or journal knowledge. Depending on the topic and the overarching research question this can be an advantage or a limitation.

5.3.6 A failed attempt at incorporating a quantitative method

Early on in the process it was decided that a survey would be created, the intention was to provide some quantitative metrics strengthening the study. Findings from interviews would

be used to see which stakeholders had what common motivations and goals for going into certain partnerships and the overall network they were a part of.

The author attempted at using the findings from the interviews to develop a survey investigating the stakeholders underlying motivations relating to PSS theory (Mont, 2002; Tukker, 2004; Tukker and Tischner, 2006), since this is the initial theoretical framing. A survey was developed incorporating a synthesis of multiple common PSS benefits and approximately three iterations made before sending the survey to stakeholders. Laster, it was discovered that the questions and statements in the survey did not fully capture the construct the author attempted at measuring (Abowitz and Toole, 2010). Some (not all) of the responses indicated both misunderstandings in that respondent did not answer to what the question intended to ask, and contradictory motivations in that answers regarding motivations and benefits was clearly different from what was stated during interviews or plain wrong. Despite several iterations, it was not enough to ensure measurement validity. Therefore, the data was not carried over in the form of descriptive statistics and statistical tests, as was intended. In the end it was used for triangulation purposes since it provided a more composite picture of the stakeholders interpretations and perceptions. A systematic error would have been transferred and produced a result with errors (Abowitz and Toole, 2010) if survey analysis had been conducted according to intentions. This illustrates several important considerations to make to ensure research validity when mixing methods. It is important to firstly ensure measurement validity through proper operationalization of hypotheses or definition of the constructs to be measured. Secondly, to be critical to data collected and to return and compare those to previous findings. Thirdly, to continue to reflect upon if the method selected serves the purpose and if it is producing the output required for continuing to use it. Further considerations include if the method should continue, be stopped and replaced with another, more suitable method.

6 The quantitative approach (paper 3)

6.1 Introduction to the quantitative approach – statistical inference

This section introduces the third article, which by conducting statistical tests of differences between two independent groups have a quantitate approach to investigating companies use of a business model canvas. It aims at placing the paper context in the overall business model context and provide details on the quantitative approach, particularly regarding statistical interference and testing.

6.1.1 The business model phenomena on a company level and from an element perspective

The business model in itself can be the competitive advancement a company have to remain competitive in the market. For example; Chesbrough (2010) states that *a mediocre technology pursued within a great business model may be more valuable than a great technology exploited via a mediocre business model*. Morris (2005) states that the business model can aid as a focusing device for entrepreneurs and employees in a company. These reasons fuel much of research on business models, which is challenging due to a number of reasons, on of which is the lack of a common definition or any other means of benchmarking. Using the same model or template across companies could aid in understanding what it is that separates the various business models and perhaps provide an indication to what it is that makes some companies and business models more successful than others. As described in the introduction, the Business Model Canvas and the Lean Canvas have been widely adopted. It may be due to ease of use and theoretical grounding in the business model element research perspective (Ritter and Lettl, 2018). The ease of which a business model can be developed by filling information in nine building blocks may be alluring to many. Of course, it is the managerial decisions and corresponding actions one must conduct to fulfill what one has written that actually enables competitive advancements.

Using such canvas templates to investigate if and how business models change and develop can be seen as a natural consequence since business model development is their intended purpose and such tools are said to facilitate business model development. Wide recognition and adoption is an advantage her as it provide measurements on the same metrics (information in each of the elements) for a multitude of companies. In turn, this enable a

quantitative and empirical research approach, which business model research has called for. This can aid our understanding of the business model phenomena on a company level, both in investigating the workings of a single company and making comparisons across companies.

6.1.2 The quantitative approach

A quantitative research approach in general is any empirical investigation collecting, analyzing and displaying data in numerical form (Given, 2008) rather than the narrative form inherent to much qualitative research. Quantitative research is often seen as opposite to qualitative research though the two approaches overlap, for example when qualitative research attempt at quantifying by using terms search as *sometimes*, *often* and *never*. Techniques used in quantitative research includes statistical, mathematical and computational techniques.

Statistical methods are designed to aid the process of making scientific judgements when faced with uncertainty and/or variation. Statistics uses fundamental mathematical laws of probability and statistical inference to draw a conclusion about a system represented by a collection of data, a sample. Descriptive statistics can be used to describe data contained in a sample in a condensed way. Associations, correlations and relationships between variables, predictions, group differences and reliability can be explored and tested by using a variety of statistical tests and techniques (Walpole, 2012). Conclusions from statistical interference are often used in decision and policy making, medical testing, process and manufacturing quality and more (Gonick, 1993; Walpole, 2012). Determining statistical interference involve data collection by making observational or experimental studies, developing one or more hypotheses regarding data information or interpretations, before applying a suitable statistical test. Reporting level of significance, confidence interval estimation, effects size, potential errors in addition to sampling procedure and the statistical test selected is common to illustrate to what the degree the study is reliable. It is worth noting the growing interest in big data analysis since large sample sizes have higher reliability than small ones provided data is free from error and bias (Kaplan et al., 2014).

6.1.3 Introducing the paper

The third paper is; “A Dynamic and a Static Approach to the Business Model - Investigating the potential difference in business model focus”. It is submitted to ICE-Conference 2019.

It encompasses a quantitative research approach where quantitative data from two independent samples (in total 34 companies) were statistically tested for differences. The context is lean business model development, investigating potential differences between companies who have a dynamic approach to their business model development and companies with a static approach conducting only initial business model planning. To do so, information on *actions conducted* in nine different elements in a business model canvas tool have been utilized.

6.2 The third paper: “A Dynamic and a Static Approach to the Business Model - Investigating the potential difference in business model focus”

The paper is enclosed in the following pages

A dynamic and a static approach to the business model

-

Investigating the potential difference in business model focus

Yngve Dahle
Norwegian University of Science and Technology (NTNU)
Faculty of Engineering (IV)
Norway
yngve.dahle@ntnu.no

Martin Steinert
Norwegian University of Science and Technology (NTNU)
Faculty of Engineering (IV)
Norway
martin.steinert@ntnu.no

Henrikke Dybvik
Norwegian University of Science and Technology (NTNU)
Faculty of Engineering (IV)
Norway
henrikke.dybvik@ntnu.no

***Abstract*—This white paper is conducted for Lean Business encompassing a first attempt at empirical testing of data extracted from the Lean Business database. Lean Business adhere to the Lean Startup Movement, where the core principle is that static business planning should be replaced with a continuous dynamic business model development and that doing so increases chances of success. To begin to understand lean startup and entrepreneurial behavior the potential differences in focus on the different business model canvas elements have been investigated. Based on two samples, one for dynamic use and one for static use of the business model canvas, differences were investigated statistically. A distinction in focus on the different business model elements between a dynamic and a static approach could not be found. However, statistical testing of quantitative data represents an important step towards understanding entrepreneurship.**

Keywords— Business Model; Business Model Development; Lean Startup; Business Model Canvas; Lean Canvas; Quantitiative Data.

I. INTRODUCTION

This is a white paper conducted for Lean Business encompassing a first attempt at empirical testing of data extracted from the Lean Business database, the Entrepreneur Platform. An in-depth depth description of the Entrepreneurship Platform and database development can be found in previous work [1]. The Entrepreneurship Platform has been built as a structured model for entrepreneurship consisting of a clearly defined terminology [1]–[3] and it is a part of an attempt at developing a quantitative and longitudinal approach to entrepreneurship research. In an attempt to further understand entrepreneurship, entrepreneurial and startup behavior, this paper begins an empirical investigation on quantitative company data regarding the business model [4]–[6].

Lean Business adhere to the Lean Startup Movement, which advocates an agile behavior in terms of business model development and product development. The core principle here is that static business planning or development should be replaced with a continuous dynamic development, and that by doing so increasing the chances of success [1], [7]–[10]. In the Entrepreneurship Platform [1], Lean Business incorporate and illustrate these notions as follows; *Companies with frequent changes in their Business Idea and Model, Project Development and Customer Interaction will have a greater chance for success:*

$$\Delta S = \Delta f(BI) + \Delta f(BM) + \Delta f(PD) + \Delta f(CI)$$

Where S = Success, BI= Business Idea, BM = Business Model, PD = Project Development and CI = Customer Interaction. This means that any positive change in BI, BM, PD and CI improves the level of Success. It is of Lean Business interest to investigate if there are differences between companies, organizations or entrepreneurs that iterate and change often, (that is; have a dynamic approach to developing a business) and those who do not change as frequently (that is; have a static approach). This can increase our understanding of what it is that make companies, organizations, entrepreneurs, or even startups successful. Moreover, it could begin to investigate if the underlying principles in the Lean Startup Movement hold.

To begin such an investigation of differences between a dynamic and a static approach, this paper focuses on investigating the Business Idea and Model. In the Entrepreneurship Platform, the Business Idea is a part of the Business Model, which consists of nine elements. These nine elements, from here on out called the business model elements have been created by combining the elements from the Business Model Canvas of Osterwalder and Pigneur [5], [9] and the Lean Canvas created by Maurya [8]. The elements are; KeyContribution, KeyMarket, Distinction, EarlyMarketCustomer, UniqueValueProposition, ProductFeature, Partner, HowToSell and HowToGetPaid, and

978-1-5386-1469-3/18/$31.00 ©2018 IEEE

they contain similar information as the Business Model Canvas and Lean Canvas. We refer to Dahle et al. [1], [2] for a thorough description of their development and what they encompass. A sample have been constructed to represent a static approach and a dynamic approach to using the nine business model elements in the Entrepreneurship Platform. Case companies were extracted from the Lean Business database and information on actions conducted in each of the nine elements analyzed. The amount of actions conducted in each business model element is used as a proxy for the attention and time a company spend on that element, that is to what extent they focus on that element. The hypothesis is that focus on the business model elements will be different in companies with a static approach compared to companies with a dynamic approach. Therefore, the amount of actions in each Business Model Element have been statistically tested for differences between a companies representing a static and a dynamic approach.

Following this introduction to the remainder of the paper consists of a theoretical background, where the basic assumptions of the Lean Startup Movement in addition to theoretical underpinning of those are presented. The research question is stated and operationalized into testable hypotheses before presenting the results. Lastly, results are discussed, including limitations before providing concluding remarks are made.

II. THEORETICAL BACKGROUND

Lean Business adhere to the Lean Startup Movement, which is based on notions from Blank's "Customer Development process" [7] that were incorporated Ries' "Lean Startup" methodology [10]. Furthermore, it utilizes ideas and tools from Maurya's "Running Lean" [8] and "Business Model Generation" by Osterwalder and Pigneur [5], [9].

In a Lean Startup, an agile behavior in terms of business model development and product development is promoted, aided by iterations and learning as fast as possible. Mantras such as "learn fast, fail fast" [8] and "fail early, fail cheap" occur frequently and describes the mindset and methodology to the Lean Startup Movement. The core principle is that static business planning or development should be replaced with a continuous dynamic development and that by doing so increasing the chances of success. Therefore, the success of a startups depends on the following abilities inherent to the startup. First, a startup must have the ability to change the business idea and subsequently its business model this is necessary [1]. This needs to be recognized by the entrepreneurs, who have to make necessary changes accordingly. Key to the process of recognition is seeking feedback through frequent customer interaction and continuously iterating on business offerings and business model by incorporating feedback. In an effort to make changes in the business model easier and facilitate the business model development, Ostewalder and Pigneur [5], [9] and Maurya [8] have developed a visual one-page tool. These tools, labeled Business Model Canvas and Lean Canvas respectively, have been widely accepted and adopted due to their flexibility. Both described the business model as a series of elements and have a strong focus on the interrelations of elements, seeing how conducting changes in one element affects the other. Making decisions and taking actions necessary to realize what is described in the canvas is what should cause success. As mentioned in the introduction, Lean Business combines these two tools in their Entrepreneurship Platform, which provides a visual tool with a defined terminology [1], [2], a canvas available for startups and entrepreneurs.

Theoretical foundation for investigating differences in a dynamic and a static approach can be found in an argument of investigating a company's dynamic capabilities [11] and dynamic entrepreneurial learning capabilities [12]. An ability to dynamically adapt to changing customer and market requirements is necessary to sustain a position in the market place and to continue to serve value to the customer [4], [6]. This can be aided by experimentation, organizational change or product development, and should be reflected in the business model. Business model development represent changes in the business model. Therefore, indications of a company's dynamic capabilities could be provided by understanding the elements, the relations between the elements and how changes affect one another. Using the description of a business model as a series of elements and interrelations is in line with the business-model elements research perspective (Ritter and Lettl, 2018), which is an additional theoretical grounding of business model research. The ability to learn and change are likely to be among the most important capabilities a firm can possess and therefore empirical attention should be devoted to the topic [11], [13].

III. THE RESEARCH QUESTION AND OPERATIONALIZATION

A. The Research Question

Our interest is to investigate whether there are differences between the case companies have a dynamic approach to business model and the ones that have a static approach to business model, with regard to their focus on the different Business Model elements. This led to an overarching research question.

Research Question:
Do entrepreneurs or companies that have a dynamic approach to the business model focus on other elements of the business model than companies that have a static approach?

B. The Hypotheses

The research question stated above is operationalized into testable hypotheses. To create hypotheses so, the ratio of actions conducted in the nine business model elements have been used as a proxy for the focus of the entrepreneur. Therefore, we have used the percentage of actions conducted in the nine business model canvas elements. This percentage-value corresponds to the ratio of actions in the element, which we argue could indicate the time and attention the entrepreneur devotes to that specific element. For all nine business model elements, the null hypotheses and alternative hypotheses are stated below.

KeyContribution:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "KeyContribution-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "KeyContribution-element" between dynamic and static use of the canvas.

KeyMarket:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "KeyMarket-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "KeyMarket-element" between dynamic and static use of the canvas.

Distinction:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "Distinction-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "Distinction -element" between dynamic and static use of the canvas.

EarlyMarketCustomer:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "EarlyMarketCustomer-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "EarlyMarketCustomer -element" between dynamic and static use of the canvas.

UniqueValueProposition:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "UniqueValueProposition-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "UniqueValueProposition-element" between dynamic and static use of the canvas.

ProductFeature:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "ProductFeature-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "ProductFeature-element" between dynamic and static use of the canvas.

Partner:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "Partner-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "Partner-element" between dynamic and static use of the canvas.

HowToSell:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "HowToSell-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "HowToSell -element" between dynamic and static use of the canvas.

HowToGetPaid:

H_0: There is not a statistically significant difference in the percentage of actions conducted in the "HowToGetPaid-element" between dynamic and static use of the canvas.

H_A: There is a statistically significant difference in the percentage of actions conducted in the "HowToGetPaid -element" between dynamic and static use of the canvas.

C. Sampling two Independent Groups

From the database we sampled cases for statistical analysis. The total database population was separated in two independent groups, the static group and the dynamic group. The static group contained all companies having used the canvas over the course of one, initial 24-hour period. The dynamic group contained all other companies, which used the canvas over multiple 24-hour periods. An additional requirement was the actual existence of the company, which had been manually controlled by the database operators.

Random sampling of static cases:

From the static group, a random sample of 17 companies was made. Company data for each of the companies in the sample was checked again, to ensure that the assumptions were met.

Stratified sample for dynamic cases:

From the dynamic group, the companies having the most 24-hour periods were selected. This translates to those companies having revisited and iterated on the canvas the most. Company data for each of the companies in the sample was checked again, to ensure that the assumptions were met.

IV. Results

The results from the statistical tests are presented here. Data from a total of 34 cases (17 for dynamic sample and 17 for static sample) were analyzed in SPSS Statistics [14] to investigate the potential statistical differences in percentage of actions conducted in the nine business model canvas elements between dynamic and static use. The percentage of actions conducted in each of the elements was compared between the dynamic group and the static group. Differences in percentage-values between the two groups were the foundation for the statistical tests. Statistical tests were chosen based on the characteristics of the data, i.e. outliers, normal distribution, homogeneity of variances, and distribution shape. Independent-Samples T-Test was used for normally distributed data, without outliers and exhibiting homogeneity of variances. For data violating homogeneity of variances Welch T-Test was used. For data violating the assumption of outliers or normality Mann-Whitney U Test was used. Outliers are defined by SPSS Statistics as values more than 1.5 box-lengths from the edge of a box in a box plot. The box plots were visually inspected. Shapiro-Wilk test for normality

was used to assess whether data were normally distributed, where significance values larger than 0.05 indicate a normal distribution. Similarly shaped distribution was inspected visually using histograms. An exact sampling distribution was used for U [15]. Independent-Samples T-Test and Welch T-Test evaluates differences in means between independent groups. Mann-Whitney U Test evaluates differences in medians between independent groups, if the groups have a similar distribution shape. Table 1 contains descriptive statistics. Table 2 contains metrics associated with assumptions deciding which statistical tests to use, along with the corresponding test and result. As shown in Table 2, the nine elements are not statistically significant. Thus, the alternative hypotheses are not accepted, and the null hypotheses are retained.

TABLE I. DESCRIPTIVE STATISTICS

	Dynamic			Static				
Variable (BM Element)	**Samples**	**Mean ± SD**	**Median**	**Samples**	**Mean ± SD**	**Median**	**Difference Mean ± Std. Error**	**Difference Median**
KeyContribution	17	24.7% ± 12.8%	23.1%	17	21.7% ± 8.3%	21.1%	2.9% ± 3.7%	2.0%
KeyMarket	17	17.2% ± 5.6%	17.2%	17	20.8% ± 10.8%	16.7%	-3.6% ± 2.9%	0.5%
Distinction	17	21.0% ± 11.6%	19.5%	17	22.8% ± 6.9%	22.2%	-1.7% ± 3.3%	-2.7%
EarlyMarketCustomer	15	5.6% ± 2.8%	6.3%	15	6.2% ± 3.5%	5.6%	-0.7% ± 1.2%	0.7%
UniqueValuePropositi on	15	6.6% ± 4.6%	6.1%	16	5.9% ± 2.7%	6.4%	0.7% ± 1.3%	-0.3%
ProductFeature	14	7.4% ± 4.2%	6.6%	13	7.4% ± 3.6%	6.8%	0.0% ± 1.5%	-0.2%
Partner	17	9.3% ± 6.8%	7.9%	13	6.8% ± 3.6%	5.6%	2.5% ± 2.1%	2.3%
HowToSell	16	7.3% ± 9.2%	5.2%	16	7.2% ± 3.0%	8.1%	0.2% ± 2.4%	-2.9%
HowToGetPaid	15	4.6% ± 3.6%	3.7%	16	5.9% ± 3.3%	4.6%	-1.3% ± 1.2%	-0.9%

TABLE II. STATISTICAL TESTING FOR DIFFERENCE IN PERCENTAGE OF ACTIONS CONDUCTED IN THE BUSINESS MODEL (BM) ELEMENTS, BETWEEN DYNAMIC AND STATIC SAMPLE

	Dynamic		Static					
Variable (BM Element)	**Outliers**	**Normality (Shapiro-Wilk test)**	**Outliers**	**Normality (Shapiro-Wilk test)**	**Homogeneity of variances (Lavenes's Test for Equality)**	**Similarly shaped distributions**	**Statistical test**	**Sig. (2-tailed)**
KeyContribution	Yes	No (p<0.05)	Yes	Yes (p=0.198)	N/A	Yes	Mann-Whitney U Test	0.563
KeyMarket	No	Yes (p=0.968)	No	Yes (p=0.074)	No (p<0.05)	N/A	Welch t-test	0.236
Distinction	Yes	No (p=0.001)	Yes	Yes (p=0.238)	N/A	Yes	Mann-Whitney U Test	0.193
EarlyMarket-Customer	No	Yes (p=0.322)	No	Yes (p=0.422)	Yes (p=0.506)	N/A	Independent-Samples T-Test	0.578
UniqueValue-Proposition	Yes	Yes (p=0.051)	No	Yes (p=0.606)	N/A	Yes	Mann-Whitney U Test	0.83
ProductFeature	No	Yes (p=0.107)	Yes	Yes (p=0.108)	N/A	Yes	Mann-Whitney U Test	0.905
Partner	No	Yes (p=172)	No	No (p=0.020)	N/A	Yes	Mann-Whitney U Test	0.385
HowToSell	Yes	No (p<0.001)	No	No (p=0.44)	N/A	Yes	Mann-Whitney U Test	0.287
HowToGetPaid	Yes	No (p=0.006)	No	No (p=0.031)	N/A	Yes	Mann-Whitney U Test	0.358

V. DISCUSSION

As the results describe statistical tests for all nine variables, i.e., the nine business model elements constituting the business model turned out not be statistically significant. This was based on two samples, one stratified sample for dynamic use of the business model canvas and one random sample for static use of the business model canvas.

As such, a distinction in focus on the different business model elements between those who use the business model canvas in a dynamic manner, compared to those who used the canvas only once could not be found. This might be due to a similar focus among entrepreneurs with a dynamic and iterative approach and entrepreneurs with a static approach to the business model. Their consideration of what the important elements are might be similar and therefore both groups have devoted similar attention to it. This could indicate that the behavior is not so different in the two groups of entrepreneurs in this aspect of developing a business. One can also speculate if this is an appropriate way to distinguish between companies who exhibit a lean behavior and those who don't. There could be

other aspects that are more representative of a lean behavior and therefore more interesting to investigate.

The study is limited by a small sample selection. Though results were not statistically significant, the sample size does not allow for generalizations had this been the case.

Despite the limitations, it is shown how one can begin to use data from the Entrepreneurship Platform in combination with existing entrepreneurial and business model research to further investigate entrepreneurs and startups. We do believe that careful statistical testing as we have conducted it can be used to analyze entrepreneurial behavior provided that the research question is properly operationalized into testable hypotheses. As such, it is and represents a step towards understanding entrepreneurship.

VI. Conclusion

This white paper conducted for Lean Business encompassing a first attempt at empirical testing of data extracted from the Lean Business database. Lean Business adhere to the Lean Startup Movement, which advocates an agile behavior in terms of business model development and product development. Here, static business planning or development should be replaced with a continuous dynamic development, since it increases the chances of success. It is of Lean Business interest to test if the principle holds in a series of quantitative studies based on data collected from their database. This paper begins such an investigation by examining potential differences between companies with frequent changes, that is one with a dynamic approach, and companies who do not change frequently, that is a static approach. A sample have been constructed to represent a static approach and a dynamic approach to using the Business Model elements in their online platform. Information from case companies were extracted from the database, and statistically tested for differences. Statistical tests investigating differences in mean and median for independent groups were conducted. Statistical tests for all nine elements constituting the business model turned out not statistically significant. As such, we did not find a distinction in focus on the different business model elements between those who use the business model canvas in a dynamic manner (by an iterative development) compared to a static manner (single occasion use of the canvas). The statistical analysis conducted as a small, yet important step as a starting discussion for how data and empirical evidence can aid in understanding business modelling, how business models develop and by doing so, understand entrepreneurship.

Acknowledgment

We would like to thank Lean Business for access to data.

References

[1] Y. Dahle, A. N. Duc, M. Steinert, and P. Abrahamsson, "Building an entrepreneurship data warehouse," in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Funchal, 2017, pp. 100–110.

[2] Y. Dahle, A. N. Duc, M. Steinert, and R. Chizhevskiy, "Resource and Competence (Internal) View vs. Environment and Market (External) View When Defining a Business," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018, pp. 1–9.

[3] A. Nguven-Duc, Y. Dahle, M. Steinert, and P. Abrahamsson, "Towards Understanding Startup Product Development as Effectual Entrepreneurial Behaviors," in *Product-Focused Software Process Improvement*, vol. 10611, M. Felderer, D. Méndez Fernández, B. Turhan, M. Kalinowski, F. Sarro, and D. Winkler, Eds. Cham: Springer International Publishing, 2017, pp. 265–279.

[4] H. Chesbrough, "Business Model Innovation: Opportunities and Barriers," *Long Range Plann.*, vol. 43, no. 2, pp. 354–363, Apr. 2010.

[5] A. Osterwalder, Y. Pigneur, and C. L. Tucci, "Clarifying Business Models: Origins, Present, and Future of the Concept," *Commun. Assoc. Inf. Syst.*, vol. 16, p. 28, 2005.

[6] C. Zott and R. Amit, "Business Model Design: An Activity System Perspective," *Long Range Plann.*, vol. 43, no. 2, pp. 216–226, Apr. 2010.

[7] S. Blank, "Why the lean start-up changes everything," *Harv. Bus. Rev.*, vol. 91, no. 5, pp. 63–72, 2013.

[8] A. Maurya, *Running lean: iterate from plan A to a plan that works*. O'Reilly Media, Inc., 2012.

[9] A. Osterwalder and Y. Pigneur, *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons, 2010.

[10] E. Ries, *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books, 2011.

[11] J. Barney, M. Wright, and D. J. Ketchen, "The resource-based view of the firm: Ten years after 1991," *J. Manag.*, vol. 27, no. 6, pp. 625–641, Nov. 2001.

[12] O. Jones, A. Macpherson, and D. Jayawarna, *Resourcing the start-up business: Creating dynamic entrepreneurial learning capabilities*. Routledge, 2013.

[13] G. George and A. J. Bock, "The Business Model in Practice and its Implications for Entrepreneurship Research," *Entrep. Theory Pract.*, vol. 35, no. 1, pp. 83–111, Jan. 2011.

[14] *IBM SPSS Statistics 25*. NY: IBM Corp.: IBM Corp., 2017.

[15] L. C. Dinneen and B. C. Blakesley, "Algorithm AS 62: A Generator for the Sampling Distribution of the Mann- Whitney U Statistic," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 22, no. 2, pp. 269–273, 1973.

.

6.3 Discussing the quantitative approach and statistical inference

This discussion concerns the methodological approach used throughout the paper "A Dynamic and a Static Approach to the Business Model - Investigating the potential difference in business model focus", namely a quantitative research approach enabled by statistical testing of company data. Experiences and reflections from using the method will be presented and discussed, including known advantages and limitations. Some key outtakes are emphasized.

6.3.1 Quantitative research in general

The ability to conduct statistical tests on quantitative data is one of the most valuable advantages of quantitative research as seen by the author. Revealing potential correlations and relationships, or non-existing relationships can aid greatly in understanding how any construct work. One could say that statistical testing limits "incorrect results" since one commonly discloses exactly how certain one can be that the results presented are correct and have not occurred by chance (Walpole, 2012).

By incorporating objective instruments for data collection, we are able to objectively measure and assess a phenomena, which minimizes bias from the researcher. Incorporating objective instruments, for example a sensor with continuous sampling frequency, can potentially bring in large amounts of data. In turn this could provide thorough and accurate research results since it is based on a larger sample. Result from such a statistical analysis would be closer to reality, provided that error and bias is minimized by ensuring data reliability and conducting the statistical procedure correctly. Minimizing error and bias is essential for any study to hold reliability.

Reporting level of certainty in combination with minor influence from researchers (compared to qualitative research where researcher interpretation plays an essential role) may be what has given qualitative research a favorable reputation.

6.3.2 Reliability of raw data

A ground assumption vital to making correct judgements from statistical inference is reliability of the data tested (Given, 2008; Walpole, 2012). Since the statistical process in itself is a mathematical calculation it cannot assist and it is for the researcher to assess if the raw data is reliable. Such an assessment is in the authors experience not straightforward

nor easy, though it may have appeared so in the beginning. Ensuring reliability of raw data can be done by correct design and development of instrument and procedure for data collection, as well as careful attention to the collection procedure. Designing and developing a means or an instrument to collecting quantitative data poses challenges in itself, and the use of instruments developed by other does not ease the challenge if the researcher aim at truly understanding the data and the data structure on a foundational level. For the author this notion is self-evident, - one should understand the way one collects, organizes and treat the data as an ethic essential to research and to actually be able to make correct assessments of the data and from there hope to draw a useful conclusion.

One should ask oneself if the data collection method is able to capture the desired data while considering what data is desirable to collect.

6.3.3 Explorative data analysis contra confirmatory data analysis

Having access to a large data set with multiple variables, which was the case with the Lean Business database, renders possible a multitude of statistical analysis as well as other approaches. This is a great opportunity to engage in an explorative data analysis (Tukey, 1977), which essentially concerns itself with how one can and should look at data to see what it says. With such an exploratory nature of an analysis comes opportunity and ambiguity, first of all since there is a multitude of analytical and numerical methods possible. Secondly, there are many, many variables to investigate each with a separate set of data points. If starting at square one, a researcher does not know what methods are interesting to apply, nor to which variables they should be applied to. One may begin such an analysis with a research hunch, however this is the only guiding methodology. Tukey (1977) places emphasis on using data to develop interesting hypotheses and not solely focus on confirmatory data analysis. Exploratory data analysis is more of a methodology than a method, focusing on understanding and challenging assumptions in developing hypotheses regarding the cause of an observed phenomena. This mindset can be helpful in big data analysis, since challenging initial assumptions and data collected can aid in systematic exclusion of errors, necessary to ensure validity in big data analysis (Kaplan et al., 2014).

Ending an exploratory data analysis can pose a great challenge on a researcher, since there is no clear ending or final answer. Closing this exploratory phase some time before one must (due to for example time) may result in a hypothesis with a statistical test providing a greater contribution to the research field than the hypothesis that was tested last. As such,

the researcher must be jointly capable of handling the opportunities and challenges inherent to the exploratory data analysis and balance it with confirmatory data analysis.

6.3.4 A note on software tools

As a rule of thumb in data analysis in general is that 80% of the time is spent on organizing data, rather than conducting any actual analysis. The author hypothesiese that this is the case when conducting an exploratory data analysis as well. In this regard, the selection and use of software is essential. Incorrect software can limit the researcher's capacity and study development significantly, while in the same way correct software that incorporate particular features tailored to the data and study requirements can lift productivity to a much higher level.

7 The experimental approach (paper 4)

7.1 Introduction to the experimental approach

This section introduces the fourth and final paper, which describes how one can conduct user testing by an experimental approach. It aims at placing the paper context in the overall business model phenomena context and provide details on the experimental methodology.

7.1.1 A deeper level of the business model phenomena – the customer offerings

A business must necessary provide an offering to the customer, whether that be a stand-alone physical product, a service, an experience, an application etc, or some integrated bundle of the mentioned offerings. Offerings is one part of the business model at a deeper level, as it details what one delivers to the customer. Offerings truly fulfilling user needs and requirements can create a seemingly unique or superior customer value, gaining a competitive advantage for the company's whose offerings it is. Human centered design (Sanders and McCormick, 1987; Woodson and Conover, 1970) has a focus on thorough understanding of user needs and integrates this with development of offerings, arguing that the resulting product will have an increased chance of success due to truly providing the customer with value. To ensure that offerings are developed according to user needs and requirements, one can conduct user testing of said offering. This article describe how one can conduct user tests of a product offering using an experimental approach. This could aid in understanding the business model phenomena from a product perspective which can be seen as a deeper and more detailed level of the business model, namely through the offering provided by the company to the customer. Learning about the experienced customer value from an offering, how one can improve this value by continuing to develop the offering is important for a business model and can be aided by conducting a user (or customer) experiment.

7.1.2 The experimental or quasi-experimental approach

A research question can be investigated by testing of a hypothesis, which is essentially what the experimental approach does. Experimental research is comparative as it investigates the relationship between input and output variables. By inducing changes in independent variables (one or more input variables), changes in dependent variables (one or more output variables) may occur. As an experiment aims at determining causality

(chronological order of concepts, covariance and exclude other factors) such a relationship must be established statistically to be valid.

In a classical experiment there are a number of requirements. The context or environment in which the phenomena is to be investigated must be under control by the researcher, participants must represent the target population and be randomly assigned to experimental groups, and the experiment must be repeatable. As it is difficult to fulfill all these requirements, one may conduct a quasi-experiment, where one compromise on some control while retaining the underlying logic (Blessing and Chakrabarti, 2009).

Experiments play a key role in design research, which concerns itself with improving the given situation. Design research investigates questions such as what successful products entails, how successful or unsuccessful products are created, and how one can improve the chances of being successful (Blessing and Chakrabarti, 2009). The hypotheses tested can concern if an entity meet the goals said entity. This entity could encompass a company's offerings, which could be a physical product, a service, an experience or application depending on the company. The entity can also be a business model or a subset of it. An experiment can be used to investigate entities already on the market and entities under development. One could for example investigate how to introduce a newly developed business model or if a prototype of product offerings is providing customer value.

7.1.3 Introducing the paper

The forth paper is "Dybvik, H., Wulvik, A., & Steinert, M. (2018). STEERING A SHIP-INVESTIGATING AFFECTIVE STATE AND WORKLOAD IN SHIP SIMULATIONS. In DS92: Proceedings of the DESIGN 2018 15th International Design Conference (pp. 2003-2014).". It was presented at the 15th International Design Conference, "DESIGN 2018" hosted by the Design Society in May 2018.

This paper encompasses a description of procedure, test environment and an explicitly states hypotheses necessary for conducting a user experiment. It describes how an experiment can be set up, conducted and used to investigate how to design and develop a successful product. The context is maritime, where the working conditions for ship captains are tested to provide foundation for subsequent design and development of ship bridges. Here, the product could be the interface design on the ship bridge, where product success is measured by the ship captains task performance and user experience.

7.2 STEERING A SHIP - INVESTIGATING AFFECTIVE STATE AND WORKLOAD IN SHIP SIMULATIONS

The paper is enclosed in the following pages.

INTERNATIONAL DESIGN CONFERENCE - DESIGN 2018
https://doi.org/10.21278/idc.2018.0459

STEERING A SHIP - INVESTIGATING AFFECTIVE STATE AND WORKLOAD IN SHIP SIMULATIONS

H. Dybvik, A. Wulvik and M. Steinert

Abstract

We present an experiment investigating concepts of affective state and workload in a large ship manoeuvring context. It is run on a consumer ship simulator software where student participants (N=31) perform two ecologically valid scenarios: sailing on open sea and in a harbour. Results from surveys show highly significant changes in terms of both affect and workload between the scenarios. Thus, one should consider varying affects and workloads from users in varying contexts, consequently demanding new design paradigms for product development, such as dynamically adaptive interfaces.

Keywords: human behaviour, emotional engineering, engineering design, empirical studies, ocean space

1. Introduction: The human element and the ship bridge

The ship bridge is where the captain and his crew controls the ship. Navigation, monitoring systems, and communicating with both internal and external personnel are important activities. Sea piloting, i.e. sailing on open sea normally consists of monitoring tasks and no active navigation at all. Harbour piloting, i.e. sailing in harbours, requires continuous adjustment of speed and course, monitoring ship systems, and communicating with both crew and external contacts. These scenarios range from the monotone to the highly complex (Norros, 2004; Nilsson et al., 2009).

Maritime accidents occur in either scenario (Nilsson et al., 2009), mostly as the result of human error. Research shows that 49 to 96 percent of all shipping incidents or marine causalities are caused by human error (Rothblum, 2000; Hetherington et al., 2006; Tzannatos, 2010). Given the large share of maritime accidents caused by human error, this paper aims to direct attention towards the human users and their mental state during ship operation with the goal of identifying opportunities for reducing accidents. The notion of human centred design (Woodson and Conover, 1970; Sanders and McCormick, 1987) has existed since the 1960s. When considering humans in engineering, they are usually represented by generic models based on certain boundary conditions (Balters and Steinert, 2017). Models often represent the "average" human, with a general and stable behaviour response. Kahneman and Tversky (1979, 1984) show that this is indeed not the case. They show that humans are not rational with stable behavioural responses to stimuli. Human behaviour is influenced by psychological, physiological and situational factors. This could be issues in personal life, lack of sleep, or suddenly demanding tasks that needs to be solved. Following the fact that humans are not static entities with known responses, but rather change over time and contexts, efforts should be made to gain insights about what might influence behaviour. Two potentially influential topics are the constructs of affect and workload. Knowledge about how affective state and workload influence operator performance could potentially aid engineers in their work to design and test new product solutions for the maritime industry. We believe that by

taking these parameters into consideration, human error could be reduced by designing the system around the human, and not make the human adapt to the system.
The paper proposes and demonstrates an experimental setup to investigate differences in affective state and workload between two ecologically valid scenarios within the domain of large ship navigation. The goal of this paper is to show that there are measurable differences in affective state and workload between the two scenarios. This may influence new ship bridge designs. These tasks have been developed in cooperation with ship simulator instructors with extensive experience as ship navigators. The experiment is run in a consumer ship simulator software (N=31) where participants from a student population are tasked to steer a ship in the following scenarios: cruising on open sea and navigating a busy harbour. Data was collected through video, self-assessment surveys and physiology sensors. The paper is part of a larger study investigating the relationship between physiological data, affect, and workload. The foundation, description and analysis of the physiological data is not within the scope or aim of this paper, and will be discussed elsewhere.
The results from the self-assessments show highly significant differences in terms of both affect and workload for the two scenarios. Consequently, one will have to consider distinctly varying affects and workloads from users in varying contexts, which, if translated into GUI and UI design suggest new design paradigms such as dynamically adaptive interfaces.

2. Theoretical foundation

2.1. Affect

Psychology presents emotion or affect as a set of variables that may moderate behaviour (Balters and Steinert, 2017). There are two main schools of thought when describing affect. The first describe emotions as a set of discrete categories (Tomkins, 1962; Ekman and Friesen, 1971; Ekman, 1992). The second describe emotions as a combination of multiple dimensions (Thayer, 1967; Russell, 1980; Watson and Tellegen, 1985; Russell and Barrett, 1999). In this paper, we adopt the description of emotions or affect of Russell (1980), the Circumplex Model of Affect, later named the Affect Grid (Russell et al., 1989). Affect is described as a construct made up of the combination of the two dimensions, arousal-sleepiness and pleasure-displeasure, see Figure 1.
Several researchers have considered how stress might influence human performance (Westman and Eden, 1996; Healey and Picard, 2005; Balters and Steinert, 2017). Russell et al. (1989) describe the construct of stress as the combination of arousal and displeasure. This is also referred to as distress as opposed to eustress which is the combination of arousal and pleasure (Healey and Picard, 2005; Balters and Steinert, 2014, 2017). Baddeley (1972) shows that increased arousal seems to narrow attention, which in term increases performance on the task that is deemed most important, but decrease performance on all other tasks.

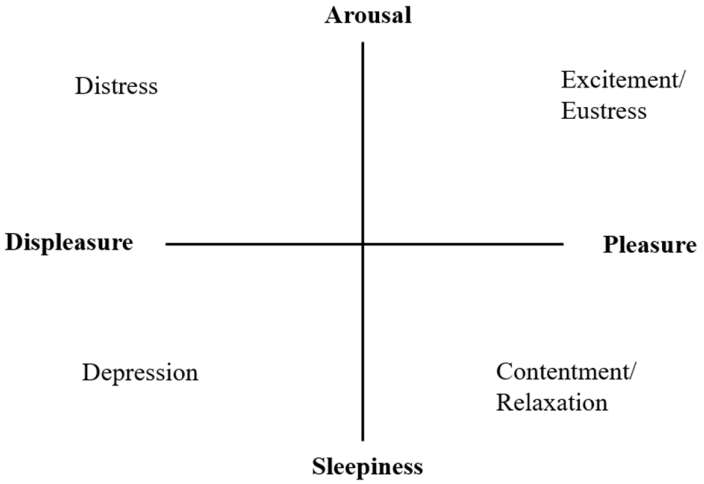


Figure 1. The Affect grid (adapted from Russell, 1980; and Russell et al., 1989)

2.1.1. Subjective measurements of affect

Assessing the subjective experience of affect is commonly done through self-report surveys. Affect can be evaluated through survey questions asking participants to evaluate levels of pleasantness and arousal (Russell, 1980), or through the single-item Affect grid (Russell et al., 1989). Positive and negative affect can be evaluated through the Positive and Negative Affect Schedule (PANAS) scales (Watson et al., 1988; Thompson, 2007). The Activation-Deactivation Adjective Check List (AD ACL) measures levels of activation (Thayer, 1967, 1986). Surveys provide a simple and low cost manner of gathering data of affective states. When using surveys in an experiment, they either interrupt participants, or must be used after tasks are finished. This might influence results, either because of the effect of an interruption, or that participants must recall how they felt during a task. Due to the subjective nature of surveys, there might be issues of self-filtering and different interpretations of questions.

2.1.2. Behavioural measurements of affect

Behavioural measurements of affect are typically concerned with measuring components of facial expression (Ekman and Friesen, 1978; Gottman and Krokoff, 1989), pitch of voice (Russell et al., 2003), and body posture and gesturing (Coulson, 2004; Wulvik et al., 2016). Advantages with behavioural measurements of affects is a very fine grained analysis of behaviour by trained experts, partially avoiding self-filtering of results, such as might be the case when answering surveys. Drawbacks are that these analyses are very labour- and time intensive, and that there might be issues of inter-coder reliability.

2.1.3. Physiological measurements of affect

The autonomic nervous system (ANS) is in charge of modulating peripheral functions of the body (Öhman et al., 2000; Mauss and Robinson, 2009). The ANS consists of the sympathetic and parasympathetic system. The sympathetic system is dominant during periods of activation, or "fight or flight", while the parasympathetic system is dominant during resting periods of the body. Changes in affective state are linked to physiological responses through the ANS. These responses can be through heart rate, heart rate variability, breathing rate, pupil dilation, muscle tension, galvanic skin response, body temperature, blood pressure, and brain activity to mention some. Healey and Picard (2005) showed a relation between levels of stress and metrics derived from galvanic skin response and heart rate variability. Baltaci and Gokcay (2016) differentiates affective states from relaxation to stress through pupil dilation and facial temperature. For a more comprehensive overview we refer to Balters and Steinert (2017), Mauss and Robinson (2009) and Levenson (2014). Physiology sensors have the advantage of providing continuous data without interrupting the person being measured, as opposed to subjective measurements through surveys. One limitation is that human physiology is very complex, and it is difficult to control all influencing factors. Another challenge with physiology data is interpreting results. How does e.g. a change in measured voltage between two sensors placed on the chest translate into affect? We recommend reading Balters and Steinert (2017) for a more complete overview.

2.2. Workload

Workload or cognitive load refers to the mental effort imposed on working memory by a particular task. (Sweller, 1988; Paas and Van Merriënboer, 1994; Paas et al., 2003) Cognition is related to our perception, in that perceptual activity, such as thinking, deciding, calculation, remembering, looking, searching increases the perceptual load, thereby the workload (Hart and Staveland, 1988). As the working memory is limited, it can be overloaded by increasing the requirements for perceptual activity. Wierwille and Eggemeier (1993) provide an overview of methods to measure workload. These can be divided into Subjective, performance-based and physiological.

2.2.1. Subjective measurements of workload

The NASA Task Load Index (NASA-TLX), a multi-dimensional scale designed for obtaining workload estimates (Hart and Staveland, 1988; Hart, 2006). NASA-TLX consists of rating six sub-scales, mental demand, physical demand, temporal demand, performance, effort, and frustration, from

low to high. Participants filling out the survey are also asked to pairwise compare the six dimensions in terms of how important they are for the performed task. An estimate of total workload is then calculated from the weighted average. Another multi-dimensional scale of subjective workload is the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988). It uses three levels (low, medium, high) along three dimensions, time load, mental effort load, and psychological stress load, to assess workload.
Overall Workload (Vidulich and Tsang, 1987) is a single scale measurement of subjective workload, ranging from very low to very high. Vidulich and Tsang (1987) show that the single-dimension scale of Overall Workload has higher sensitivity than the multi-dimensional scale of NASA TLX. Hill et al. (1992) showed that both the single-dimension scale of Overall Workload and NASA TLX was superior to SWAT in terms of sensitivity.

2.2.2. Performance based measurements of workload

Performance is expected to decrease with increases in workload through reduction in speed and accuracy (Wierwille and Eggemeier, 1993). Two strategies of evaluating workload through performance are common, primary and secondary task performance. Primary task performance, e.g. steering a ship might be insensitive to variations in workload, due to the operator recruiting extra resources to maintain performance (Hart and Wickens, 1990). Secondary task performance can both be assessed through external tasks and embedded tasks. External tasks are not part of the system being tested, e.g. calculating arbitrary arithmetic, while embedded tasks have a logical connection to the primary task, e.g. communicating via radio on a ship.

2.2.3. Physiological measurements of workload

Changes in physiological states has been shown to correspond with changes in workload (Galy et al., 2012). Common physiological measurements to evaluate workload are heart rate variability (HRV) (McDuff et al., 2014), electroencephalography (EEG) (Wilson and Russell, 2003), pupillary response (Iqbal and Bailey, 2005), and galvanic skin response (GSR) (Nourbakhsh et al., 2012).

3. Ship navigation experiment

An experiment was created to investigate concepts of affect and workload in two different ecologically valid scenarios in the context of large ship navigation. One task concerned steering a large ship on open water. The other task concerned steering a large ship through a busy harbour. These tasks can be described as low and high activity respectively. The aim of the experiment was to identify potential differences in affective state and workload in the different scenarios. The implication of different affective states and levels of workload for the various scenarios is that users could have changing capabilities, and that this should be addressed through the design of systems in the future. For the experiment, we formulate the following research question:
Is there a measurable difference in affective state and workload between low and high activity scenarios in the context of large ship navigation?

3.1. Scenarios

Two ecologically valid scenarios were created in the commercial ship simulator software Ship Simulator Extremes (*Ship Simulator Extremes*, 2010), replicating two typical situations in large ship navigation. Ecologically validity is obtained by the nature of the primary and the secondary task and the nature of the environmental stimuli i.e. sounds. Scenarios describe common activities on board large ships in daily operation. The scenarios and stimuli were developed in cooperation with several ship navigators with long experience as professional navigators.

3.1.1. Ship navigation on open sea – low level of activity

The first scenario was designed to recreate a low-activity situation where the task was to navigate on open sea. This is typically an uneventful task with long periods of time spent monitoring systems. The environment was set to *Dover*, and ship set to *Pride of Rotterdam,* a 215-meter long car ferry. The ship

was placed close to the exit of Dover harbour with the front of the ship pointing towards the English Channel. Participants were instructed to steer the ship straight ahead towards Calais, France. The task lasted for 15 minutes, but the duration was unknown to participants. The monotonous sound of a ship engine was added to create a realistic backdrop.

3.1.2. Ship navigation in a busy harbour – high level of activity

The second scenario was designed to simulate a high-activity situation where the task was to navigate a busy harbour under a time constraint with additional secondary in the form of radio communication. The environment was set to *Rotterdam*, and *Pride of Rotterdam* was again used at ship. The participants were instructed to steer through narrow channels to a designated berth for docking. Upon leaving the starting position, a ten-minute timer would start and be displayed in the top left corner of the screen, instructing participants to reach their destination within this time limit. At regular intervals throughout the ten minutes, participants were prompted to answer eight pre-recorded questions via radio from immigration, customs and the ship's main office. These questions were voiced by three different people unfamiliar to the participants. Answers to the questions could be found in two lists provided to the participants, a cargo manifest and a crew list. These were consciously designed to be hard to read, with small letters and lots of superfluous information. Questions were repeated after 90 seconds if no answer had been given, or upon request of the participants. If participants reached the designated berth, a new destination was given. The task was designed in such a way that the final destination would be next to impossible to reach in the available ten minutes.

3.2. Physical environment

The aim of the physical environment was a controlled, static, physical space for conducting the abovementioned ecologically valid scenarios in the context of large ship navigation. A honeycomb cardboard cubicle was built (similar to the one made by Leikanger et al. (2016), equipped with a 27" computer screen mimicking the window view. A keyboard had the numerical pad marked with stickers indicating what ship functionality they controlled, e.g. rudder, thruster, etc. Today, a ship bridge control interface consists of button arrays resembling a keyboard. Additionally, much of monitoring tasks are conducted using information conveyed on a computer screen. Headphones eliminated external noise, ensuring exposure to the sound introduced by the experimenters only, i.e. ship engine noise, radio chatter and the task-specific questions. Effects from changes in external light was controlled by obscuring ambient light and illuminating the cubicle artificially with an LED strip and normal ceiling lights. Additional equipment included a mouse for answering the surveys, two web cameras for recording and monitoring the participant, a Bluetooth antenna hidden close to the devices, lists with information regarding the questions in the second scenario and marking tape indicating the area for placing the left hand. Figure 2 shows the experiment environment.

Figure 2. Experiment environment, both physical and virtual. ECG (top) and GSR (bottom) sensors highlighted in red rectangle

3.3. Collecting data from participants

A combination of self-report surveys and physiology sensors were used for data collection in the experiment. In addition, video was recorded to allow in-depth analysis of collected data.

3.3.1. Self-report surveys

To evaluate subjectively experienced affect, participants were asked to evaluate their state of arousal, awakeness, alertness, pleasantness, and stress on scales from 0 to 10. Arousal and pleasantness was taken directly from the Circumplex Model of Affect. Awakeness and alertness were added after pilot studies uncovered that participants had trouble understanding the meaning of arousal to triangulate their meaning. A question of stress was included in the surveys to capture the participants' notion of stress directly, and not only as a combination of arousal and pleasantness.
For self-assessment of workload, the single-dimension Overall Workload scale (Vidulich and Tsang, 1987) and NASA TLX (Hart and Staveland, 1988) was used. Overall Workload was evaluated on a scale from 0 to 10, and the six dimensions of the NASA TLX survey was evaluated on scales from 1 to 7, as well as 15 pairwise comparisons. All survey answers were collected through Google Forms.

3.3.2. Physiology sensors

Two types of physiology data were collected in this experiment, electrocardiography (ECG) and galvanic skin response (GSR). Electrocardiography measures electric potentials over the heart through sensors placed on the skin. The Shimmer3 ECG unit (Shimmersense, 2017a) was used in this experiment, with a sampling rate of 512 Hz. Five sensors were placed on the skin of participants per the instructions provided by Shimmer, with the V_x lead placed on position six. Data collected through ECG is measured in millivolts [mV], and can be translated into variables such as heart rate and heart rate variability. Galvanic skin response (GSR) is a measurement of conductance over the skin. The Shimmer3 GSR+ Unit (Shimmersense, 2017b) was used to measure skin conductivity. Two sensors were connected to the underside of the medial phalanx on the index and middle finger of the left hand. Sampling rate was set to 128 Hz.

3.3.3. Organising stimuli and synchronizing data

iMotions 6.4 (*iMotions*, 2017), a software platform for biometric research was as framework for presenting stimuli and synchronizing data. The sequence of instructions, surveys and simulator tasks were pre-defined in iMotions. Physiology data and video were given a common timestamp from iMotions, syncing data for future analysis.

3.4. Stating the hypotheses

The analysis of results in this paper concerns the change of self-reported affective state and workload. We operationalise the research question stated above into testable hypotheses.
Is there a measurable difference in affective state and workload between high and low activity scenarios in the context of large ship navigation?

3.4.1. Affect hypotheses

Affect is measured by asking participants to evaluate their level of arousal, awakeness, alertness, pleasantness, and stress. This leads to the following five hypotheses:

- **Affect H1:** *There is a significant change in self-reported arousal between low and high activity tasks.*
- **Affect H2:** *There is a significant change in self-reported awakeness between low and high activity tasks.*
- **Affect H3:** *There is a significant change in self-reported alertness between low and high activity tasks.*
- **Affect H4:** *There is a significant change in self-reported pleasantness between low and high activity tasks.*

- **Affect H5:** *There is a significant change in self-reported stress between low and high activity tasks.*

3.4.2. Workload hypotheses

Workload has been evaluated by participants assessing their overall workload on a single scale and through filling out the NASA TLX survey. This leads to the two following hypotheses:

- **Workload H1:** *There is a significant change in overall workload between low and high activity tasks.*
- **Workload H2:** *There is a significant change in TLX workload between low and high activity tasks.*

3.5. Running the experiment

This section aims to display how the experiment was run, and give a detailed description of the data foundation.

3.5.1. Participants

Participants in this experiment came from an engineering background (N=31). Age ranged from 19 to 33 years (24.0 ± 2.74). Out of 31 participants, 18 were male and 13 female. In addition, there were eleven participants were excluded from the analysis due to technical errors and failure to follow instructions. In the invitation to the experiment, participants were asked to participate in a study concerning "Ship Manoeuvring Behaviour". They were asked to wear a loose top for convenient connection of physiology sensors.

3.5.2. Experimenters

Two researchers conducted the experiment. The first experimenter would greet, brief, and attach sensors to participants. All interactions were scripted in advance to ensure that every participant was exposed to the same stimuli. The experimenter read all instructions from a manuscript, wore similar clothing (black jeans, light coloured dress shirt, hair pulled back in pony-tail, and no make-up). The second experimenter would sit behind a wall controlling the stimuli. After the experiment finished, the first experimenter debriefed the participant and removed sensors.

3.5.3. Protocol

Participants were greeted, introduced to the experiment, and informed about what kind of data that would be recorded. A consent form was signed by the participant, agreeing to have video, physiology data (electrocardiography and galvanic skin response) and survey answers recorded. Physiology sensors were attached by the experimenter. Participants were then instructed to sit down in front of a computer screen, and place their left hand on the table, making sure their arm was resting comfortably. They were told that instructions may be given both on-screen and through audio. In the case of audio instructions, answers should be given through a radio handset. Usage of the radio handset was explained and demonstrated. Participants were instructed to keep their left hand still throughout the experiment to ensure the quality of GSR data recorded. After instructions were given, the experimenter left the room and joined the second experimenter behind a wall. The computer screen showed a black image with white crosshairs in the middle when participants entered the room. When participants were ready to start, the second experimenter manually started the sequence of stimuli in iMotions. Participants were first presented with neutral stimuli. Participants then filled out a survey on their affective state to serve as a reference baseline. Information about the experiment was given in writing with a white background. They were informed that they would be controlling the ship *Pride of Rotterdam* and execute various tasks. Participants were informed that there were two printed lists, a crew list and cargo manifest, to their right side. These lists should not be used before instructed to do so. Following the initial brief, participants were shown a video giving instructions for how to control the ship with the keyboard. All keys to be used on the keyboard were physically labelled with a short explanatory name. After receiving instructions, participants were informed that the first task

would begin and the computer screen switched to the simulator software for the low activity task. The second experimenter manually unpaused the software and gave over control to the participants. After 15 minutes from leaving the harbour in Dover the software would display a loading screen, initialising the high activity task. The second experimenter would manually change the view to the second survey, concerning affective state and workload. After completing the survey, the view was manually switched back to the simulator software, starting the high activity task. The ten-minute timer would start after the ship had started moving, and pre-recorded radio questions were manually played at pre-defined intervals by the second experimenter. After the ten minutes passed, a screen telling participants that they failed their mission (no participants were able to complete the mission, as expected). The third survey was presented to participants, asking about affective state and workload. When completed, participants were prompted to answer background questions, e.g. age, gender, occupation, in a fourth survey. After completing the final survey, they were informed that the experiment was finished, and were thanked for their participation. The first experimenter would walk back to debriefing the participants, thank them for their contribution, remove the sensors and ask them not to share content or details about the experiment to others. All equipment was cleaned and printed lists were replaced after each participant.

4. Survey results

Survey results from the 31 participants completing the experiment was analysed in SPSS Statistics (IBM, 2016) to investigate potential statistical differences in affective states and workload. A total of seven variables were tested for statistically significant change in values on an 11-point scale, between low and high activity tasks. Statistical tests were selected based on the properties of recorded data, i.e. outliers, normal-, and symmetric distributions. Difference in values between the two scenarios is the foundation for the tests. Paired samples t-test was used for normally distributed date without significant outliers. For data violating the assumptions of normal distribution or no significant outliers, the Wilcoxon signed-rank test was used if the data was symmetrically distributed. For non-symmetric distributions, the Sign test was used. The Wilcoxon signed-rank test and the Sign test evaluates median differences as opposed to mean differences in the paired samples t-test. Outliers are defined by SPSS statistics as values more than 1.5 box-lengths from the edge of a box in a box plot. Shapiro-Wilk's test for normal distribution was used to assess whether values were normally distributed, where significance values larger than 0.05 indicates normally distributer variables. Symmetricity of distribution was evaluated visually using histograms. Data are mean ± standard deviation, unless otherwise stated. Table 1 contains descriptive statistics, and Table 2 contains metrics associated with assumptions that decide which statistical tests to use along with the corresponding results. As shown in Table 2, all seven variables are significantly different in the two scenarios.

Table 1. Descriptive statistics

Variable	S1	S1 Median	S2	S2 Median	Diff.	Diff. Median
Arousal	5.61 ± 2.38	6	6.68 ± 2.70	8	1.06 ± 1.75	1
Awakeness	6.48 ± 2.05	7	7.61 ± 2.04	8	1.13 ± 1.09	1
Alertness	6.26 ± 1.95	7	7.42 ± 1.86	8	1.16 ± 1.61	1
Pleasantness	6.35 ± 1.62	7	4.68 ± 1.80	4	-1.68 ± 1.54	-2
Stress	3.23 ± 1.94	3	6.39 ± 2.14	7	3.16 ± 1.88	3
Overall Workload	2.03 ± 1.70	2	8.03 ± 1.78	8	6.00 ± 2.93	7
TLX Workload	2.86 ± 1.23	2.73	5.60 ± 0.74	5.80	2.73 ± 1.37	2.86

Table 2. Testing for statistical difference change in variables between low and high activity scenarios

Variable	Outliers	Shapiro-Wilk's test	Symmetric	95% CI Lower	95% CI Upper	Sig. (2-tailed)
Arousal	Yes	0.063	Yes	0.42	1.71	< 0.01 [a]
Awakeness	No	< 0.01	No	0.73	1.53	< 0.01[c]
Alertness	Yes	0.014	No	0.57	1.75	< 0.01[c]
Pleasantness	No	0.214	Yes	-2.24	-1.11	< 0.01 [a]
Stress	No	0.112	Yes	2.47	3.85	< 0.01 [a]
Overall Workload	Yes	< 0.01	No	4.93	7.07	< 0.01[c]
TLX Workload	No	0.48	Yes	2.23	3.24	< 0.01[a]

a: Paired samples t-test, b: Wilcoxon signed-rank test, c: Sign test

5. Discussion: Interpreting the results and the way forward

Results in the above tests show that there are significant changes in all seven variables. TLX Workload is a weighted sum of the six dimensions: Mental demand, physical demand, temporal demand, performance, effort, and frustration level. With the exception of performance (p=0.69), all dimensions had significantly changes. This might be due to difficulties related to comparing performance in two very different and unfamiliar scenarios. Changes are quite small for the variables of arousal, awakeness and alertness, with mean changes of around one on an eleven-point scale. Variables of pleasantness, stress, overall workload and TLX workload have a larger change (see Table 1). We are not sure whether the differences in magnitude of change is due to real differences, or due to how participants interpret the survey questions. One can speculate e.g. that participants did not have a clear understanding of the concepts of arousal, awakeness, and alertness, or at least had difficulties evaluating them. Pleasantness, stress and workload might be more intuitively understandable for the participants, which might be the reason for the difference in magnitude of change. This finding is interesting, as it contrasts with the fact that Russell (1980) defines stress as a combination of arousal and pleasantness. We know from literature that there is supposed to be a link between physiological data and arousal, e.g. heart rate variability and skin conductance. Further work will include analysing physiological data and comparing results with subjective assessment of affective state and workload, investigating the relationship between the two. One limitation of our study is that participants were sampled from a student population. Results might have been influenced by this fact, due to being unfamiliar with the situation of ship piloting. We believe that the findings that show a difference in affective state and workload between the two scenarios are valid for the context of ship navigation, although the effect size should be verified through testing with professional navigators in more realistic contexts, i.e. professional ship simulators or real ships.

The results nevertheless show that there is a clear difference in affective state and workload in the two scenarios tested in this experiment. Consequently, one should consider distinctly varying affects and workloads from users in varying contexts. This, if translated into product development, GUI, and UI design suggest new design paradigms such as dynamically adaptive interfaces.

References

Baddeley, A.D. (1972), "Selective attention and performance in dangerous environments", *British Journal of Psychology*, Vol. 63 No. 4, pp. 537–546. https://doi.org/10.1111/j.2044-8295.1972.tb01304.x

Baltaci, S. and Gokcay, D. (2016), "Stress Detection in Human–Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features", *International Journal of Human–Computer Interaction*, Vol. 32 No. 12, pp. 956–966. https://doi.org/10.1080/10447318.2016.1220069

Balters, S. and Steinert, M. (2014), "Decision-making in engineering-a call for affective engineering dimensions in applied engineering design and design sciences", *Proceedings of the 2014 International Conference On*

Innovative Design and Manufacturing (ICIDM 2014), August 13-15, 2014, Montreal, Canada, IEEE, pp. 11–15. https://doi.org/10.1109/IDAM.2014.6912663

Balters, S. and Steinert, M. (2017), "Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices", *Journal of Intelligent Manufacturing*, Vol. 28 No. 7, pp. 1585 - 1607. https://doi.org/10.1007/s10845-015-1145-2

Coulson, M. (2004), "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence", *Journal of Nonverbal Behavior*, Vol. 28 No. 2, pp. 117–139. https://doi.org/10.1023/B:JONB.0000023655.25550.be

Ekman, P. (1992), "An argument for basic emotions", *Cognition and Emotion*, Vol. 6 No. 3-4, pp. 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. and Friesen, W.V. (1971), "Constants across cultures in the face and emotion", *Journal of Personality and Social Psychology*, Vol. 17 No. 2, pp. 124-129. https://doi.org/10.1037/h0030377

Ekman, P. and Friesen, W.V. (1978), *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto, California.

Galy, E., Cariou, M. and Mélan, C. (2012), "What is the relationship between mental workload factors and cognitive load types?", *International Journal of Psychophysiology*, Vol. 83 No. 3, pp. 269–275. https://doi.org/10.1016/j.ijpsycho.2011.09.023

Gottman, J.M. and Krokoff, L.J. (1989), "Marital interaction and satisfaction: a longitudinal view", *Journal of Consulting and Clinical Psychology*, Vol. 57 No. 1, pp. 47-52.

Hart, S.G. (2006), "Nasa-Task Load Index (NASA-TLX); 20 Years Later", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* Vol. 50 No. 8, pp. 904–908. https://doi.org/10.1177/154193120605000909

Hart, S.G. and Staveland, L.E. (1988), "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research", *Advances in Psychology*, Vol. 52, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hart, S.G. and Wickens, C.D. (1990), "Workload Assessment and Prediction", In: Booher, H.R. (Ed.), *Manprint,* Springer, Dordrecht, pp. 257–296. https://doi.org/10.1007/978-94-009-0437-8_9

Healey, J.A. and Picard, R.W. (2005), "Detecting stress during real-world driving tasks using physiological sensors", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6 No. 2, pp. 156–166. https://doi.org/10.1109/TITS.2005.848368

Hetherington, C., Flin, R. and Mearns, K. (2006), "Safety in shipping: The human element", *Journal of Safety Research*, Vol. 37 No. 4, pp. 401–411. https://doi.org/10.1016/j.jsr.2006.04.007

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklade, A.L. and Christ, R.E. (1992), "Comparison of Four Subjective Workload Rating Scales", *Human Factors*, Vol. 34 No. 4, pp. 429–439. https://doi.org/10.1177/001872089203400405

IBM (2016), *IBM SPSS Statistics for Mac.* [online] IBM, New York, USA. Available at: https://www.ibm.com/products/spss-statistics

iMotions (2017), *iMotions biometric research platform*. [online] iMotions. Available at: https://imotions.com

Iqbal, S.T. and Bailey, B.P. (2005), "Investigating the Effectiveness of Mental Workload As a Predictor of Opportune Moments for Interruption", *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05), Portland, USA, April 2-7, 2005*, ACM, New York, USA, pp. 1489–1492. https://doi.org/10.1145/1056808.1056948

Kahneman, D. and Tversky, A. (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica,* Vol. 47 No. 2, pp. 263–292.

Kahneman, D. and Tversky, A. (1984), "Choices, values, and frames", *American Psychologist*, Vol. 39 No. 4, pp. 341-350. https://doi.org/10.1037/0003-066X.39.4.341

Leikanger, K.K., Balters, S. and Steinert, M. (2016), "Introducing the Wayfaring Approach for the Development of Human Experiments in Interaction Design and Engineering Design Science", *Proceedings of the DESIGN 2016 / 14th International Design Conference, Dubrovnik, Croatia, May 16-19, 2016*, The Design Society, Glasgow, pp. 1751–1762.

Levenson, R.W. (2014), "The Autonomic Nervous System and Emotion", *Emotional Review*, Vol. 6 No. 2, pp. 100–112. https://doi.org/10.1177/1754073913512003

Mauss, I.B. and Robinson, M.D. (2009), "Measures of emotion: A review", *Cognition and Emotion,* Vol. 23 No. 2, pp. 209–237. https://doi.org/10.1080/02699930802204677

McDuff, D., Gontarek, S. and Picard, R. (2014), "Remote measurement of cognitive stress via heart rate variability", *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2014), Chicago, Illinois, August 26-30, 2014*, IEEE, pp. 2957–2960. https://doi.org/10.1109/EMBC.2014.6944243

Nilsson, R., Gärling, T. and Lützhöft, M. (2009), "An experimental simulation study of advanced decision support system for ship navigation", *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 12 No. 3, pp. 188–197. https://doi.org/10.1016/j.trf.2008.12.005

Norros, L. (2004), Acting under uncertainty: The core-task analysis in ecological study of work, VTT Technical Research Centre of Finland, Espoo, Finland.

Nourbakhsh, N., Wang, Y., Chen, F. and Calvo, R.A. (2012), "Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks", *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12), Melbourne, Australia, November 26-30, 2012,* ACM, New York, USA, pp. 420–423. https://doi.org/10.1145/2414536.2414602

Öhman, A., Hamm, A. and Hugdahl, K. (2000), "Cognition and the autonomic nervous system: orienting, anticipation, and conditioning", In: Cacioppo, J.T., Tassinary, L.G. and Berntson, G.G. (Eds.), *Handbook of psychophysiology*, 2nd ed., Cambridge University Press, New York, pp. 533–575.

Paas, F., Tuovinen, J.E., Tabbers, H. and Gerven, P.W.M.V. (2003), "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory", *Educational Psychologist*, Vol. 38 No. 1, pp. 63–71. https://doi.org/10.1207/S15326985EP3801_8

Paas, F.G. and Van Merriënboer, J.G. (1994), "Instructional control of cognitive load in the training of complex cognitive tasks", *Educational Psychology Review,* Vol. 6 No. 4, pp. 351–371. https://doi.org/10.1007/BF02213420

Reid, G.B. and Nygren, T.E. (1988), "The subjective workload assessment technique: A scaling procedure for measuring mental workload", *Advances in Psychology*, Vol. 52, pp. 185–218. https://doi.org/10.1016/S0166-4115(08)62387-0

Rothblum, A.M. (2000), "Human error and marine safety", *National Safety Council Congress and Expo, Orlando, Florida, October 16-18, 2000*.

Russell, J.A. (1980), "A circumplex model of affect", *Journal of Personality and Social Psychology*, Vol. 39 No. 6, pp. 1161–1178. https://doi.org/10.1037/h0077714

Russell, J.A. and Barrett, L.F. (1999), "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant", *Journal of Personality and Social Psychology*, Vol. 76 No. 5, pp. 805-819. https://doi.org/10.1037/0022-3514.76.5.805

Russell, J.A., Bachorowski, J.-A. and Fernández-Dols, J.-M. (2003), "Facial and Vocal Expressions of Emotion", *Annual Review of Psychology*, Vol. 54, pp. 329–349. https://doi.org/10.1146/annurev.psych.54.101601.145102

Russell, J.A., Weiss, A. and Mendelsohn, G.A. (1989), "Affect grid: A single-item scale of pleasure and arousal", *Journal of Personality and Social Psychology,* Vol. 57, pp. 493–502. https://doi.org/10.1037/0022-3514.57.3.493

Sanders, M.S. and McCormick, E.J. (1987), *Human factors in engineering and design*, McGraw-Hill.

Shimmersense (2017a), Shimmer3 ECG/EMG Unit, Available at: http://www.shimmersensing.com/products/shimmer3-ecg-sensor

Shimmersense (2017b), Shimmer3 GSR+ Unit, Available at: http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

Ship Simulator Extremes (2010), *Ship Simulator Extremes.* [online] ShipSim.com. Available at: https://www.shipsim.com/products/shipsimulatorextremes

Sweller, J. (1988), "Cognitive load during problem solving: Effects on learning", *Cognitive Science*, Vol. 12 No. 2, pp. 257–285. https://doi.org/10.1207/s15516709cog1202_4

Thayer, R.E. (1967), "Measurement of Activation through Self-Report", *Psychological Reports*, Vol. 20 No. 2, pp. 663–678. https://doi.org/10.2466/pr0.1967.20.2.663

Thayer, R.E. (1986), "Activation-Deactivation Adjective Check List: Current Overview and Structural Analysis", *Psychological Reports,* Vol. 58 No. 2, pp. 607–614. https://doi.org/10.2466/pr0.1986.58.2.607

Thompson, E.R. (2007), "Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS)", *Journal of Cross-Cultural Psychology*, Vol. 38 No. 2, pp. 227–242. https://doi.org/10.1177/0022022106297301

Tomkins, S. (1962), *Affect imagery consciousness: Volume 1: The positive affects,* Springer Publishing Company.

Tzannatos, E. (2010), "Human Element and Accidents in Greek Shipping", *The Journal of Navigation*, Vol. 63, pp. 119–127. https://doi.org/10.1017/S0373463309990312

Vidulich, M.A. and Tsang, P.S. (1987), "Absolute Magnitude Estimation and Relative Judgement Approaches to Subjective Workload Assessment", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 31 No. 9, pp. 1057–1061. https://doi.org/10.1177/154193128703100930

Watson, D. and Tellegen, A. (1985), "Toward a consensual structure of mood", *Psychological Bulletin*, Vol. 98 No. 2, pp. 219-235. https://doi.org/10.1037/0033-2909.98.2.219

Watson, D., Clark, L.A. and Tellegen, A. (1988), "Development and validation of brief measures of positive and negative affect: The PANAS scales", *Journal of Personality and Social Psychology*, Vol. 54 No. 6, pp. 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Westman, M. and Eden, D. (1996), "The inverted-U relationship between stress and performance: A field study", *Work Stress*, Vol. 10 No. 2, pp. 165–173. https://doi.org/10.1080/02678379608256795

Wierwille, W.W. and Eggemeier, F.T. (1993), "Recommendations for Mental Workload Measurement in a Test and Evaluation Environment", *Human Factors*, Vol. 35 No. 2, pp. 263–281. https://doi.org/10.1177/001872089303500205

Wilson, G.F. and Russell, C.A. (2003), "Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks", *Human Factors*, Vol. 45 No. 4, pp. 635–644. https://doi.org/10.1518/hfes.45.4.635.27088

Woodson, W.E. and Conover, D.W. (1970), *Human engineering guide for equipment designers*, University of California Press, California.

Wulvik, A., Erichsen, J. and Steinert, M. (2016), "Capturing Body Language in Engineering Design – Tools and Technologies", *Proceedings of the NordDesign 2016, Trondheim, Norway, August 10-12, 2016*, The Design Society, Bristol, pp. 165-174.

Andreas Simskar Wulvik, PhD Student
Norwegian University of Science and Technology - NTNU, Department of Mechanical and Industrial Engineering
Richard Birkelands Veg 2B, 7034 Trondheim, Norway
Email: andreas.wulvik@ntnu.no

7.3 Discussing the experimental approach

This discussion concerns the methodological approach used throughout the paper "STEERING A SHIP-INVESTIGATING AFFECTIVE STATE AND WORKLOAD IN SHIP SIMULATIONS", namely the experimental approach. It discusses the researcher's learnings and experiences from conducting the method, with known advantages and limitations. Some key outtakes are emphasized.

7.3.1 User experiments in general

The experimental approach in general hold the possibility of investigating and testing the relationship between multiple variables. It is therefore a good way to test theory or generate new theory. Provided that the variables and their dependencies have been thoroughly considered, they can be structured into well-formulated hypotheses before tested experimentally, where potential causal relationships can be established. Such a study usually receives high credibility provided it is well reported (Given, 2008; Johnson and Onwuegbuzie, 2004).

Of course, the experimental environment and stimuli intentionally effects participants, however this can also affect measurements. Their influence should be eliminated before starting the experiment or adjusted for in the subsequent data analysis (Balters and Steinert, 2015). Here, environmental stimuli were consistent for all participants and particular attention was devoted to consistent lighting and temperature of the setting.

Since an experiment tests one or more hypotheses, is must be reproducible by other researchers (Blessing and Chakrabarti, 2009). A thorough description of procedure, test environment and an explicitly hypothesis can aid here, including clear reporting of results. This should include which statistical tests were used. Advantages of doing so is generating results with high credibility and reliability. Further advantages include, that results from an experiment are relatively independent of the researcher (Johnson and Onwuegbuzie, 2004).

7.3.2 Accommodating and understanding the human user

One challenge in conducting user experiments; accommodating the human user, can also be found in qualitative research - for example an interview participant. This is challenging in multiple ways; humans change their mind, they are not always aware of their own believes or even able to articulate what they know, and they do not always make logical decisions (Kahneman and Tversky, 1984, 1979). This can be particularly challenging if the

overall aim is to create value for the user, since true understanding of user needs and requirements is desirable and even necessary for creating exceptional value for the user.

7.3.3 Surveys as a subjective means of measurement

Surveys as a mean to generate self-assessment of cognitive workload are instant measures with low variance between participants (Balters, 2017). When employed in an experiment one can observe a coherence between self-assessment and their behavioral responses, which could provide useful details in subsequent interpretation of the participants behavior. The subjective nature of survey questions could result in inconsistent answers, as the individual interpretation of the questions will be different from participant to participant. Differences in question interpretation could be caused by word interpretation (Hart and Staveland, 1988), the participants introspection and memory abilities, language fluency, etc. Including language fluency in the demographic questions could have given an impression of if this factor influenced the answer in any way. Here, this was not included which it should have been. Other potential influencing factors could and should also have been included to further eliminate or at least create an impression of the influence of rival factors.

Using a highly adopted survey, such as NASA-TLX (Hart and Staveland, 1988) claiming to be a convenient tool, bring both advantages and challenges. NASA-TLX was relatively short in length and quick to complete. Furthermore, high adoption is here linked to high acceptance and ecological validity. However, by observing the confused expression on some of our participants one could tell they had a hard time understanding and answering the questions. This impression is strengthened by discussion among HMI community who confirmed our impression, it is a complex tool and it is not user friendly.

7.3.4 Sensor as an objective means of measurement

Sensors are objective measurement tools, provided the tool is untampered. Advantages of objective measurement tools are objectiveness and allowing for sampling of continuous streams of data enabling monitoring over time. The vast majority of such tools are non-invasive, allowing for applications in situ and they are increasingly more affordable and accessible (Balters and Steinert, 2017; Sibi et al., 2017).

To ensure correct measurement, sensors must be calibrated and used correctly regardless of the nature of the sensor. Sensors used in the experiment approach targeted physiological

measurements, meaning that care must be taken when attaching the sensors to ensure correct and consistent placement, as data quality can be questioned otherwise. For example; the two electrodes of the GSR sensor requires skin contact and must therefore be secured tight enough to maintain contact throughout the experiment, however not so tight impacting blood flow to fingers, as this changes the physiological state of the fingers, thus the data. Caution was exerted when securing the sensor while retrieving feedback from the participant. GSR measurements can also be influenced by the use of cosmetic products and/or cleaning products in that it could alter skin conductivity (Balters and Steinert, 2017). To provide the same baseline for all participants alcohol prep-pads were used to clean the electrodes and finger area to attach the electrode to. Although participants may have a different skin conductivity as a result of drying their skin with alcohol, this baseline shift is consistent for all participants across the two scenarios. The difference in skin conductance levels between the low-stress and the high-stress situation should not be changed by a shift of baseline. Therefore, there is no reason to be concerned regarding the quality of the skin response data in this experiment.

Although the selected sensors should be applicable in-situ, there is of course always a possibility of having technical issues with data capturing tools. Here, a simple Bluetooth connection issue led to several iterations on where to locate the Bluetooth-antenna to maintain a continuous data stream during the experiment.

The advantage of using objective over subjective means of measurement is minimizing variance from participations interpretations and researcher bias or influence.

7.3.5 Participant sample

Participants were laboriously sampled from a student population. Students may be acceptable for initial tests of products, though results may have been influenced by them not being familiar with ship piloting and not adhering to the behavioral codex of ship captains. For testing new product solutions it would be ideal to have access to the people the design is intended for, which may not always be possible due to their job as a ship captain, a flight navigators, etc. or other causes of unavailability. An incorrect or poorly sampled participant sample could influence behavioral measurement tools and introduce bias in the results. When using user experiment results as feedback in continued development of product solutions it is important to be aware of where and from whom the results are coming, to ensure correct product development direction.

7.3.6 The experimental setting is not objective

Participants have a tendency to form an interpretation of the experiments purpose based on the experimental artifacts and alter their behavior accordingly. Such demand characteristics (Rosenthal and Rosnow, 2009), should be reduced to counteract bias. Counteracting measures included labeling the experiment "Ship maneuvering behavior study" and a carefully worded and perhaps brief introduction to the system. However, contributing demand characteristics included wording of questions in the survey, such as frustration and, stress level, and physiological sensors. GSR in particular can be related to sweaty palms, which relates to stress or physical performance.

Depending on the experimental setup, there are many aspects that could impose an effect on participant behavior. In this experiment there was an unknown situation where the participant, wearing two types of physiological sensors, was placed inside a small carboard cubicle, and asked to conduct specific tasks while having video and audio recorded. This could create a feeling of being under surveillance, make participants more conscious of their actions and reactions, body movement and physiological responses. If could also have a motivational effect, as some perform better under supervision. Furthermore, participants could attempt at exhibiting what they see as desirable behavior due to being monitored. Due to the sensitivity of GSR, the participants hand had to be kept still to limit noise in data, introducing a bodily constraint. Some combination of the abovementioned aspects could produce artificial behavior, which could affect the result of the experiment.

A participant reacting to demand characteristics, displaying signs of artificial behavior, or even playing the role of "the good subject" can of course occur in any research inquiry. The potential for incorrect influence of participants is a limitation is inherent to the experimental approach[3].

7.3.7 Incorporating *Observation*

Observing the participant during the experiment was useful as it gave a better understanding of participant behavior. Direct observation (Given, 2008; Wulvik et al., 2016) is in engineering design commonly applied prior to the data analysis since it can

[3] As mentioned previously, this applies in general to observational research – one cannot observe anything without affecting it to some extent.

provide indications of interesting data analyses. Being able to utilize an additional, qualitative method gave valuable nuances to interpreting the overall results.

7.3.8 Designing and developing a multidisciplinary experiment

Developing a multidisciplinary experimental setup is an elegant form of design and development, which the author finds to be an interesting challenge. Building a physical setup that incorporates various data capturing equipment can easily bring compliance and interrelations problems, as there is a great amount of detail that need to be in place for a smooth experimental run. One example here were noise in ECG data readings, not seemingly coming from muscle activity as one would expect (Balters and Steinert, 2015). Rather, it was result of the headphone wires resting on the chest of the pilot participant interfering with the readings. The flexible and iterative development process, as discussed below, was valuable in uncovering and solving such unexpected issues.

Developing the experiment was a flexible process (Thomke and Reinertsen, 1998). This included a quick response to external and internal changes (such as location and as discovering better technical solutions), early delivery (Jalote et al., 2004; Larman and Basili, 2003) of the ship maneuvering task and a bias towards action (Steinert and Leifer, 2012). It pushed the experimenters to conduct as many pilot experiments as possible within the timeframe, to uncover mistakes, lack of equipment and to train the first experimenter in the procedure. Furthermore, combining and testing multidisciplinary components together, such as sensors, computer, software, physical space, etc., was done early to discover and solving interference problems (Gerstenberg et al., 2015; Leikanger et al., 2016). This was valuable as it aided in handling uncertainty and complexity, and it enabled a rapid and smooth transition from piloting to running the experiment.

7.3.9 Design considerations made in preparing for the experiment

Awareness of potential pitfalls when running the experiment is important and suitable design considerations must be made when developing and preparing for the experiment. An example of a potential problem is the accessibility of the prescribed positions for the ECG electrodes, which could be inhibited by participants selection of clothing. To avoid such issues, we asked the participants to wear clothes allowing for easy access to the mentioned area, which appeared to work as there were no unexpected problem in this part of the experimental procedure. An example of insufficient design considerations can be found in the instruction prompts. The wording of these instructions left opportunity for

interpretation, where they should have been simple and non-ambiguous. Some of the participants failed to use the lists provided, in which case the data had to be discarded in the analysis.

8 Discussing mixing the methods

This section presents a joint discussion of the methods described in Introduction 4.1 – 7.1 and Discussion 4.3 – 7.3, which is a literature investigation, an in-depth case study, statistical inquiry and an experiment. The authors' experience in conducting the methods are discussed and related to literature. It includes appropriate use and mixing of methods, requirements posed on mixed method researchers, a discussion of reliability, quality of method and research design.

8.1 Appropriate use of the methods

8.1.1 Comparing their advantages and limitations

Conducting an in-depth literature investigation on one particular topic provided the author with a broad and relatively full picture of existing knowledge in the field, an ability inherent to the method (Fink, 2005; Given, 2008; Hart, 2001). Multiple perspectives of a topic can be explored (Given, 2008). Conducting a literature review is seen as a safe move, as it receives credibility and a wide audience (Johnson and Onwuegbuzie, 2004). Reviewing the literature is an activity that should be conducted on an ongoing basis throughout a project to keep up-to-date with the latest research findings (Blessing and Chakrabarti, 2009), which also should be the case when it is a part of a mixed method approach. In this case, the literature investigation was conducted prior to the other methods and in hindsight it could have been labeled a prescriptive study (Blessing and Chakrabarti, 2009). In practice, continuous review of existing and new literature is a challenge for researchers as they prioritize other, perhaps more pressing research activities. The share amount of literature contribute to the daunting feeling one may experience as one is beginning to take on a literature investigation, in addition to uncertainty in deciding when it is appropriate to stop or close the investigation.

The qualitative in-depth-case study provides an excellent opportunity to explore a broad research question, or a topic. A range of perspectives can be investigated and the method allow for an attention to detail, which could aid in understanding large complex phenomena. This is the case also for a literature investigation since multiple and detailed perspectives are available. Data collection is time consuming in qualitative research (Johnson and Onwuegbuzie, 2004; Yin, 2017), due to many data sources and data collection methods, and a lack of guidelines for when to freeze data collection. A challenge or

limitation to qualitative research is bias, which should be limited to ensure validity, while it simultaneously continues to be essential for the qualitative inquire excelling its performance in interpretive research questions.

The ability to conduct statistical tests on quantitative data is one of the most valuable advantages of quantitative research as seen by the author. By incorporating objective instruments for data collection we are able to objectively measure and assess a phenomenon while indicating the level of certainty, which minimizes bias from the researcher (Johnson and Onwuegbuzie, 2004). This is distinct from qualitative methods where researcher interpretation plays an essential role. Incorporating objective instruments, for example a sensor with continuous sampling frequency, can potentially bring in large amounts of data for exploratory data analysis of the measured phenomena. A challenge here, which is also the case for a literature investigation is the lack of clear guidelines for when to stop a study.

The experimental approach in general hold the possibility of investigating and testing the relationship between multiple variables, thus testing theory. Here, cause-effect relationships can be established with more credibility (Johnson and Onwuegbuzie, 2004), compared to other methods. A crucial advantage of using objective over subjective means of measurement is minimizing variance from participations interpretations and researcher bias. Despite including objective measurements as a as part of the experimental approach, the user experiment is not objective. Participants have a tendency to form an interpretation of the experiment's purpose based on experimental artifacts and alter their behavior accordingly (Rosenthal and Rosnow, 2009), and efforts should be made to counteract this bias. As mentioned previously, one cannot observe a phenomena without somewhat affecting it. Acknowledging this fact, a general discussion on the required level of objectivity for the study's purpose should be present to support selecting methods.

8.1.2 The output of one method can and is the input of another

A qualitative approach can be used for theory building and concept development (Eisenhardt, 1989; Gioia et al., 2013; Yin, 2017). The aim is often to generate novel testable theoretical constructs. Such theoretical construct can be formulated as a hypotheses to enable testing. Empirical data, for example in a narrative form can also be the output of a qualitative endeavor.

A literature investigation can inform empirical studies (Blessing and Chakrabarti, 2009), most commonly done by identifying research gaps and unfinished theories (Fink, 2005), from which new research directions can be proposed.

Using big data to engage in an exploratory data analysis can generate hypotheses for further testing, in addition to test already formulated hypotheses. By having hypotheses and suitable data, statistical inquiry is one means of testing this theory. Emerging or established theory can also be tested or tired by qualitatively finding additional empirical examples of a phenomena in the field.

As can be seen, entire research projects covering a topic can be governed by considering the input and output of the methods, matching required output from one method to the input of another. For example, to begin exploring a new research direction single-or multiple case studies can be utilized. A qualitative case study can lead to a testable hypothesis. Hypothesis testing require reliable quantitative data, which can be obtained by an experiment incorporating user testing. Statistical tests assess to what degree the hypothesized relationship exists and based on results here one could continue plotting a research direction.

All methods have their specialization, advantages making them highly applicable and appropriate for investigating areas of a phenomena, while falling short to cover other areas. The author acknowledges the inherent strengths and limitations to all methods, as such they should be used together to take full advantage of the depth provided by their advantages while cowering their weaknesses by using other methods. This is also the argument of mixed methods researchers (Creswell and Clark, 2007; Johnson and Onwuegbuzie, 2004). The output of one method can and is the input of another. Making this consideration could assist in selecting and mixing methods. Exactly how to integrate or mix the methods further is up to the research design.

8.1.3 Effective communication between methods

Morgan (Morgan, 1998) raises questions regarding the notion of an inability for effective communication between methods due to their incommensurable nature. Morgan states that this is a pure empirical question, therefore by investigating what it takes to combine qualitative and quantitative methods we would know if and what is possible. This thesis have begun to collect some insights in this respect. The author believes that effective communication between methods is possible. As discussed, each method have one or more

possible types of input and output. By acknowledging that other methods perform better in certain aspects, one can correctly formulate a studys' output to be handed over to serve as the input of another method.

8.1.4 How to not mix methods

A failed attempt at incorporating a quantitative method exemplifies how the lack of a carefully formulated theoretical construct caused a surveys' inability to capture the construct the author attempted at measuring. Therefore, the answers could not be used as intended. Explicit definition and operationalization of theoretical concepts is important for study reliability and validity (Abowitz and Toole, 2010), the survey provides an example of how to not mix methods. Procedure, rigor and craftsmanship is important in conducting any method and should not be compromised. Data collection is extremely important as the validity of the result inherently depend on the validity of the data and data collection method and results are only as good as these.

8.1.5 Mixed methods for complex phenomena – followed by a shortage of research designs

Complex phenomena in a real-world setting are just that – complex. There are multiple levels of investigation, where one can go in great depth or consider the phenomena from an overarching perspective. Having a range of methods available to investigate each level thoroughly was found to be valuable by the author as complementary approaches provide complementary insights. A mixed methods approach is recommended for investigating complex phenomena due to this reason, which the author adheres to. However, the established research designs are not necessarily suited to do so. There are many models and research designs in mixed methods research, most of which aim to cover one specific part of the "research journey", while a minority of research designs have a broader aim. If a topic is new or relatively unexplored there is an entire research journey to complete. To do so one must use a multitude of established models to complete the research journey since they only cover parts of the journey by themselves or take a different approach to research design by true mixing of methods. Either way, there is a shortage of mixed method research design for complex and uninvestigated phenomena. Complex phenomena and context require a genuine mixed method approach, where a topic is investigated from different perspectives and multiple methods, as each method hold potential to add unique and complementary insights.

8.1.6 Select an appropriate research design according to topic and research question

To some degree the author advocates a pluralistic approach to true mixed methods research. By that we mean that we believe in a diversity of views, standards and methods, rather than one single method or approach. However, awareness of the methods appropriateness; what they require as input, what they produce as output, advantages and limitations is crucial. An appropriate approach to a mixed method research design depend on the topic and the research question.

If a topic is known in a field the existing knowledge is somewhat saturated, however some research gaps could remain. Here, one can select an established research design with a purpose that complements the method already used and that corresponds to the research gap. This is a distinctly different case than the complex and uninvestigated phenomena discussed above.

Thorough reflections and considerations must be made with research question and study purpose as the primary guiding principle for selecting research designs and mixing methods. One should plan a set of methods carefully tailored to the overarching aim of the research, while maintaining an openness to that alternative approaches, interpretations, analysis techniques etc, may assert itself as a better fulfillment of the research purpose as the study progresses. If this is the case, plans should be changed by adapting or changing the method. Furthermore, the author advocates a courageous attitude towards making such a changes in plans. It may be easier and more practical to stick with initial plans, despite greater potential gains in terms of research outcome by making changes. In other words, the risk related to sticking to the initial yet inadequate method may be much greater than the risk related to change. Therefore, if one is uncertain if it is 'worth' changing or adapting a method, the question of risk could be used as a guide. An effective adaptation of methods and research design can enable *true triangulation* (Morgan, 1998) and should be sought after.

8.2 Skill requirement for researchers

8.2.1 A variety of skill requirements for a variety of methods

All research approaches require a separate set of skills and tools, from the overarching methodology and accompanying methods, down to use of specific software and programming language. To master an individual research method there are procedures and

strategies to apply, in addition to certain skills and abilities the researcher possesses or acquires. In conducting a literature investigation; the ability to quickly read academic articles, distil crucial information benefits the research process and the outcome of it in that it is quicker and/or concise. The ability to jointly consider different perspectives could aid in synthesizing information and forming constructs for further investigation, while a purist attitude could inhibit the researcher since it limits the process of forming a construct from multiple points of view. In a qualitative semi-structured interview, the interviewer (the researcher most often) must be able to adapt to the interview participant to shift the line of inquiry and allow for new interpretations and unexpected insights, while balancing a fine line of bias caused by reflexivity. Decision making is inherently a challenge. In qualitative work the researcher must return to their data in an iterative manner, to compare and corroborate constructs, explanations and interpretations (Given, 2008) to establish validity. There are no clear guidelines for when to end this process. The same is true for ending an exploratory data analysis, since there is not clear ending or final answer. As such, the researcher must be jointly capable of handling the opportunities and challenges inherent to the exploratory data analysis and balance it with confirmatory data analysis.

8.2.2 Requirements posed on researcher by mixing the methods

When conducting two or more methods simultaneously, there are additional requirements posed on the researcher compared to conducting a monomethod. The procedures, strategies, skills and capabilities for each single method must simultaneously be kept at the front of the mind, there is most likely several theoretical fields to consider, posing a challenge for the cognitive capabilities and the working memory of the researcher. The mental energy required for making shifts in methodology, method, theory, software and syntax etc., is not to be diminished. This can be, and for the author it was, at times really challenging. Despite this, the author finds the gains of using multiple techniques to outweigh the discussed disadvantages. When struggling with a problem seemingly without a solution, it can be comfortable to shift to a different method where workflow is better. The ability to continue performing productive work, while letting the other problem rest in the back of the mind, is comfortable compared to having to continue working on that problem. This is assuming that one is able to make such a compartmentalization, which is not always the case. An additional incentive for conduction distinct methods simultaneously is research supporting the ‘the Aha! moment’. In this thesis, the project for paper 2 and 3 were conducted simultaneously and the author used their distinct software to compartmentalize the case

study from statistical testing. Neuroscience proposes that new and diverse stimuli, whether its sport, traveling, new languages, or perspectives on a complex issue are promote the creation of new brain mass and the establishment of new neural connections.

There is a variety of skill requirements for a variety of methods and additional requirements are posed on the researchers by mixing methods, in terms of theory, procedure and cognitive capabilities.

8.3 Simultaneous vs sequential use of methods

The case study and the statistical testing was conducted simultaneously, bringing advantages and challenges as discussed above. In doing so the author was able to make comparisons that would not have been the same if all methods had been conducted in a *sequential* way (Creswell and Clark, 2007). For example, the author would filter the bad experiences for the good ones and devote an unproportionate amount of attention to the latter method(s) in writeup. The possibility to jointly conduct and consider the methodological approaches, with its requirements for skills, procedure, detail level, etc., have been enriching, in that much have been learned in a short timeframe while providing some key insights to which research direction to pursue and what methods to focus on. Conducting the methods simultaneously in a mixed method approach can be valuable in a state where a researcher can decide which direction to pursue and which methodology, methods and tools to bring along their research journey. Naturally, this adds complexity in terms of balancing two distinctly different sets of techniques and in that capability requirements (skills to be learned) more than double.

8.4 Using the principle of flexibility across methods and in mixed methods research design

When conducting interviews for the case study the interview questions naturally changed as the research process progressed, both within-interview-setting and between each interview. This flexibility enabled shifts in focus and zooming in on themes, (Gioia et al., 2013; Yin, 2017), allowing to probe in on emergent themes and take advantage of arising opportunities (Eisenhardt, 1989). The author found this to be of high value for the research progress. Since the stakeholders interviewed exhibited such distinct characteristics having to adhere to a standardized set of questions would have left out important insights and details necessary for interpreting stakeholders motivation for participation in the network.

In turn, the research outcome would have suffered. The flexibility principle is claimed to be what makes interpretive research good at uncovering new concepts and theories (Eisenhardt, 1989; Gioia et al., 2013; Yin, 2017).

Developing the experiment was a flexible and iterative process (Thomke and Reinertsen, 1998). Quick response to external and internal changes, combining and testing multidisciplinary components together enabled early discovery and solution of interference problems (Gerstenberg et al., 2015; Leikanger et al., 2016). This was valuable as it aided in handling uncertainty and complexity, and it enabled a rapid and smooth transition from piloting to running the experiment.

Mixed methods could be suited to integrating multidisciplinarity in a study, as it is able to handle the multiple levels of a phenomena by utilizing many principles – one of which is the flexibility one. The flexibility principle is a part of mixed methods by being inherent to the qualitative approach, which it should be as it brings the abovementioned advantages. Since the flexibility principle is a part of the mixed method approach it may make sense to also utilize it in the research design, guiding the integration or mixing of methods based on responding and adapting to emerging insights or changes. Paired with continuous reflections regarding the methods performance with regards to study purpose, input, output, etc., applying the flexibility principle to research design would result in a research design that is not preset, but that adapts to the research process as it progresses. Such a flexible research design could further enhance the mixed method research approach's ability to investigate broad and complex phenomena such as the context in this thesis.

True triangulation (Morgan, 1998) can be enabled by effective adaptation of methods and research design. A flexible research design as described can therefore enable true triangulation as a part of a genuine mixed method approach, if it is effective.

8.5 Reliability of raw data to ensure validity

A ground assumption vital to making correct judgements from statistical inference is reliability of data tested (Given, 2008; Walpole, 2012). Ensuring reliability of raw data in quantitative research can be done by correct design and development of data collection instrument and procedure, as well as careful attention to the collection process. In qualitative research some of this reliability issue is handled by interview subjects being treated as knowledgeable agents (Gioia et al., 2013). Furthermore, it is important to deliberately consider what amount of attention and emphasis should be devoted to which

participant or interpretation, while also considering the other data sources. When surveys are employed in an experiment one can observe a coherence between self-assessment and their behavioral responses, which could provide useful details in subsequent interpretation of the participants behavior and in ensuring validity of their self-reported data. Here, the experiment participants answers are treated as "true" to a similar extent as the knowledgeable agent assumption. Potential influencing or rival factors could and should also have been included in the survey to create an impression of their level of influence.

Reliability of raw data is vital for validity of results in any research approach including mixed methods. Since there are multiple and distinct data sources and means for data collection, reliability must be ensured for each individual method. When making inferences across the methods reliability of raw data remains crucial as errors here would be magnified as interpretations are carried over from method to method, and since the method is not consistent the order of magnification is unknown. There is no telling what kinds of errors may be in the result. Validity of such a study would be highly questionable, if not non-existent.

8.6 The notion of the greener grass in selection research method

Having conducted four research projects with distinct methods and one particular contextual framing, a discussion of which method is more favorable have become apparent, one that discusses aspects beyond the mere purpose of the method and to what degree the method serves that purpose. The purist approach to research methodology exists in both qualitative and quantitative domains, both on a top level and down to the level of one specific analytic technique. Both research domains seem to have a notion that one of the two general approaches is, in the lack of a better word, *easier* to conduct. What it interesting is that it is not always clear if this notion is used as an argument for adhering to the selected approach or against selecting another approach. Quality assessment of methods are made without regard to the purpose and functionality inherent to the method. Examples of impressions left with the author include several variations of "The grass is greener on the other side since it is easier to conduct that method, however I'm adhering to my method because it is the better method." or "The grass is greener on this side, because my method is better" or "This seems like the easier method to conduct, therefore I will select it and adhere to it from now on out". This discussion is extremely interesting to tap into, though

the author believes this should not be the primary discussion to have when one selects methods.

Descriptive statistics will tell what method is more common and accepted in the operating research field. This provide some indication of what methods currently serve the research questions in the field and what is generally acceptable, though it in principle does not address if a method is suited to answering the research questions. A researcher may have an inherent preference towards one particular method, which can and should be considered in determining the research methods that is to be a part of their academic endeavor. However, it is crucial to considering other aspects. Research domain and topic, previous coverage of the topic, the aim of the study and is this a desirable aim, will the methods serve the purpose if the study, what may the outcome from a different method be and could that be interesting etc., are aspects to be considered in various degrees when selecting methods.

Careful selection of method, with a continued reflection of how the method performs throughout the research project is advocated by the author. In other word, the author advocates a deliberate selection of methods to mix.

9 Concluding remarks

This section encompasses concluding remarks from the discussion. Results for researchers, limitations and future work are presented.

9.1 Results regarding mixed methods for researchers

Conducting multiple methods poses great requirements on a researcher in terms of skillset, and cognitive abilities. In practice it may be next to impossible for one single researcher, or even research group to conduct the methods properly. This is not an excuse to exclude methods, nor compromise on academic rigor. Therefore, the handover between the methodological approaches and the specific methods must be appropriate.

There is a shortage of mixed methods for complex and uninvestigated phenomena since established research designs usually are targeted towards one part of a research process or have a solitary purpose. For uninvestigated phenomena one does not yet know what insights are to be found and preselecting a research design may jeopardize results.

The author advocates a pluralistic approach to true mixed methods research. By that we mean that we believe in a diversity of views and methods, rather than one single method or approach. Awareness of the methods appropriateness; with regards to required input, produced output, advantages and limitations is crucial for careful initial selection of a method. Deliberate selection of methods to mix should be combined with a continued reflection of how the method performs throughout the research project. Combined with applying the flexibility principle to research design the resulting research design is one that is not preset, but that adapts to the context and the progress of the research process. An effective flexible research design enable true triangulation as a part of a genuine mixed method approach. Such a flexible research design can further enhance the mixed method research approach's ability to investigate broad and complex phenomena such as the context in this thesis.

9.2 Limitations

The specific context contextual background is not consistent throughout the paper as they were conducted based on projects available and suited to the method selected. Usage, comparison and discussion of the different methodological approaches is the focus. Of course, a longitudinal study investigating one consistent context could have provided a

better foundation for method comparison. The master thesis and the considerations made here are limited by this.

Limitations inherent to the studies conducted for paper 1 – 4 comprising this thesis, can be found in each individual paper.

9.3 Future work

The author recommends a longitudinal study with a complex phenomena as a consistent contextual background that adopts a flexible research design. A better assessment of a true mixed method approach can be made by doing so.

Bibliography

Abowitz, D., Toole, T., 2010. Mixed Method Research: Fundamental Issues of Design, Validity, and Reliability in Construction Research. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000026

Ausrød, V.L., Sinha, V., Widding, Ø., 2017. Business model design at the base of the pyramid. J. Clean. Prod. 162, 982–996. https://doi.org/10.1016/j.jclepro.2017.06.014

Balters, S., Steinert, M., 2017. Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. J. Intell. Manuf. 28, 1585–1607. https://doi.org/10.1007/s10845-015-1145-2

Balters, S., Steinert, M., 2015. Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. J. Intell. Manuf. 1–23.

Barney, J., Wright, M., Ketchen, D.J., 2001. The resource-based view of the firm: Ten years after 1991. J. Manag. 27, 625–641. https://doi.org/10.1016/S0149-2063(01)00114-3

Baron, R.A., Henry, R.A., 2010. How entrepreneurs acquire the capacity to excel: Insights from research on expert performance. Strateg. Entrep. J. 4, 49–65.

Blessing, L.T., Chakrabarti, A., 2009. DRM, a design research methodology. Springer Science & Business Media.

Bocken, N.M.P., Short, S.W., Rana, P., Evans, S., 2014. A literature and practice review to develop sustainable business model archetypes. J. Clean. Prod. 65, 42–56. https://doi.org/10.1016/j.jclepro.2013.11.039

Borseman, M., Tanev, S., Weiss, M., Rasmussen, E.S., 2016. Lost in the canvases: Managing uncertainty in lean global startups. Presented at the ISPIM Innovation Symposium, The International Society for Professional Innovation Management (ISPIM), p. 1.

Browning, T.R., 2003. On customer value and improvement in product development processes. Syst. Eng. 6, 49–61. https://doi.org/10.1002/sys.10034

Chesbrough, H., 2010. Business Model Innovation: Opportunities and Barriers. Long Range Plann., Business Models 43, 354–363. https://doi.org/10.1016/j.lrp.2009.07.010

Creswell, J.W., 1999. Mixed-method research: Introduction and application. Handb. Educ. Policy 455–472.

Creswell, J.W., Clark, V.L.P., 2007. Designing and conducting mixed methods research.

Creswell, J.W., Miller, D.L., 2000. Determining Validity in Qualitative Inquiry. Theory Pract. 39, 124–130. https://doi.org/10.1207/s15430421tip3903_2

Demil, B., Lecocq, X., 2010. Business Model Evolution: In Search of Dynamic Consistency. Long Range Plann., Business Models 43, 227–246. https://doi.org/10.1016/j.lrp.2010.02.004

Donaldson, K.M., Ishii, K., Sheppard, S.D., 2006. Customer Value Chain Analysis. Res. Eng. Des. 16, 174–183. https://doi.org/10.1007/s00163-006-0012-8

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. Acad. Manage. Rev. 14, 532–550. https://doi.org/10.2307/258557

Fink, A., 2005. Conducting Research Literature Reviews: From the Internet to Paper. SAGE.

Gemmell, R.M., Boland, R.J., Kolb, D.A., 2012. The Socio-Cognitive Dynamics of Entrepreneurial Ideation. Entrep. Theory Pract. 36, 1053–1073. https://doi.org/10.1111/j.1540-6520.2011.00486.x

George, G., Bock, A.J., 2011. The Business Model in Practice and its Implications for Entrepreneurship Research. Entrep. Theory Pract. 35, 83–111. https://doi.org/10.1111/j.1540-6520.2010.00424.x

Gerstenberg, A., Sjöman, H., Reime, T., Abrahamsson, P., Steinert, M., 2015. A Simultaneous, Multidisciplinary Development and Design Journey – Reflections on Prototyping, in: Entertainment Computing - ICEC 2015, Lecture Notes in Computer Science. Presented at the International Conference on Entertainment Computing, Springer, Cham, pp. 409–416. https://doi.org/10.1007/978-3-319-24589-8_33

Gioia, D.A., Corley, K.G., Hamilton, A.L., 2013. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organ. Res. Methods 16, 15–31. https://doi.org/10.1177/1094428112452151

Given, L.M. (Ed.), 2008. The Sage encyclopedia of qualitative research methods. Sage Publications, Los Angeles, Calif.

Gonick, L., 1993. The cartoon guide to statistics. HarperPerennial, New York.

Hamari, J., Sjöklint, M., Ukkonen, A., 2016. The sharing economy: Why people participate in collaborative consumption. J. Assoc. Inf. Sci. Technol. 67, 2047–2059. https://doi.org/10.1002/asi.23552

Hart, C., 2001. Doing a literature search: a comprehensive guide for the social sciences. Sage.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, in: Hancock, P.A., Meshkati, N. (Eds.), Advances in Psychology. North-Holland, pp. 139–183.

Jalote, P., Palit, A., Kurien, P., Peethamber, V.T., 2004. Timeboxing: a process model for iterative software development. J. Syst. Softw. 70, 117–127. https://doi.org/10.1016/S0164-1212(03)00010-4

Jensen, M.B., Semb, C.C.S., Vindal, S., Steinert, M., 2016. State of the Art of Makerspaces - Success Criteria When Designing Makerspaces for Norwegian Industrial Companies. Procedia CIRP 54, 65–70. https://doi.org/10.1016/j.procir.2016.05.069

Johnson, R.B., Onwuegbuzie, A.J., 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. Educ. Res. 33, 14–26. https://doi.org/10.3102/0013189X033007014

Kahneman, D., Tversky, A., 1984. Choices, values, and frames. Am. Psychol. 39, 341–350.

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47, 263–291. https://doi.org/10.2307/1914185

Kaplan, R.M., Chambers, D.A., Glasgow, R.E., 2014. Big Data and Large Sample Size: A Cautionary Note on the Potential for Bias. Clin. Transl. Sci. 7, 342–346. https://doi.org/10.1111/cts.12178

Larman, C., Basili, V.R., 2003. Iterative and incremental developments. a brief history. Computer 36, 47–56.

Leikanger, K.K., Balters, S., Steinert, M., 2016. Introducing the wayfaring approach for the development of human experiments in interaction design and engineering design science. Presented at the DS 84: Proceedings of the DESIGN 2016 14th International Design Conference.

Maurya, A., 2012. Running lean: iterate from plan A to a plan that works. O'Reilly Media, Inc.

Mont, O.K., 2002. Clarifying the concept of product–service system. J. Clean. Prod. 10, 237–245. https://doi.org/10.1016/S0959-6526(01)00039-7

Morgan, D.L., 1998. Practical strategies for combining qualitative and quantitative methods: Applications to health research. Qual. Health Res. 8, 362–376.

Morgan, J.M., Liker, J.K., 2006. The Toyota product development system : integrating people, process, and technology. Productivity Press, New York.

Morris, M., Schindehutte, M., Allen, J., 2005. The entrepreneur's business model: toward a unified perspective. J. Bus. Res., Special Section: The Nonprofit Marketing Landscape 58, 726–735. https://doi.org/10.1016/j.jbusres.2003.11.001

Onwuegbuzie, A.J., Teddlie, C., 2003. A framework for analyzing data in mixed methods research. Handb. Mix. Methods Soc. Behav. Res. 2, 397–430.

Osterwalder, A., Pigneur, Y., 2010. Business model generation: a handbook for visionaries, game changers, and challengers. John Wiley & Sons.

Osterwalder, A., Pigneur, Y., Tucci, C.L., 2005. Clarifying Business Models: Origins, Present, and Future of the Concept. Commun. Assoc. Inf. Syst. 16, 28.

Östlund, U., Kidd, L., Wengström, Y., Rowa-Dewar, N., 2011. Combining qualitative and quantitative research within mixed method research designs: A methodological review. Int. J. Nurs. Stud. 48, 369–383. https://doi.org/10.1016/j.ijnurstu.2010.10.005

Priem, R.L., Li, S., Carr, J.C., 2012. Insights and New Directions from Demand-Side Approaches to Technology Innovation, Entrepreneurship, and Strategic Management Research. J. Manag. 38, 346–374. https://doi.org/10.1177/0149206311429614

Ries, E., 2011. The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses. Crown Books.

Ritter, T., Lettl, C., 2018. The wider implications of business-model research. Long Range Plann. 51, 1–8. https://doi.org/10.1016/j.lrp.2017.07.005

Sanders, M.S., McCormick, E.J., 1987. Human factors in engineering and design. McGRAW-HILL book company.

Sarasvathy, S.D., 2001. What makes entrepreneurs entrepreneurial?

Sibi, S., Baiters, S., Mok, B., Steiner, M., Ju, W., 2017. Assessing driver cortical activity under varying levels of automation with functional near infrared spectroscopy. Presented at the Intelligent Vehicles Symposium (IV), 2017 IEEE, IEEE, pp. 1509–1516.

Steinert, M., Leifer, L.J., 2012. "Finding One's Way": Re-Discovering a Hunter-Gatherer Model based on Wayfaring. Int. J. Eng. Educ. 28, 251.

Thomke, S., Reinertsen, D., 1998. Agile product development: Managing development flexibility in uncertain environments. Calif. Manage. Rev. 41, 8–30.

Tortorella, G.L., Marodin, G.A., Fettermann, D. de C., Fogliatto, F.S., 2016. Relationships between lean product development enablers and problems. Int. J. Prod. Res. 54, 2837–2855. https://doi.org/10.1080/00207543.2015.1106020

Tukey, J.W., 1977. Exploratory data analysis. Reading, Mass.

Tukker, A., 2015. Product services for a resource-efficient and circular economy – a review. J. Clean. Prod., Special Volume: Why have 'Sustainable Product-Service Systems' not been widely implemented? 97, 76–91. https://doi.org/10.1016/j.jclepro.2013.11.049

Tukker, A., 2004. Eight types of product–service system: eight ways to sustainability? Experiences from SusProNet. Bus. Strategy Environ. 13, 246–260. https://doi.org/10.1002/bse.414

Tukker, A., Tischner, U., 2006. Product-services as a research field: past, present and future. Reflections from a decade of research. J. Clean. Prod., Product Service Systems: reviewing achievements and refining the research agenda 14, 1552–1556. https://doi.org/10.1016/j.jclepro.2006.01.022

Walpole, R.E., 2012. Probability & statistics for engineers & scientists, 9th ed. ed. Pearson, Boston, Mass.

Welo, T., 2011. On the application of lean principles in Product Development: a commentary on models and practices. Int. J. Prod. Dev. 13, 316–343. https://doi.org/10.1504/ijpd.2011.042027

Welo, T., Ringen, G., 2017. Lean assessment and transformation strategies in product development: a longitudinal study. 87-2 Proc. 21st Int. Conf. Eng. Des. ICED 17 Vol 2 Des. Process. Des. Organ. Manag. Vanc. Can. 21-25082017.

Wernerfelt, B., 1984. A resource-based view of the firm. Strateg. Manag. J. 5, 171–180. https://doi.org/10.1002/smj.4250050207

Woodson, W.E., Conover, D.W., 1970. Human engineering guide for equipment designers. Univ of California Press.

Wulvik, A., Erichsen, J., Steinert, M., 2016. Capturing Body Language in Engineering Design – Tools and Technologies. 85-1 Proc. Nord. 2016 Vol. 1 Trondheim Nor. 10th - 12th August 2016.

Yin, R.K., 2017. Case study research and applications: Design and methods. Sage publications.

Zott, C., Amit, R., 2010. Business Model Design: An Activity System Perspective. Long Range Plann., Business Models 43, 216–226. https://doi.org/10.1016/j.lrp.2009.07.004