ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

MATHEUS JARDIM BERNARDES

ON THE UNDERSTANDING OF THE ROLE OF EXPERIMENTATION IN TECHNOLOGY-BASED STARTUPS

Porto Alegre

2021

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica do Rio Grande do Sul

PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL
SCHOOL OF TECHNOLOGY
COMPUTER SCIENCE GRADUATE PROGRAM

ON THE UNDERSTANDING OF THE ROLE OF EXPERIMENTATION IN TECHNOLOGY-BASED STARTUPS

MATHEUS JARDIM BERNARDES

Thesis submitted to the Pontifical Catholic University of Rio Grande do Sul in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Prof. Sabrina Marczak, PhD

Porto Alegre
2021

Ficha Catalográfica

B522o Bernardes, Matheus Jardim

On the Understanding of the Role of Experimentation in Technology-based Startups / Matheus Jardim Bernardes. – 2021.
83.
Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Sabrina dos Santos Marczak.

1. Software Engineering. 2. Agile Software Development. 3. Lean Startup. 4. Continuous Experimentation. 5. Technology Entrepreneurship. I. Marczak, Sabrina dos Santos. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051

Matheus Jardim Bernardes

On the understanding of the role of experimentation in technology-based startups

This Master Thesis/Doctoral Thesis has been submitted in partial fulfillment of the requirements for the degree of Doctor/Master of Computer Science, of the Graduate Program in Computer Science, School of Technology of the Pontifícia Universidade Católica do Rio Grande do Sul.

Sanctioned on the 24th of August, 2021.

COMMITTEE MEMBERS:

Prof. Tiago Silva da Silva, PhD (ICT/UNIFESP)

Prof. Afonso Sales, PhD (PPGCC/PUCRS)

Prof. Sabrina Marczak, PhD (PPGCC/PUCRS - Advisor)

“If I have seen further it is by standing on the shoulders of Giants.”
(Isaac Newton)

ACKNOWLEDGMENTS

I want to thank all of those who, somehow, helped me advance through this challenging endeavor.

This work certainly would not be possible without the support of the marvelous people that surround me. I thank my wife, parents, sister, and nephew for all their enormous patience, understanding, and love throughout this journey.

To my advisor, Sabrina Marczak, for sharing valuable and lasting advice during the past months.

To all my research colleagues for their caring contributions and support.

To PUCRS and the Technology School, for all the infrastructure offered during this work and in the last decade.

To the interviewees that participated in this research, by the time they shared with me and their precious contribution.

The past few months were especially arduous for unprecedented reasons. I do hope better times will come for all of us.

COMPREENDENDO O PAPEL DA EXPERIMENTAÇÃO EM STARTUPS DE BASE TECNOLÓGICA

RESUMO

Startups de tecnologia surgem constantemente para tentar criar soluções inovadoras em ambientes de extrema incerteza e, por enfrentar inúmeros desafios, possuem altos índices de falência. Entre os principais desafios estão a escassez de recursos e falta de aderência do produto com as necessidades do mercado. Na tentativa de suavizar estes desafios surgem iniciativas como a Experimentação Contínua. Esta abordagem faz com que hipóteses sejam testadas de forma sistematizada, ajudando equipes a priorizar as entregas que agregam maior valor aos usuários. Neste contexto, esta dissertação apresenta uma revisão da literatura seguida de um estudo de campo onde se procurou identificar como a Experimentação Contínua está sendo adotada e como ela interfere no ciclo de desenvolvimento de software em startups de base tecnológica. Para tanto, foram entrevistados 16 membros de *startups* de base tecnológica de maneira semi-estruturada a fim de explorar e caracterizar papéis, processos, ferramentas, desafios, benefícios, entre outros. Os resultados coletados nas entrevistas foram confrontados com as informações encontradas na revisão da literatura para formular respostas às questões de pesquisa e propor trabalhos futuros para o tópico seguir sendo explorado. Sendo assim, o objetivo desta dissertação foi o de caracterizar o uso dessa abordagem em *startups* de base tecnológica, além de identificar desafios e benefícios percebidos por profissionais que atuam nesse cenário.

Palavras-Chave: Engenharia de Software, Desenvolvimento de Software Ágil, Lean Startup, Experimentação Contínua, Empreendedorismo Tecnológico.

ON THE UNDERSTANDING OF THE ROLE OF EXPERIMENTATION IN TECHNOLOGY-BASED STARTUPS

ABSTRACT

Technology startups are constantly emerging, trying to create innovative solutions in environments of extreme uncertainty, and because they face numerous challenges, they have high failure rates. The scarcity of resources and the product's lack of adherence to market needs are among the main challenges. In an attempt to alleviate these challenges, initiatives such as Continuous Experimentation arise. This approach supports systematical tests of hypotheses, helping teams prioritize deliveries that increase perceived value by the users. In this context, this thesis presents a literature review followed by a interview-based study that focused on identifying how Continuous Experimentation is being adopted and how it underlies software engineering activities throughout the product development cycle of technology-based startups. We conducted semi-structured interviews with 16 members of technology-based startups to explore and characterize roles, processes, tools, challenges, benefits, among others. Results were compared with the information found in the literature review to formulate answers to the research questions and propose future work to further explore the topic. Thus, this thesis aimed to characterize the use of this approach in technology-based startups and identify challenges and benefits perceived by professionals working in this scenario.

Keywords: Software Engineering, Agile Software Development, Lean Startup, Continuous Experimentation, Technology Entrepreneurship.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

B2B – Business-to-Business

BML – Build-Measure-Learn

CD – Continuous Deployment

CE – Continuous Experimentation

CI – Continuous Integration

LR – Literature Review

MVF – Minimum Viable Feature

MVP – Minimum Viable Product

OCE – Online Controlled Experiments

OEC – Overall Evaluation Criteria

RE – Requirements Engineering

RQ – Research Question

SE – Software Engineering

US – User Story

XP – Extreme Programming

CONTENTS

1. INTRODUCTION

This chapter presents the motivations for carrying out this research. It describes the research goal and research questions that guided the formulation of the adopted methodology, which is also briefly portrayed here, followed by the main contributions accomplished.

1.1 Context and Motivation

In the past few decades, software engineering practitioners have adopted various project management tools, practices, and methodologies. The initial methodologies had heavy processes that usually impose costly changes to requirements. But in fact, organizations need to continually adjust their workflows and demands to compete in the market. This scenario required more flexible alternatives to project management, and companies started to propose new project management options.

Independent initiatives started to gain attention around the '90s with new approaches to tackle project management more flexibly. Those initiatives later received the name of 'Agile' and nowadays there are diverse agile methodologies such as Scrum and Extreme Programming (XP). Pressman [50] highlights that agile methodologies tend to focus on collaboration, delivering early versions of the final product, and engaging the customer in the development process.

Even though Agile methodologies have shown significant benefits against previous methodologies, it is still noticeable that some challenges need to be addressed. Pressman [50] highlights that practitioners of these methodologies seek, among other benefits, to accelerate the development cycle and enable faster deliveries. But as Yaman et al. [71] suggest leaning on agility does not guarantee that the solution created will add value to the customers. In some cases, the delivered software generates negative or no value, even though the agile tools and methodologies are correctly followed.

In a traditional Software Engineering (SE) environment, product development begins by eliciting requirements. Sommerville [59] highlights that the goal of this activity is to create a clear and well-defined definition of the desired solution. Eventually, some requirements will come directly from users. Kujala et al. [34] state that user involvement brings positive impacts in the requirement engineering process, helping to create a product that generates more value to the users.

Technology-based startups are constantly developing products with a defined target group in mind but with no specific user to get involved in the process. To overcome this challenge, entrepreneurs formulate hypotheses about potential user needs and design solutions to test these ideas through experimentation[1]. Different approaches to executing such tasks appeared to help startups evaluate their business propositions.

In 2011, Eric Ries [52] received attention in the entrepreneurship ecosystem when releasing his book entitled 'The Lean Startup'. Ries and some other authors started the now called 'Lean Startup' movement: a series of initiatives that emerged intending to reduce waste of resources while producing a product while coping with extreme uncertainty and increasing the perceived value. To achieve that, practitioners implement sequential 'experiments' in a systematic and iterative approach, collect data, and plan the next steps with those findings. Previously proposed approaches for achieving an experiment-driven development, including the ideas brought by Ries, are also known as 'Continuous Experimentation' (CE).

CE is defined, in a nutshell, as an approach based on continuously identifying critical assumptions, transforming them into hypotheses, prioritizing and testing them with 'experiments' following the scientific method to support or refute them [39].

The fast-paced environment that technology-based startups reside in is imposing changes on the well-established SE practices to accommodate CE. The proper execution of CE fosters the adoption of CD [22, 56], new specialties (e.g., data scientists and DevOps engineers) are becoming essential team members [21], sales and 'Customer Success' personal are more frequently involved in the engineering process, among others. These companies are gradually merging CE into their SE activities while validating business hypotheses.

Although such practices have become highly popular with entrepreneurs, it is noticeable that a significant number of startups fail either because they have exhausted their resources or have developed products with no market fit. In both cases, CE is expected to reduce the chances of failure due to such reasons.

This research fits in this scenario of elevated startup mortality rates due to factors that could be softened with popular CE practices [36]. Earlier studies on the subject revealed the need to investigate the startup ecosystem to characterize the adoption of CE and its impacts on the software engineering process.

1.2 Research Problem

Emerging countries such as Brazil have alarming numbers regarding startup bankruptcy. Besides the entrepreneurial challenges, a 2020 study placed Brazil as 124 out of 190 countries when comparing the difficulty in business regulation [4]. The study highlights both labor market and tax regulation as significant obstacles for the entrepreneurship

[1] We are using the term 'experiment' (surrounded by single quotes) to denote the idea of *experimentation* or experimental studies. In fact, experiments represent a fraction of the *experimental methods*. By doing so, we expect to emphasize the common misuse as a synonym by industry members.

ecosystem. Among others, these issues contributed to approximately 30% of all Brazilian technology-based startups ending their activities within the first year of existence [36].

The two main reasons why startups end activities are “Product or solution does not meet a real market need” (lack of market fit) and “Shortage of capital or lack of planning when using resources” (resource wasting) [8].

Considering CE became highly popular in the past few years and was meant to address issues exactly like these two aforementioned, this scenario seems contradictory: either startup companies are not adequately using experimentation, or the method itself is not very useful in solving those issues.

1.3 Research Goals and Questions

This thesis aims to characterize the adoption of CE by technology-based startups. More specifically, we aim to identify how experimentation is currently used by such startups, which practices are popular among practitioners, and challenges and benefits perceived by startup members. We defined a set of Research Questions (RQ) to guide the planning and execution of our activities while conducting this research, starting with:

RQ1 How do technology-based startups implement Continuous Experimentation?

By answering RQ1, we intend to identify the conditions in which entrepreneurs are currently implementing CE. This RQ mainly focuses on listing methodologies of adoption CE alongside technologies that support such practices. Therefore, RQ1 will guide us in understanding ‘how’ CE is currently being adopted. To deepen our comprehension of the research topic, we propose two secondary RQs, as follows:

RQ2 What are the perceived benefits while adopting Continuous Experimentation for technology-based startups?

We expect to comprehend the benefits of adopting CE observed by technology-based startups while answering RQ2. As a result, we expect to formulate a list of relevant positive impacts that CE may cause when adopted. Such a list can support future entrepreneurs when deciding whether they should embrace these practices in their endeavors.

RQ3 What are the challenges while implementing Continuous Experimentation for technology-based startups?

The third RQ seeks to complement RQ2 by identifying the challenges perceived in CE’s adoption by our interest group. Similarly to RQ2, we expect to formulate a comprehensive list of the challenges experienced when adopting CE. Once again, entrepreneurs can benefit from such an outcome by analyzing our findings and considering the implications of embracing such practices in their projects.

1.4 Research Methodology Overview

To accomplish the aforementioned research goal and define answers to the posed research questions, we conducted a three-activity research process based on a qualitative approach. It started with a Literature Review (LR) based on the Snowballing guidelines [65] (Activity 1). The goal of this activity was to comprehend the current understanding of authors about Lean Startup and CE in the Software Engineering (SE) industry.

The second activity was based on semi-structured and exploratory interviews with technology startup members (Activity 2), which aimed to qualitatively explore the research problem from the perspective of those who act upon and live in it in the industry. Such interviews were recorded, and their audio was transcribed for posterior analysis. This data set was analyzed following the Thematic Synthesis approach [9].

Finally, conclusions were drawn by contrasting the results of such analysis with the LR findings, which was essential to help us formulate answers to the posed RQs.

1.5 Main Contributions

The primary contribution of this research concerns characterizing how technology-based startups are currently adopting CE, the perceived benefits experienced by practitioners when conducting CE, and the perceived challenges on the subject.

This research compares results found in the literature with data gathered on our industry-based study aiming to provide a comprehensive list of models used to implement CE, commonly used technologies and services that support such practice, popular perceived benefits and challenges, sources of information where entrepreneurs get educated on the topic, among other aspects.

These results collected from real cases are relevant for understanding the impacts of such practices in industry. They can potentially serve as the foundation for future research. Such results can also be used as valuable input for entrepreneurs assessing the consequences of adopting CE in their businesses.

1.6 Document Outline

The remainder of this document is organized as follows:

- **Chapter 2 - Theoretical Foundation**: presents the theoretical background on the research topic;
- **Chapter 3 - Research Methodology**: describes how this research was organized and the purpose of each activity;
- **Chapter 4 - Literature Review**: describes the results obtained during the execution of the literature review;
- **Chapter 5 - Interview-Based Study**: provides details on our findings during the interview-based study;
- **Chapter 6 - Result Analysis**: presents the answers to the research questions using the results obtained in the interview-based study while comparing it to the findings of the literature review;
- **Chapter 7 - Final Consideration**: concludes this thesis by summarizing our findings and outlining possible future research on this topic.

2. THEORETICAL FOUNDATION

This chapter presents background concepts that are related to the topic of this thesis, which are: Requirements and Hypothesis Engineering (Section 2.1), The Lean Startup (Section 2.2) and CE in Software Engineering (Section 2.3).

2.1 From Requirements to Hypotheses Engineering

Preece et al. [49] define requirement as a statement about an intended product that specifies what a system should do and how it should be performed. In a traditional SE scenario, projects are conducted in a requirement-driven manner: it means that a project's success is conditioned upon which degree it fits the users' expectations [43]. To meet such expectations, it is essential to elicit the user's desired behaviors and restrictions that the system must attend to. Sommerville [59] states that the process in which system requirements are identified, analyzed, and defined is known as Requirements Engineering (RE).

Alexander [1] highlights that developing software without involving those who will use it, especially during the requirements elicitation phase, puts the whole project at risk. However, in the highly innovative domain of technology-based startups, companies are constantly developing disruptive products based on hypotheses about market needs instead of actual direct customer demands.

Tripathi et al. [63] argue that due to restricted time and resource constraints, startups constantly enumerate requirements mostly in a self-invented manner, rarely documented, and validated only after product releases. Besides that, software startups frequently avoid following traditional RE practices or even build their own customized processes, adapting them according to the faced circumstances [39].

In this entrepreneurial scenario, practitioners came up with numerous approaches for designing their business assumptions as requirements that can be implemented and posteriorly used to conduct 'experiments'.

Melegati et al. [41] state that in this experiment-driven development scenario, such hypotheses guide the development and execution of 'experiments' to allow entrepreneurs to learn more about the market. Moreover, this approach could soften the lack of a user's presence during RE activities, potentially reducing the waste of resources while implementing solutions with a poor market fit.

This intricate relation between requirements and the hypotheses can also be found in some CE implementation models. For example, when proposing the RIGHT model for conducting CE, Fagerholm et al. [21] state that requirements should evolve in real-time based on actual users' data. Melegati, Wang and Abrahamsson [41] propose the discipline of ‘Hypotheses Engineering’ compared to classical requirements engineering, to start reducing the gap among those related topics.

2.2 The Lean Startup

Stavru [61] highlights that agile methodologies have become popular among software development organizations nowadays. But as Yaman et al. [71] suggest, the delivered software generates negative or no value in some cases, even though the appropriate agile methodologies are ‘process-wise’ followed correctly.

In that sense, new approaches for developing products while reducing waste of resources have emerged, such as The Lean Startup. The Lean Startup is an entrepreneurial set of principles for developing products proposed by Eric Ries in 2011. It relies on a systematic and iterative process in which hypotheses are tested to generate ‘validated learning’. Ries [52] claims that with his methods companies can create order from chaos by using its tools to test a vision continuously.

This methodology not only had its name inspired on Toyota’s ‘Lean Manufacturing’ system [44], but it also has the reduction of waste in its core value similar to its predecessor. Womak et al. [67] define waste as “*[..] any human activity which absorbs resources but creates no value*”. The five principles of The Lean Startup are:

1. Entrepreneurs are everywhere;
2. Entrepreneurship is management;
3. Validated learning;
4. Build-Measure-Learn;
5. Innovation accounting.

The first principle proposes that entrepreneurship can happen at any place, not only in a startup environment. According to this, smaller groups inside a more significant organization can use The Lean Startup to deal with high uncertainty scenarios.

The second principle states that even though the problem being solved is unknown and unpredictable, management should not be chaotic. Ries [52] suggests that rigorous processes should be used to convert hypotheses into knowledge.

The process proposed in The Lean Startup in which ideas can be tested and converted into ‘knowledge’ is called the ‘Build-Measure-Learn loop’. That concept is expressed in the fourth principle, and it is visually represented in Figure 2.1. It starts with an idea

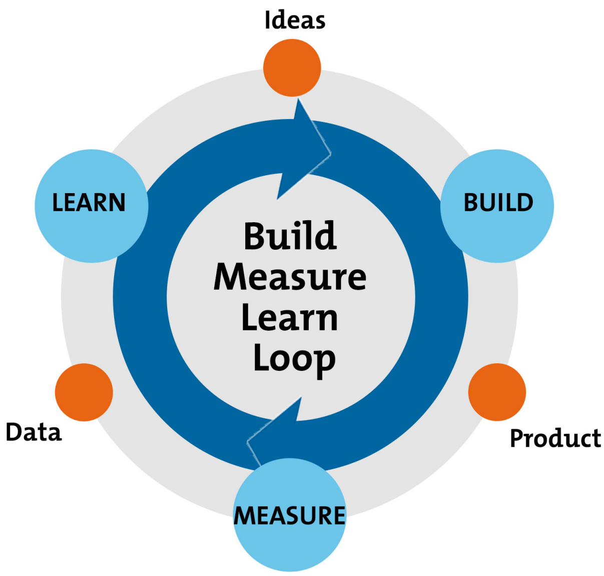

Figure 2.1 – Build-Measure-Learn Loop
Diagram adapted from 'The Lean Startup' [42].

(hypotheses) that needs to be tested. To execute those tests, a piece of software well-instrumented to enable such tests is built, i.e., a Minimum Viable Feature (MVF) or a Minimum Viable Product (MVP). The product is then put into test with real users, and as much data as possible is collected to measure the acceptance and effectiveness of the proposed solution. Then the data is analyzed to define if the initial assumptions are confirmed, and the 'experiment' should be kept as part of the solution (persevere), or it shows that the initial idea was incorrect, and the solution should be put aside (pivot). In either case, the 'experiment' generates knowledge as an output.

The third principle expresses how the author calls the output of each Build-Measure-Learn (BML) loop. 'Validated learning' [52] is the empirical result obtained from experimentation that can be considered the truth. This acquired knowledge helps the organization decide when persevering into an assumption or start testing a new one.

The outcomes of validated learning are separated into four categories:

- Pivoting: After conducting an 'experiment' and discarding a hypothesis, pivoting is the action to radically change one or more aspects of the business model in order to formulate a new hypothesis and test it through new 'experiments'.

- Iterating: It is a less radical change when compared to pivoting. Due to the lessons learned one can improve one or more changes in the business model or product to test new hypotheses.

- Escalating: In this situation the entrepreneurs believe that they have found a sustainable business model and are willing to invest more in the business.

- Giving up: It happens when experimentation shows that the business vision set is not able to generate a sustainable business model.

The last principle regards the systematic and disciplined approach in which the company can establish metrics to measure the whole progress, and it provides inputs for the future's organization planning.

2.3 Continuous Experimentation in Software Engineering

The modern economy requires companies to continually invest in technology and digitalization of processes to attend with the rapid changes in business scenarios. Many companies are born highly dependent on software as its main competitive advantage. Automatization through software turns the work activities more flexible, and services can be delivered faster to customers. This high demand for software technology made supply grow as well, lowering the cost of creating software.

The main challenge most companies face nowadays is no longer how to solve technical issues, but rather how to prioritize development to convert efforts efficiently into value to their customers [20]. In the entrepreneurial environment, the procedure to identify and prioritize potential customer's demands were traditionally based on guesswork or in centralized planning with no empirical data to support it. However, some successful companies have come up with systematic and methodological approaches to address this challenge.

Yaman et al. [71] define as Continuous Experimentation, all systematic and iterative approaches based on field experiments and data collection during the development of products or services. One popular approach proposed to deal with this endeavor is the aforementioned BML loop. Once the concepts present in The Lean Startup became popular, many authors proposed alternatives or modifications to the original BML loop (e.g. [5,21,46]).

Kevic et al. [26] cite that multiple tools support conducting experimentation and gather data from users. Münch et al. [35] defend that proper tools and techniques should be selected according to the goal of the 'experiment', the type of hypothesis that is under tests, and its context. Table 2.1 presents a summary of popular tools for conducting experimentation and corresponding brief description.

Olsson et al. [46] suggest that hypotheses should be organized and prioritized in a process similar to feature requests in a product backlog. Such hypotheses are generated and managed by the product management and product development staff based on their understanding of customer needs and strategic business goals.

Kohavi et al. [31] demonstrate that to implement a CE process multiple challenges need to be addressed, and those challenges can be grouped into 3 areas: cultural and organizational, engineering, and trustworthiness. In short, the reason and aim for executing

Table 2.1 – Definition of Tools Supporting CE

Name	**Definition**
A/B testing	Experiment conducted with 2 variations of the same software where the effectiveness of a single change is measured [33].
Canary Release	Technique which includes gradual software release while testing new features on a small subset of customers [56].
Dark launches	Approach in which a new software version is released to production, but initially no actual user access it directly. Instead, user requests are both sent to a server running the older version and the new one, the results are compared without exposing users to the new version [56].
Fake door tests	Technique that displays call-to-actions of non-existing features to the user and then measures how many users are actually interested in the subject, before actually developing it [35].
Landing pages	The first web page that a customer lands (usually coming from search engines) that presents a product or a tease of a future product and captures access metrics to validate users interest in a given topic.
Mockups and wireframes	Prototyping techniques used to represent future products and, in the CE context, can be used to to validate data from user interaction [21].
MVP and MVF	Minimum portion of a product (MVP) or of a feature (MVF) which is used in the experimentation process and has the necessary instrumentation to collect data for posterior data analysis [20].
Multivariate tests	Technique used to test the effectiveness of multiple variations in a given software with users. It's similar to A/B testing but it is not restricted to test single variations at once [33].

CE should be apparent across the organization, proper technology must be in place to collect data in a way it can be analyzed *à posteriori*, and the 'experiments' must be conducted with proper rigor to guarantee that the obtained results are reliable.

Dmitriev et al. [12] indicate that big companies such as Amazon, Facebook, and Google can afford to invest in in-house experimentation systems. At the same time, small startups need to lean on more affordable alternatives. Fabijan et al. [16] list some of these experimentation systems that are available on the internet to instrument software such as Optimizely.com, Mixpanel.com, and Oracle Maxymiser[1].

Distinct technologies and methodologies are intertwined into the practices of CE. As previously mentioned, RE is directly related to CE, considering entrepreneurs should propose their 'experiments' to optimize learning about customers with minimum waste of resources. Simultaneously, the BML loop present in The Lean Startup is constantly referred to as a base for CE models proposed by various authors. Continuous Integration (CI) and Continuous Deployment (CD) are concepts not covered in this research; however, these subjects are also widespread among CE publications. Many authors mention CI/CD as tools to support CE implementation [21, 22, 56].

[1] Available at https://www.oracle.com/br/cx/marketing/personalization-testing/

3. RESEARCH METHODOLOGY

This chapter presents the research design and methodology used in this thesis research work. Section 3.1.1 presents the Literature Review activity conducted using the Snowballing approach [65]. Section 3.1.2 details the procedure adopted in the interview-based study, where interviews with startup members were carried out in a semi-structured manner. Such interviews were later consolidated with the aid of the Thematic Synthesis approach [9], and findings were contrasted with the ones of the LR to formulate answers to the posed research questions.

3.1 Research Design

To accomplish the research goal and answer the RQs, both described in Section 1, we conducted a research process that started with a literature review followed by an interview-based study. All the data collected in both activities served as input while drawing conclusions. Figure 3.1 presents a visual representation of the overall process.

Activity 1, the Literature Review, was conducted according to the guidelines of the Snowballing method [65], and its goal is to comprehend the current understanding of authors about Lean Startup and CE in the Software Engineering (SE) industry. Such activity was fundamental in the process of formulating the interview script.

Activity 2, an Interview-based activity focusing on technology startups, aimed to qualitatively explore the research problem from the perspective of those who act upon and live in it in industry. We took advantage of our location at PUCRS and interviewed represen-

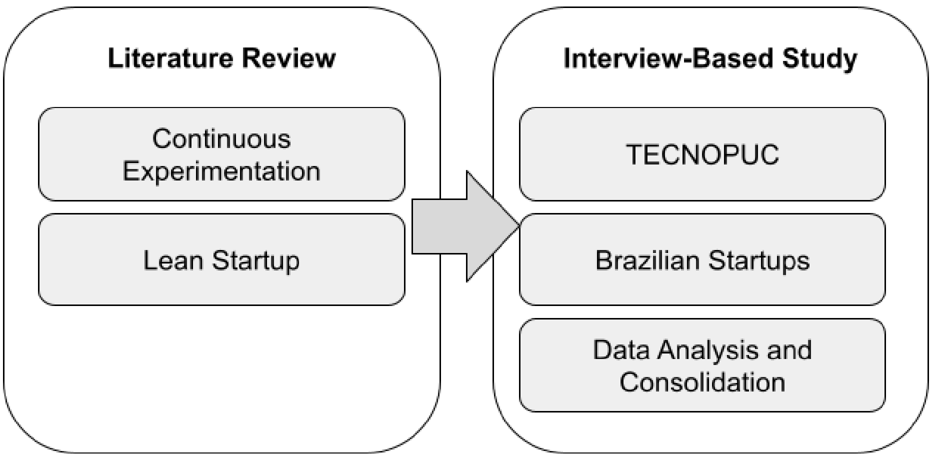

Figure 3.1 – Research Plan Proposal Activities

tatives of technology-based startups at the University Technology Park, TECNOPUC. We later expanded our investigation to companies at a national level (e.g., Brazilian startups).

We analyzed the data gathered during the interviews following the Thematic Synthesis approach, proposed by Cruzes et al. [9]. The findings of this process were then compared and contrasted with the ones of the literature review, aiming to answer our RQs.

3.1.1 Literature Review

This activity aimed to identify the state-of-the-art in the topic of study and get familiar with recent publications and best practices in the topic. The literature review also contributed to answering the proposed research questions.

The Snowballing method was chosen to guide such activity. Snowballing is the literature research method proposed by Wohlin [65] in which a systematic and rigorous procedure is executed to cover existing literature regarding one topic.

In short, the method consists of executing an iterative process such as in each cycle a group of papers is evaluated following predefined methods to determine if they should be included or not for posterior data extraction. It starts with an initial set of papers as candidates, and then new candidates are elected from related ones. These related papers may come from two sources which can be:

- backward snowballing: in this stage, candidates are listed from the reference list of previously selected papers.
- forward snowballing: in this stage, papers that cite previously selected ones are evaluated as candidates.

Wohlin [66] describes that the main advantage of this approach is that the search for papers starts with relevant publications, and these papers guide the search for new potentially relevant ones. The Snowballing procedure is easy to comprehend and to reproduce, which makes it possible to extend the original literature review in future updates. Besides that, our research group has previously executed this method with positive results, with colleagues highlighting the time efficiency obtained with the technique.

3.1.2 Interview-Based Study

In this activity, a qualitative approach was taken in an interview-based study aiming to collect data from real projects assessing practitioners of technology-based startups. These interviews were essential to contribute to answering the proposed RQs.

The interview candidates must have had experience working with currently active startups or even with ceased ones. By listening to participants of ceased startups, valuable data can be collected regarding the reason for bankruptcy.

To maximize efforts, we initially looked for startup incubation centers to obtain candidates. A high concentration of currently operating startups and contacts from previously incubated companies can be found in such places. Companies may leave incubation by failure or extreme success: both cases could be included in the interviews.

The interviews were conducted in two phases, as follows:

I. We started looking for startups with relations to the Tecnopuc park. This incubation center was selected due to its national relevance and proximity to the university in which this research is being conducted.

II. We later looked for companies sited in multiple locations within national borders. To be more productive, we used the existing relations of our research group with the PulseHub incubation center. Furthermore, to reach out to more candidates for interviewing, we replicated the approach also used in our research group of using the professional-oriented social network, LinkedIn[1] [51].

There are two reasons for splitting the interviews: first, the interview itself could be tested and adjusted (if necessary) with a smaller group of companies. This concept relates directly to the BML loop so that we could *build* a set of questions, execute interviews and *measure* effectiveness, and then *learn* from previous experience to better execute with a bigger audience. The second reason for splitting the activity is that it becomes possible to contrast regional results with national results.

The interviews were carried out in a semi-structured and exploratory manner. Seaman [58] highlights that these interviews use a mixture of open-ended and specific questions to elicit not only the information that is foreseen but also unexpected information. This approach will allow participants to discourse more openly about the central topic. The main script will guide the interview sessions, but the participants were instigated to express themselves in an open environment.

After executing the interviews and transcribing their audio, we analyzed the gathered data following the Thematic Synthesis approach proposed by Cruzes et al. [9]. This method provides a systematic way of identifying, analyzing, and reporting patterns using large qualitative databases in SE. The overall process consists of reviewers moving iteratively among 5 steps, as displayed in Figure 3.2: Extract data, Code data, Translate codes into themes, Create a model of higher-order themes, Assess the trustworthiness of the synthesis. The method proposes a series of checklists that guide the execution of each step.

[1]Available in: https://www.linkedin.com.

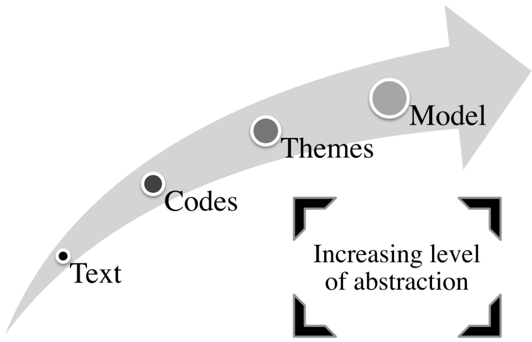

Figure 3.2 – Levels of Interpretation in Thematic Synthesis
Adapted from [9]

Cruzes and Dyba [9] suggest that the coding process should be either:

1. *Deductive or A Priori Approach*: where data is coded following a predefined code set;
2. *Inductive or Grounded Theory Approach*: in this approach becomes, data is thoroughly analyzed, and, as a concept becomes apparent, a code is assigned. An iterative process can be used to review later created codes;
3. *Integrated Approach*: both predecessors approaches are combined in this third one. A start list of codes is improved with an inductive procedure.

We chose to conduct this activity using the integrated approach. The literature review previously executed generated initial answers to our RQs. Such answers were then translated into an initial set of codes that could be used in the thematic analysis. Since the initial set of codes was not enough to represent our data collected, new codes emerged while performing the activity, which is an inherent step of the chosen integrative approach.

4. LITERATURE REVIEW

This chapter presents the process adopted while executing the literature review activity, details metrics collected while performing the Snowballing, and details the findings achieved in this activity.

4.1 Literature Review Procedure

Figure 4.1 illustrates the steps followed while conducting this activity. The RQs listed in Section 1.3 were achieved after conducting an unstructured literature review, in which multiple papers in the corresponding subject were evaluated in an exploratory manner. This step allowed us to identify initial definitions and understanding between the author of this thesis and her advisor.

After that, the Snowballing procedure was defined as a guideline to conduct the literature review. This approach was chosen based on two major aspects: firstly, due to the positive feedback provided by our research group's colleagues, highlighting the time efficiency obtained while adopting this technique. Secondly, as Wohlin [66] points out that, the Snowballing procedure is easy to comprehend and reproduce, which makes it possible to extend the original literature review in future updates.

The literature review initially identified 142 papers in database searches, which narrowed to 54 candidates after more careful exclusion criteria and concluded the process with 33 papers selected.

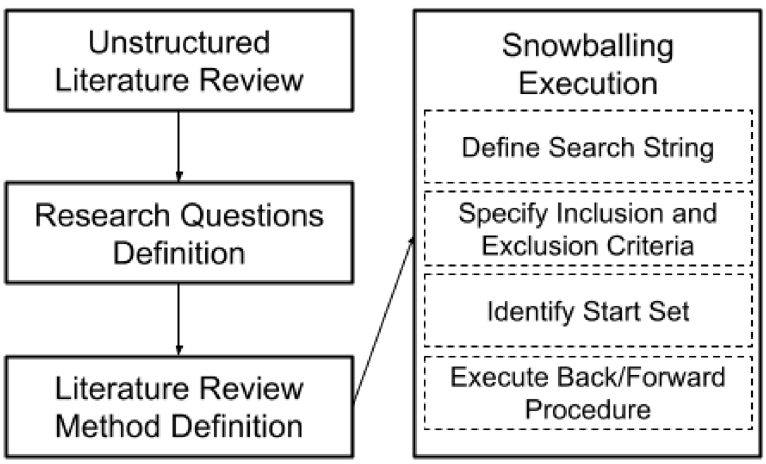

Figure 4.1 – Literature Review Methodology

Diagram adapted from [65].

Table 4.1 – Search String

Population	(Startup OR Software Engineering OR Software Development)
	AND
Intervention	(Experimentation OR Continuous Experimentation)
	AND
Outcome	(Experiment-driven development OR Data-driven development)

Next, metrics were extracted and later analyzed (see Section 4.6). Such metrics can be useful to measure efficiency and evaluate papers' coverage over a given subject. Finally, the selected papers were thoroughly evaluated, and relevant information was collected to compose the answers to the proposed research questions alongside our interview-based study findings in Section 6.

The aforementioned study produced a technical report [24], which was published in partial fulfillment of the requirements for the degree of Master in Computer Science.

4.2 Database and Search String

Wohlin [65] proposes that to start the Snowballing procedure, it is necessary to define a start set of papers. This can be achieved by testing search strings in tools such as Google Scholar. Additionally, Wohlin [65] suggests that such a search string can be formulated following the guidelines proposed by Kitchenham et al. [28].

Kitchenham et al. [28] propose a set of criteria that frame the structure of such search strings. These criteria were adapted from medical studies guidelines to match the viewpoint of software engineering.

The first aspect adopted in this search string is 'Population', which relates to the target group of interest. The second one, 'Intervention', links to the software methodology/tool/technology/procedure aligned with our topic. And the last aspect, 'Outcome', links to the relevant outcomes expected to be found in the literature to be analyzed. Table 4.1 shows the final search string.

4.3 Inclusion and Exclusion Criteria

As previously mentioned, Snowballing requires criteria for inclusion and exclusion of candidates while evaluating the papers. The inclusion criteria for this research is:

- contains at least one of the terms: 'Software Engineering' or 'Software Development';
- must have a title, keywords, or an abstract explicitly aligned with the research topic;

- paper must be written in English;
- contains the terms: ‘Experimentation’ and ‘Startup’;
- must have its content fully accessible;
- must present clear contributions to the research question somehow.

The criteria for exclusion follows as:

- must not be redundant (already evaluated previously);
- must not address CE as a tool for testing hypotheses during development process;
- must not be considered grey literature (i.e., must be peer-reviewed).

4.4 Initial Set

It is essential to create a good starter set of papers to perform the Snowballing method properly. But as Wohlin [65] states, there is no silver bullet for identifying a good one. After applying the search string into Google Scholar, we have selected eight papers as candidates. These are the five candidates that were included for posterior analysis:

1. ‘Introducing Continuous Experimentation in Large Software-Intensive Product and Service Organisations’ by Yaman at al. [71].
2. ‘Raising the Odds of Success: The Current State of Experimentation in Product Development’ by Lindgren et al. [35].
3. ‘User Involvement in Experiment-Driven Software Development’ by Yaman et al. [68].
4. ‘Hypotheses Engineering: First Essential Steps of Experiment-Driven Software Development’ by Melegati et al. [41].
5. ‘Current State of Research on Continuous Experimentation: A Systematic Mapping Study’ by Auer et al. [3].

And these are the 3 papers discarded for the start set:

1. ‘Challenges and Strategies for Undertaking Continuous Experimentation to Embedded Systems: Industry and Research Perspectives’ by Olsson [37].
2. ‘Initiating the Transition Towards Continuous Experimentation: Empirical Studies with Software Development Teams and Practitioners’ by Yaman [70].
3. ‘From MVPs to Pivots’ by Khanna et al. [27].

Table 4.2 – Papers by Iteration

Iteration	Snowballing	Papers
0	Start set	[3,35,41,68,71]
1	Backward	[5,20,21,31,45,53]
1	Forward	[15,19,30,32,38,55–57,64]
2	Backward	[12,23]
2	Forward	[11,16–18,22,26,32,40,54]
3	Forward	[39]
4	Backward	[14]

4.5 Iterations

After defining the start set, we started executing the backward and forward snowballing. The overall process lasted for five iterations (start set and more four iterations) to complete the literature review.

Table 4.2 displays which papers were reached in each iteration. As expected, the first and second iterations have appended a significant amount of papers. As a result, the subsequent iterations reached many papers that were already previously selected.

4.6 Overall Efficiency

Wohlin [65] points out that efficiency is an important aspect of literature review approaches. In this context, efficiency is measured by the ratio between the total number of selected papers and the total number of candidates.

Considering all the iterations, the overall efficiency in the process was 60%. The efficiency achieved by iteration was:

- Start set: 5 / 8 = 62.5%;
- Iteration 1: 15 / 25 = 60%;
- Iteration 2: 8 / 12 = 66.7%;
- Iteration 3: 1 / 4 = 25%;
- Iteration 4: 1 / 1 = 100%;
- Iteration 5: no new papers.

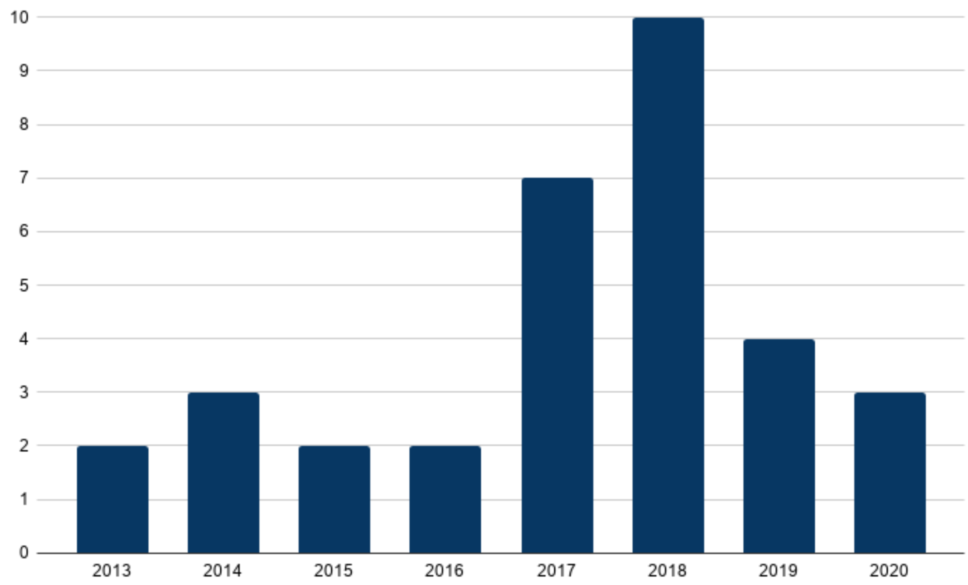

Figure 4.2 – Distribution of Papers by Year of Publication

4.7 Citation Matrix, Publications Timeline and Categories

Wohlin [65] defends the importance of creating a citation matrix, which is a visual representation of the relationship between the selected papers. The author proposes that a matrix with many empty cells may indicate a lack of papers in the literature review.

Table 4.3 represents the relationship between the papers selected for this Snowballing. For example, it reads that the paper [53] cites the papers [20,31], seen in the marking with ‘X’ in the appropriate cells. Cells that contain a dash symbol (‘-’) are greyed out and represents that a paper cannot be referenced because it was not published yet.

This table offers 2 contributions to the original proposal of Wohlin [65]. The first one is regarding the fact that the papers are presented orderly according to their publishing year. This modification helps to visually identify the newer papers, which will evidently receive fewer citations. The publication year can be seen in the very first row of the table.

The second modification is regarding the last row and column added to the table, representing the sum of cells containing an ‘X’ both horizontally and vertically, respectively. This contribution helps the reader to faster identify, for example, that papers [20] and [21] were both cited 10 times and paper [14] cited 10 other papers in this selection.

Figure 4.2 offers a visual representation of the publishing year of the papers selected in this literature review. We highlight the year 2018, which had the most papers (total of 10), representing 30.3% of all years.

Table 4.3 – Citation Matrix

Pub.	2013		2014			2015		2016		2017							2018										2019				2020			
Ref.	[5]	[31]	[20]	[45]	[18]	[53]	[22]	[35]	[11]	[71]	[21]	[15]	[16]	[26]	[12]	[14]	[3]	[30]	[57]	[56]	[55]	[69]	[17]	[54]	[32]	[23]	[41]	[19]	[40]	[39]	[68]	[64]	[38]	
[5]			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
[31]			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
[20]						-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
[45]						-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
[18]		X				-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
[53]		X	X					-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
[22]			X					-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
[35]		X	X	X		X	X			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
[11]			X				X			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
[71]	X	X	X	X		X					X						-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6
[21]																	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
[15]		X						X							X		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
[16]			X														-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
[26]		X									X						-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
[12]		X			X												-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
[14]		X	X		X	X	X	X			X	X	X		X		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10
[3]		X	X				X				X		X	X						X							-	-	-	-	-	-	-	7
[30]						X	X			X	X																-	-	-	-	-	-	-	4
[57]								X					X	X							X						-	-	-	-	-	-	-	4
[56]					X			X																	X		-	-	-	-	-	-	-	3
[55]						X	X	X			X																-	-	-	-	-	-	-	4
[69]			X			X		X																			-	-	-	-	-	-	-	3
[17]											X															X	-	-	-	-	-	-	-	2
[54]		X									X																-	-	-	-	-	-	-	2
[32]		X	X			X				X																	-	-	-	-	-	-	-	4
[23]		X																							X		-	-	-	-	-	-	-	2
[41]								X																							-	-	-	1
[19]														X	X		X			X						X					-	-	-	5
[40]				X							X																				-	-	-	2
[39]					X											X													X		-	-	-	3
[68]	X					X				X											X													4
[64]										X	X																							2
[38]								X		X		X							X															4
	2	12	10	3	4	8	6	8	0	5	10	2	3	3	3	1	1	0	1	2	2	0	0	0	2	2	0	0	1	0	0	0	0	**T.**

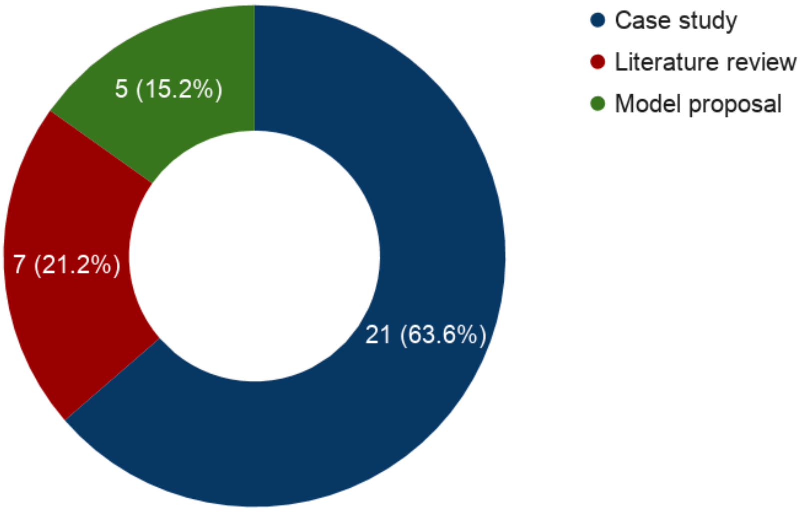

Figure 4.3 – Distribution of Papers by Subject

This chart also suggests that most papers are fairly recent: 48.5% of all papers were published in the past 3 years. It is clearly expected that papers published recently cannot be cited a substantial number of times, leaving the citation matrix more sparse.

This research is being carried out during the Sars-CoV-2 (COVID-19) global pandemic outbreak, and it is apprehensible that such circumstances may impact the productivity of some authors and Journal publishing companies. Companies like Elsevier are already assuming that this unprecedented situation might delay their activities[1]. It can be assumed that the growth of publishings found during 2020 may be impacted by the current scenario.

All papers included in the final selection were categorized according to their research proposal into 3 groups. The groups are:

1. Case study: representing all publications that involve gathering data from real projects and interviewing participants;
2. Literature review: characterizing publications that use the current literature to formulate answers to its research questions;
3. Model proposal: groups of all papers that present models and techniques for implementing Continuous Experimentation.

Figure 4.3 demonstrates the distribution of papers in such groups visually. The majority of the selected publications (approximately 63.6%) are dedicated to studying CE's current implementation in technology-based startups and large companies. In this context, future publications based on this work can contribute by expanding the second biggest category of 'Literature review'.

[1]Available in: https://www.journals.elsevier.com/expositiones-mathematicae/news/impact-of-covid-19-on-journal-publishing-processes

Table 4.4 – Popularity of Tools Supporting CE according to Literature Review

Name	Citations	Total
A/B Testing	[3,5,12,14–19,21–23,26,31,32,35,38–41,45,53–57,68]	27
CI & CD	[3,11,14–16,18,20–22,26,35,38,45,53,55–57,69]	18
MVP & MVF	[5,20,21,31,35,40,41,45,53,68,71]	11
Canary Release & Dark Launches	[14,19,23,26,38,40,56,57]	8
Mockups & Wireframes	[17,21,26,31,35,41,68,71]	8
Multivariate Tests	[17,19,31,35,45,53,55]	7
Landing Pages	[35,40,68]	3
Fake Door Test	[35]	1

4.8 Literature Review Findings

This section presents answers to the research questions (see Section 1.3) achieved based on the findings obtained during the LR. Each of the following sections formulates answers using bullet points, sorted by the frequency in which it was found in the literature and indicating the corresponding authors.

4.8.1 CE Adoption in Technology-Based Startups (RQ1)

The literature selected during Snowballing points that there is a variety of tools used to support the proper execution of experimentation while implementing CE in a project. Such tools can help teams to test hypotheses more efficiently and to increase the trustworthiness of gathered data.

Table 4.4 shows how many times the literature mentions these mechanisms for supporting CE. The group's most popular technique is 'A/B Testing', which was mentioned at least once in a total of 27 papers (presence of 81.8%). The second most cited technology mentioned is CI and CD, which are not specifically meant to conduct experimentation. Still, literature considers it a good practice that automatically supports experiments' deployment, thus reducing human interference in the results [26].

Besides tools, RQ1 is also concerned about models of implementing CE available in the literature. In the selected papers, we identified 5 models:

1. Early Stage Software Startup Development Model (ESSSDM), by Bjork et al. [5]

   This model proposes mechanisms to support testing multiple assumptions in parallel while identifying those worth scaling [5]. The model has 3 parts: idea generation (opportunities for generating value to customers), a prioritized idea backlog, and a process based on the BML loop to validate ideas systematically. In short, experiments are conducted, and the results and learnings are then documented to fed back into the business model. This process can potentially generate new hypotheses. Each BML iteration ends with a debate with team members to discuss whether assumptions tested should persevere or not.

2. Rapid Iterative value creation Gained through High-frequency Testing (RIGHT), by Fagerholm et al. [21]

   This approach is also based on the BML loop, with additional steps supported by a technical infrastructure [21]. In this approach, experiments are proposed from the product vision, which in turn comes from business strategy. The business strategy is made of assumptions to support creating an ideal business model. Experimentation is used to reduce uncertainties from such assumptions, reducing development risks. Hypotheses are tested through experiments in order to validate assumptions. A proper MVP/MVF is released to collect data of usage during a certain duration of time. The collected data is then analyzed to support future business planning.

3. QUESt, by Melegati and Wang [40]

   QUESt is a quality guideline for representing USs as hypotheses for experiment-driven software development. It is not a model for conducting experimentation but rather define hypotheses in a backlog in a similar approach already used with User Story (US). This guideline proposes a new template to write hypotheses as US's, recommending that it must have a Questioning sense, be Updatable, Evaluable, and Straightforward.

4. Experimentation Evolution Model, by Fabijan et al. [16]

   This model proposes a systemic approach to guide software organizations to implement CE and thus turn its development truly data-driven [16]. The model is based on a study conducted at Microsoft and inspired by a model developed at A&E.

   The approach is composed of 'three phases of evolution': technical, organizational, and business. The first phase focus on technical issues such as the complexity of the experimentation platform and the pervasiveness of experimentation in product teams. The second phase focuses on preparing data science teams and guaranteeing their self-sufficiency for experimentation. The third phase focus on the Overall Evaluation Criteria (OEC), which is a quantitative measure of a controlled experiment's objective.

5. The Experiment Growth Model, by Fabijan et al. [14]

   This model seeks to help companies conduct their first Online Controlled Experiment (OCE) and mature their experimentation capabilities through the 4 stages of experimentation growth: 'Crawl', 'Walk', 'Run 'Fly'. This four-staged framework addresses

7 aspects of CE, ranging from 'Experimentation platform capability' to 'Experimentation Impact'. This model was proposed and tested with the help of companies like Microsoft, Booking.com and Skyscanner.

4.8.2 Perceived Benefits While Using CE (RQ2)

Using our findings of the literature review, we summarized the benefits perceived by startups after implementing CE in the following four topics.

1. Data-driven decisions

Yaman et al. [71] highlight that collecting data from users was not introduced by CE, however connecting these data to proper 'experiments' makes it possible to make development decisions supported by data rather than by assumptions. Rissanen et al. [53] affirm that using data in the decision-making process helps the team to create a solution with a higher business value by better reacting to customer needs.

Fabijan et al. [14] support that conducting 'experiments' enabled companies to make decisions based on quantified and analytical hypotheses, moving away from personal experience and preferences. Olsson et al. [45] has reported that properly adopting a data-driven decision approach had enhanced productivity and motivation in development teams.

2. Reduced development effort

Yaman et al. [71] affirm that companies can reduce development efforts when adopting an experiment-driven development. Features can be dropped from project backlog when a hypothesis proves itself wrong, and the team decides not to persevere in it.

Rissanen et al. [53] reported that in a given case study, management was wrong about customer expectations 80% of the time. In this scenario, prioritizing development over data collected from users instead of the guesswork can significantly reduce the efforts of developing unwanted features.

Vargas et al. [64] show that in a given case study, the company reported that CE helped the team abandon the development of a feature that demonstrated itself not to be worth it. By pivoting to another hypothesis, companies can increase the efficiency of converting invested resources into value to customers.

3. Decentralization of business knowledge and facilitate knowledge transfer

Yaman et al. [71] reported that teams that properly executed CE as part of the development cycle conquered a knowledge database from previous 'experiments' that can be shared across the team, thus facilitating knowledge transfer.

Fabijan et al. [14] propose that companies conquer the ability to accumulate learnings from past ‘experiments’ when implementing CE properly. In this scenario, every team member can know previous ‘experiments’ and achieved results, profoundly impacting the development roadmap for future ‘experiments’.

The knowledge database generated after conducting multiple ‘experiments’ can be shared among products in the same organization. Fabijan et al. [15] defends that such findings can help the company define what is valuable to a whole portfolio and avoid repeating similar ‘experiments’ across products in the same portfolio.

4. Predicting infrastructure needs

Experimentation can be executed to help companies validate whether the projected infrastructure can handle new features. Fabijan et al. [14,15] analyzed a case study at Skype that decided to use CE to release a new implementation of an algorithm to a small portion of users instead of all customers at once. The engineering team was capable of collecting proper data to validate whether the predictions were met. This approach prevented the company from releasing a new change that could potentially negatively impact all customers without testing it previously, damaging its product’s brand.

4.8.3 Perceived Challenges While implementing CE (RQ3)

This section presents all the challenges of implementing CE found in the literature into 3 groups: technical, organizational, and customer challenges. These groups were originally proposed by Rissanen et al. [53], and we expanded them according to our new findings in the following subsections.

Technical Challenges

1. Trustworthiness and statistical significance of data

Companies running experimentation commonly face 2 related challenges: collected data must be trustworthy, and still some level of statistical significance must be achieved. Fagerholm et al. [20] highlight that running multiple experiments simultaneously may generate distorted results.

Lindgren et al. [35] suggest teams must ensure that relevant stakeholders carefully analyze the collected customer and product data. Fabijan et al. [14] defend that issues with data quality are common in CE and must be addressed as soon as possible due to its criticality to the whole experimentation.

2. Achieving rapid release cycles

Rapid and automated release cycles become extremely important when integrating CE to the product development workflow [35]. Once hypotheses are formulated, the development team must implement and deploy it to collect data and build learning over the 'experiment'. Longer development cycles may lead the team to lose track of running 'experiments' or lose valuable competitive advantage [22].

3. Low-level interface limits the 'experiments'

Rissanen et al. [53] highlight that different types of software products impose different technical challenges. For example, a service consumed by its users via a user interface offers more freedom to the development of experimentation with the user experience. On the other hand, an API service must follow a certain specification; otherwise, the customers would not consume it.

4. Infrastructural changes on mature projects to support CE

Rissanen et al. [53] argue that architectural changes are more costly in complex and mature software, so companies adopting CE in this scenario will face an additional challenge. Implementing infrastructure for collecting data and execute experiments may damage the experience of users already operating the system.

Organizational Challenges

1. Creating a innovative culture within the company

Rissanen et al. [53] defend that companies that plan to adopt CE must address social aspects such as organizational culture. Implementing experimentation workflow impacts multiple areas of a company, including non-engineering teams such as sales.

Multiple levels of management must also be aligned with the mindset of running experiments. As Vargas et al. [64] point out, team members may fear testing hypotheses when culture does not welcome failure.

2. Low maturity of processes and low education in experimentation

Yaman et al. [71] propose that companies frequently face difficulties in defining success criteria for experimentation. Poorly defined 'experiments' may lead to inconclusive or deceptive results. Lindgren and Munch [35] confirm that companies must address the social aspect of competence in defining and executing 'experiments' besides technical challenges. Fagerholm et al. [20] define that proper and rapid design of 'experiments' is crucial for adopting experimentation.

3. Prioritization of customers requests over experiments

Rissanen et al. [53] demonstrate that leadership may lower the priority of conducting experimentation when confronted with the demands of customers. Management may prioritize direct requests from paying customers over experimentation. This sort of event may impact running 'experiments' and have negative impacts on the CE culture.

Customer Challenges

1. Legal and ethical aspects must be addressed

Fabijan et al. [16] highlight that teams conducting experimentation must address the legal and ethical aspects involved. Legislation must be taken into account when designing experiments because such aspects may impact their execution. Yaman et al. [68] show that, in some cases, the lack of clear processes regarding legal and ethical issues lead practitioners to rationalize their perceptions based on their own experiences.

2. Complex stakeholder structures

Different business models may increase complexity when conducting experimentation. Rissanen et al. [53] demonstrate that solutions that are targeted to a business-to-business scenario must implement 'experiments' that consider multiple user roles and needs. Yaman et al. [69] show that products that already have customers may have to inform such stakeholders about major changes that are part of 'experiments', thus reducing user experience degradation.

3. Interruption of service during deployment of experiments

Mattos et al. [38] proposes that teams must focus on reducing downtime when deploying 'experiments' in projects that customers are already using. Long deployment processes during working hours may lead to interruption of service and degradation of user's experience. Rissanen and Munch [53] propose that the engineering team may try to negotiate with the customer or schedule deployments for moments of lower impacts on usage.

5. INTERVIEW-BASED STUDY

This chapter presents the procedure adopted while conducting the interview-based study and showcases the participants' profiles, transcription tooling, and process. Section 5.5 presents the generated themes and the respective findings associated with the participant's identification (e.g. 'P1', which stands for 'Participant 1').

5.1 Interview-Based Study Procedure

Figure 5.1 illustrates the steps followed while conducting the interview-based study. We decided to conduct semi-structured interviews to allow us to explore the topic. By using open questions and encouraging the participant to elaborate on their answers, we could deepen our understanding of the participant's perspectives on the matter.

Our interview script was designed inspired on the results of the literature review. The script was then presented to colleagues in our research group, and adjustments were made to reflect the valuable contributions received. Appendix C shows the final form of the interview script. Once the study was designed, we then submitted the request of approval to conduct the study to the PUCRS' Ethics Committee Board (see Appendix A).

Also, concerning ethics matters, every participant must agree upon the Consent Form (available in Appendix B) before participating in the interviews. Such form was written in Portuguese because that is the language used during the interviews.

A pilot interview was then conducted, aiming to test the whole process. It happened with the same tooling and measurements that would take place in the subsequent interviews. No significant changes were identified with this pilot, so we proceeded with the interviews.

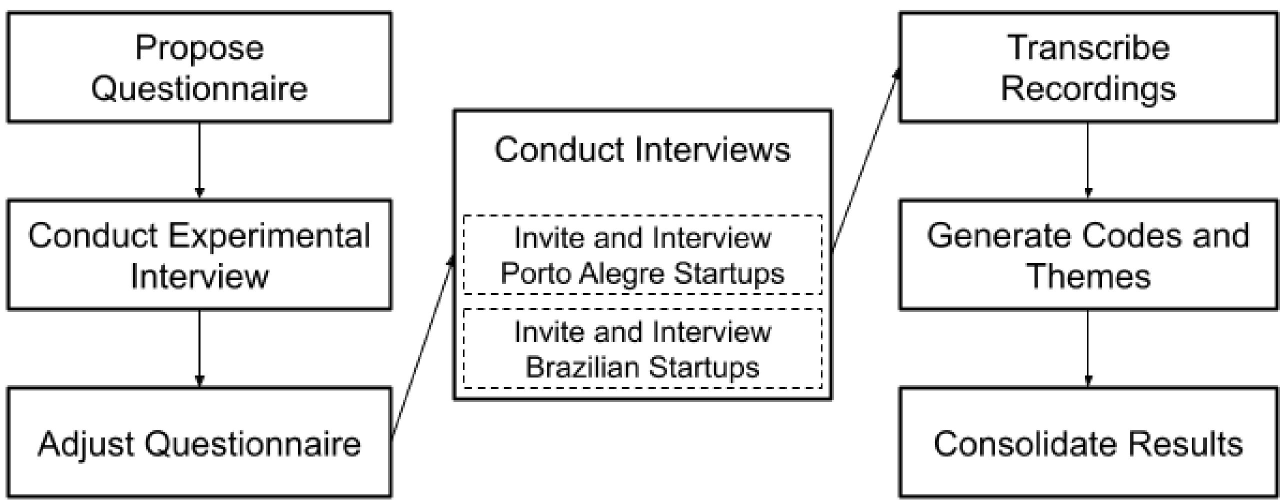

Figure 5.1 – Interview-Based Study Methodology

The remainder steps of the interview-based study are detailed as follows: Section 5.2 describe how the interviews were executed; and Section 5.4 details the transcription process; our codes and findings are analyzed in the Section 5.5.

5.2 Interviews Execution

All interviews occurred via the internet through video conferences using Google-Meet[1]. Even though many participants could attend the universities' facilities, we took this approach as security measurements to reduce risks related to the current global pandemic.

The interviews' audios were recorded using the open-source software OBS[2] for future transcription. All meetings occurred with one participant at a time. The conversations were performed and transcribed in Portuguese, considering that all researchers involved are fluent in the language.

We invited 51 members of technology-based startups to participate in the interviews. These candidates were reached using different methods, such as e-mail and direct messages on social networks (i.e., Linkedin). The participants' identities could be found on the startup's contact page or at the public portfolio of startups in the incubation centers.

A total of 16 out of the 51 invites were accepted, which means around 30% of the invites resulted in interviews. All the recordings summed up a total of 08:44 hours, resulting in an average duration of 32 minutes. The shortest interview lasted about 20 minutes, and the longest lasted 52 minutes.

We conducted debriefing sessions after each interview, looking for opportunities of improvement before executing the next interviews. This activity was composed of 8 questions and was inspired in previous research projects conducted by Prof. Cleidson de Souza, PhD (UFPA). Appendix D displays our addapted version of the debriefing questionnaire.

5.3 Characterizing the Participants' Profile

The selected participants of this study are technology-based startup members, primarily those that have involvement with product backlog management. We invited professionals who execute different software development roles to enrich our data collection through different perspectives. No underage professional participated in the interviews.

[1]Available at https://meet.google.com/
[2]Available at https://obsproject.com/

Table 5.1 – Participants' Profile

ID	Academic Background	Role	YOE in the IT Industry	Years in this Startup
P1	BS, Social Communication	Product Manager	6	3
P2	BS, Computer Science	Full Stack Developer	10	4
P3	BS, Control and Automation	iOS Developer	5	3
P4	BS, Computer Science	Product Designer	5	3
P5	BS, Biomedical and MBA, Sales Management	CEO	3	3
P6	BS, Computer Science and Postgraduate, Project Mgmt.	CTO	15	6
P7	MS, Administration and MBA, Business Management	CEO	4	3
P8	BS, Oil and Gas and MBA, HR Management	CEO	7	5
P9	BS, Mechanical Engineering	CEO	17	4
P10	BS, Mechatronics Engineering	CEO	11	5
P11	PhD, Mechanical Engineering	CEO	24	10
P12	BS, Agricultural Eng. and MBA, Business Marketing	Operations Manager	4	3
P13	BS, Agronomy and Postgraduate, Georeferencing	Data Analyst	11	1.5
P14	BS, Production Engineering	CEO	8	5
P15	BS, Marketing and Journalism	CMO	4	4
P16	PhD, Computer Engineering	CEO	8	4

Table 5.1 characterizes the profile of each participant. We gathered such data during the interviews directly from the participant. To better draw this table, we reduced 'Years Of Experience' as 'YOE'.

The 'Academic Background' column showcases that the selection of participants belong to a broad area of knowledge. Every interviewee has at least a bachelor's degree and half of them also have an additional formation, including 2 PhDs.

The 'Role' column suggests a concentration at the C-level job position[3] (10 out of 16), which was expected and foreseen prior to the interviews. This occurs mostly because these startup companies have a small number of members, and at least one of those positions is occupied by an executive officer (CEO).

While designing the interview script, we identified that some participants could work in 'single-product startups' and others only have contact with one of many solutions within their companies. To better represent the perspective of participants that may not know all projects within their company, we included questions to facilitate characterizing the participant's startup and its project.

[3]'C-level' (also known as 'C-suite') is an expression commonly used in the industry to refer to chief-level positions such as CEO, COO, CTO.

Table 5.2 – Company and Projects' Profile Data

	Company's Data						Project's Data					
ID	Co. Address	Co. Age	EC	CL	Biz. Model	Biz. Segment	PC	Project Age	Team Location	Project Methodol.	Deployment	Project Status
P1	Porto Alegre, RS	4.5	C	C	SAAS	Energy	15	3	B	A, C	Manual	Live
P2	Porto Alegre, RS	6	B	B	SAAS	Ads	7	4	A	A, C	Manual	Descontinued
P3	Porto Alegre, RS	4	A	A	Marketplace	Real state	4	3.5	A	C	CD	Descontinued
P4	Porto Alegre, RS	7	E	C	SAAS	Agro	7	0.5	B	A, C	CD	Development
P5	Porto Alegre, RS	2	A	A	Marketplace	Health	3	2.5	A	C	Manual	Live
P6	Porto Alegre, RS	6	C	B	SAAS	HR	10	6	B	C	CD	Live
P7	Porto Alegre, RS	3	B	B	Marketplace	Health	10	3	B	A	Manual	Live
P8	São Paulo, SP	5	B	C	SAAS	HR	7	5	B	A, C	Unknown	Live
P9	Piracicaba, SP	4	B	C	IAAS	Agro	10	0.5	B	A, C	CD	Live
P10	Belo Horizonte, MG	4.5	D	C	SAAS	Logistics	18	1.5	A	A, C	CD	Live
P11	Uberlandia, MG	11	C	C	SAAS	Industry	16	4	B	A, B, C	CD	Live
P12	São João da Boa Vista, SP	5	D	C	SAAS	Agro	14	5	B	A, C	Manual	Live
P13	Campinas, SP	2.5	C	C	SAAS	Agro	3	1.5	B	C	Manual	Live
P14	São Paulo, SP	5	C	B	SAAS	Real state	2	3	B	A	Unknown	Live
P15	São Paulo, SP	3	C	C	SAAS	HR	11	3	C	A	Manual	Live
P16	São Paulo, SP	4	E	C	SAAS	Agro	20	4	B	A	Manual	Live

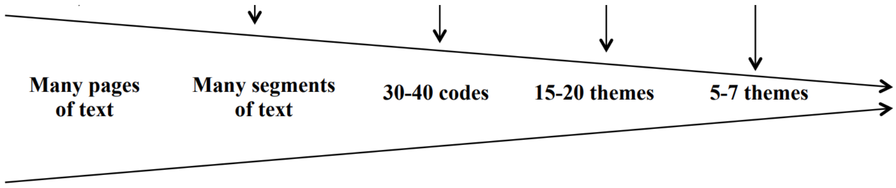

Figure 5.2 – Thematic Synthesis Process
Adapted from [9]

Table 5.2 displays such findings using the same IDs found in Table 5.1 as reference. To better draw this table, we reduced 'Employee Count' as 'EC', 'Customer Location' as 'CL', and 'Participant Count' as 'PC'. Appendix C displays the full script used during the interviews and the respective options offered to the participant at each question.

As described in Section 3.1.2, we started by interviewing companies near our university, which resulted in the first seven interviews occurring in the city of Porto Alegre. All participants are currently located or have started their operations in the TECNOPUC park. The following nine participants are located in the Brazilian southeast region, which coincidentally represents about 60% of the nationals entrepreneurial ecosystem nowadays [60].

5.4 Transcription Process

All interview recordings were manually transcribed, maintaining the speaker's native language (Portuguese), and proper translation into English was only applied when quoting the original transcript in this thesis.

The transcription process transformed the almost nine hours of recordings into 159 pages of text, representing an average of 10 pages per interview. During this process, any sensitive and personal information recorded was removed from the result text. The resulting text was later imported into the qualitative data analysis tool Atlas.ti[4] to start the coding activity described in the following sections.

5.5 Interview-Based Study Findings

This section details our findings after conducting the Thematic Analysis over the transcriptions. Each of the following subsections represents a higher-level theme in our compilation. Figure 5.2 was adapted from Cruzes and Dyba [9] and visually describes the steps taken while refining the transcriptions into codes, later summarized into themes.

[4]Available at https://atlasti.com

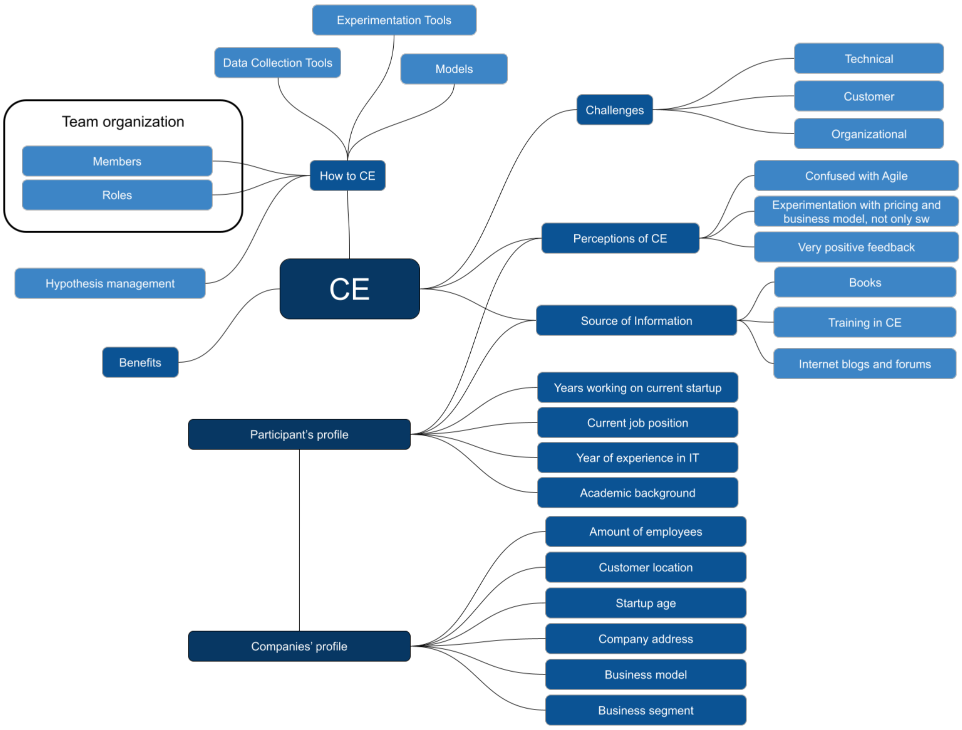

Figure 5.3 – Mind Map of Interviews' Codes

As recommended by Cruzes et al. [9], we generated a visual representation of codes we created that can be found in Appendix E. Figure 5.3 shows a reduced and more readable version of such a mind-map, displaying only higher-level codes. A mind-map representation assisted us in the process of sorting the multiple codes into themes.

Unfortunately, the online version of the qualitative data analysis tool used (i.e., Atlas.ti) does not auto-generate such visualization graph. The aforementioned mind-map was manually created using GoogleDocs (available at https://docs.google.com/) drawing tool.

5.5.1 Source of Information about CE

It is important to understand how and where the focus groups are getting information about our research topic. When participants were asked about how they got educated on the topic of CE, only three interviewees (P1, P3, P9) stated to have formal education on the topic. However, five participants (P1, P3, P4, P7, P9) stated to get informed over the topic through grey literature, such as internet blogs and forums.

One of the questions presented during the interviews listed five popular books about Lean Startup and Experimentation. Participants were asked to describe which they have read and share their perception about such literature. All participants, with no exception, declared that they knew or had read 'The Lean Startup' by Eric Ries [52]. Seven participants have provided positive feedback about such publication; P1 said that: "*I liked it a lot because I did not know anything about it, so it was my introduction in this universe*", and P10 added that this book is still relevant nowadays.

The second most popular book among the suggested list was 'The Four Steps to the Epiphany' by Steve Blank [6]. Seven interviewee [P1, P4, P10, P11, P14–P16] knew the book in which five have read it [P4, P10, P14–P16]. P10 considered it "*very dense at the beginning, so it's a good book without a doubt, but it's much more theoretical and impractical*". Participants were asked to recommend publications they considered to be good readings on our research topic, and these were the suggestions:

1. Scrum, by Jeff Sutherland [62]: P1, P12
2. Hooked, by Nir Eyal [13]: P9, P10
3. The one thing, by Gary W. Keller [25]: P1
4. Business Model Generation, by Alexander Osterwalder [47]: P2
5. Lean Software Development, by Mary Poppendieck [48]: P6
6. Principles, by Ray Dalio [10]: P15
7. When coffee and kale compete, by Alan Klement [29]: P4
8. The Mythical Man-Month, by Fred Brooks [7]: P11

5.5.2 How CE is Implemented

Many tools and practices are involved in the process of implementing CE. This section describes how the participants are approaching the research topic.

Data Collection Tools and Processes

Various techniques and platforms are available on the internet to gather users' data and behavior while using software systems. These are the data collection tools found in the interviews, sorted by frequency:

1. Google Analytics[5] (8): P1, P2, P3, P4, P6, P7, P9, P14
2. Developed own tool (4): P9, P10, P11, P12
3. Interviews with customers (3): P1, P13, P16

[5] Available at https://analytics.google.com

4. Facebook Pixel[6] (2): P3, P6
5. HotJar[7] (2): P1, P6
6. Participant unable to answer (2): P8, P15
7. MixPanel.com (2): P4, P6
8. Amplitude[8] (1): P3
9. Not collecting data (1): P5

Every interviewee currently uses at least one of the aforementioned options to track their product's usage, except P5. However, this participant said that his product "*is operating mostly in a 'manual' manner, while major fixes are being applied*" to his companies' software. When asked whether P5 intended to instrument his app with data collection tools, the participant confirmed that "*this is certainly planned as part of the next release*".

Two participants were not able to cite exactly which data collection tools their product was using. Instead, P8 stated that his team "*has a specialist dedicated to lead such data analysis tasks*". Similarly, P15 declared that his company "*has an entire area dedicated to collect usage and engagement metrics*".

Participants P1, P13, and P16 revealed that their softwares were targeted to have few but large clients. For that reason, they can afford to analyze the product's usage manually while conducting interviews with customers. P16 points that his *agrotech* startup is currently attending 15 large farm owners, so his team "*has started implementing it [analytics collection tools] into their product very recently*", yet this magnitude of users still let his team be "*very close to the customer and collect information directly in a more human and less automated manner* [through interviews]".

These results indicate that data collection tools intended to track user behavior in software solutions are well-known among practitioners. No participant was unaware of the topic, and 31% (5 out of 16) declared combining multiple tools while conducting related activities. Even participants who are not actively conducting 'experiments', such as P6 and P14, declared collecting data using such tools.

CE Models and Experimentation Tools

Although we identified 5 distinct models for implementing CE during the LR activities (see Section 4.8), no participant mentioned any of such models while conducting their projects. The results indicate that participants have developed their own procedures while

[6] Available at https://developers.facebook.com/docs/facebook-pixel
[7] Available at https://hotjar.com
[8] Available at https://amplitude.com

Table 5.3 – Popularity of Tools Supporting CE according to Interview-Based Study

Name	**Usage**	**Total**
MVP and MVF	[P1–P12, P14–P16]	15
Mockups and Wireframes	[P1–P10, P12, P14–P16]	14
A/B testing	[P2–P3, P7–P10, P15–P16]	8
Landing Pages	[P1–P2, P4–P8, P15]	8
Canary Release	[P3–P4, P16]	3
Dark launches	[P2–P3]	2
Fake Door Tests	[P10]	1
Multivariate Tests	[P16]	1

conducting CE instead of executing one found in the literature. Furthermore, 4 participants (P6, P7, P13, and P14) declared not to conduct ‘experiments’ routinely; thus, their contributions may not be considered in some of the following findings.

When asked whether ‘experiments’ were conducted sequentially or not, the participants (7 out of 12) declared to execute them simultaneously. P1 declared that their team only conducted ‘experiments’ sequentially “*because the team was not big enough to implement multiple ‘experiments’ at once*”. P10 added that their team runs ‘experiments’ “*simultaneously but not more than 2 or 3 hypotheses at once*”.

In that matter, participant P2 revealed that his team conducted multiple ‘experiments’ at once. Still, his team “*did not isolate them and, after several deploys during a single day, who knows what is being measured*”, which clearly points to untrustworthy results.

In contrast, P15 assumed that his team “*started by executing [the ‘experiments’] simultaneously and in a less organized way*”. Still, over time his team decided to start running ‘experiments’ in an “*organized and isolated way to ensure that we can say that the result is a consequence of something specific that we were evaluating*”. Participant P10 offers a similar testimony asserting that some results were only achieved due to other ‘experiments’ being conducted simultaneously.

The literature review activity also raised a comprehensive set of tools for conducting experimentation (See Table 2.1). We analyzed its adoption by practitioners and compiled it into the Table 5.3. The significant popularity of the book The Lean Startup [52], highlighted in Section 5.5.1, is accompanied by the notable adoption of the MVP and MFV approaches by the interviewees as well. These concepts were highly explored in such publication and thus becoming popular among entrepreneurs.

Developing products using the MVP/MVF approach was adopted by every interviewee during the existence of their companies at least once, except by P13. P12 stated that his team uses it frequently enough that they started naming their process as “*‘collaborative development’, because the customer works together developing the final product*”.

When participants were asked about how mockups and wireframes fit their workflow, we identified that 35% of the teams (5 out of 14) used it solely during the development process as a communication tool among developers and designers. Other participants have used such tools for prototyping and validating hypotheses with potential customers before implementing it. P3 stated to have created wireframes of the product and tested it face to face with users before developing it.

Half of the participants (4 out of 8) declared using A/B tests only with ad platforms (e.g., Google Adwords) when comparing multiple marketing campaigns. In this scenario, the experimentation happens before the user has landed on the startup's software; the test occurs with ad campaigns variables to optimize marketing investments.

Results show that participants had adopted multiple experimentation tools into their product development routines: 75% of the interviewee (12 out of 16) declared to have used 3 of the listed tools or more. The rigor and maturity of such processes are yet to be analyzed in the following section.

Hypothesis Management Process

When analyzing the participant's organization regarding hypothesis management, we identified that 11 teams have a distinct backlog exclusive to accommodate their hypotheses, separated from feature backlog. Three other participants declared to register their hypothesis alongside features in a single backlog.

Almost half of the teams with a hypothesis backlog (6 out of 11) declared that such backlog is managed by only one person in charge of it. Participants P4, P6, P9, P15, and P16 stated to assign such attribution to 'product owners'; only P11 uses its CTO to lead such task. The second half of teams with hypothesis backlog (5 out of 11) share the management of it among multiple team members. P10 reveals that "*such task is collaborative, every team member can bring new ideas*". P15 reported that his team "*has a Slack channel called 'brainstorm', and every team member can contribute with a ideas, coming it from sales, marketing, customer support [...] and later, the product management team prioritizes and compiles such ideas into testable hypothesis*".

We identified 3 sources that participants reported their hypotheses originated from. Ten participants stated that their hypotheses were originated from direct contact with customers. P16 suggests that "*direct interaction of salesperson and customer success staff have brought excellent hypothesis to be tested*".

Participants P1, P4, and P10 reported studying the solution of their business competitors (as a benchmark) during brainstorm sessions to formulate hypotheses. P5 suggested that "*due to years of experience in the market*" most of their hypothesis came from personal experience; P16 equally reproduced this concept.

We identified no clear consensus regarding how gathered data is analyzed once the experimentation is executed. Seven participants reported that such task was conducted by a single person in their team, and other 6 participants declared to execute such task with multiple people simultaneously.

Only 6 participants declared to define time duration constraints before starting their 'experiments': P5, P4, and P16 reported to limit 'experiments' to one week; P12 and P15 limit it to 2 weeks; and P8 limits to 2 months. The other participants declared not defining such time restrictions.

Regarding the confidence on the execution of the experimentation and trustworthiness of the results, almost every participant reported very positively to their previous experiences. However, only P9 and P10 declared to have previously executed 'experiments' which had negative results (i.e., hypothesis was tested and proved itself wrong). This lack of negative results to most participants may suggest that either the experimentation was poorly conducted or the hypotheses were not falsifiable.

Team Organization in CE

Considering team members' location, our dataset comprises 11 teams working remotely with all members within national territory (i.e., Brazil); 4 companies are working in an office, and 1 company has team members working remotely from more than one country.

Almost half of the teams working geographically distributed (5 out of 12) have started this working modality due to sanitary restrictions caused by the current global pandemic. Participants have reported positive feedback over the experience of working without an office and are considering adopting this modal indefinitely.

We found no evidence that team location impacted CE activities. P12 declared that "*new hires are not restricted to proximity to our office, since we are adopting this 'working from home' culture*". P15 stated that they "*realized that it is totally 'ok' to work remotely*" and they noticed no impacts to productivity so far.

Software distribution and effects on CE

Automation of software deployment (i.e., CD) is widespread among CE publications as a practice that facilitates CEs implementation [21, 22]. This practice is not considered a mandatory step for implementing CE, but it is highly recommended in the literature.

Participants were asked whether their projects are deployed manually or through an automated process. Exactly half of the participants (8 out of 16) declared to use manual deployment, followed by 6 participants who have implemented CD. The remaining 2 participants were unable to answer such questions.

The findings of the interviews suggest that even participants who have not adopted CD are still capable of conducting experimentation. However, the participants who adopted

CD are among those conducting more types of experiments (e.g., MVP, A/B testing). This finding may reinforce the understanding among the publications identified in the LR.

5.5.3 Benefits of Implementing CE

The list of benefits found in the LR is considerably smaller than the list of challenges (see Section 4.8.2), and the same result was found in the interviews. Rasmus et al. [55] suggest that there is a lack of publications regarding such topics. Similar to the list of challenges, these items served as the base of our coding process, which followed the integrated approach [9].

- Reduced development effort

Multiple participants reported that CE could lead to a reduction in the development of unnecessary features, which reduces wasting scarce resources. P1 and P12 provided feedback similar to P3, which reported that CE leads to "*'informed decisions' which makes the team saves time and energy while delivering the product faster*".

According to P2, once CE is regularly adopted it becomes a virtuous circle of improvements to the product and "*as we became specialized in experimenting, every time we did a bigger and more correct 'experiment', it becomes a continuous build on your products. It gets better and better, then the user visits it more and stays longer. When you make a mistake in the 'experiment' is to implement a new feature, and when it becomes part of your product, you will continue to experiment on top of that*".

P4 characterized the benefit as "*good for risk management*". According to P4, "*CE turns the decision of 'what to develop' more accurate, which helps to reduce the risk of developing an unwanted product*". Participants P15 and P16 provided very similar contributions.

- Data-driven decisions

Participant P3 highlights that by adopting CE, their team "*can make informed decisions, leading to faster product delivery*". P8 describes that "*data from 'experiments' usually are brought to meetings to base relevant decisions that are being discussed*". Participants P10 and P13 offered similar feedback.

According to P12, basing actions on collected data from experimentation was very impactful. P12 added that: "*no doubt this has become a routine. For us, it started without well-defined process, and then the change was extremely positive. After we started using these techniques, the result was fast. We started to have a much clearer vision of the processes, and we started to apply this policy: what we don't know, we can't improve*".

- Increased customer engagement

Among the benefits, P9 says that CE “*increases customer engagement to the platform and reduces Churn rate*[9]”. Similarly, P5 holds a similar experience to what P10, which reports that "*the main benefit is that you make a product that your customer wants, so he’s making the product indirectly too. We have this logic to make things for the client: who has to like it is the customer, they will use it*".

- Improved communication with investors

P5 highlights that experimentation assisted his team to understand the perspective of their investors, as follows: "*we started to understand what the investors’ view was: what they needed to hear to understand our business to want to invest, and we are focusing on that now. MVP helped us to understand better the metrics we should focus on*".

- Decentralization of business knowledge and facilitate knowledge transfer

Participant P4 suggests that CE assists in involving engineers in the problem, which leads to higher empathy to the customer’s issues. P4 shared that: "*I found it interesting that it doesn’t end up being just between designers and product managers, it ends up reaching the whole team. So to analyze the data together with engineers, for example, brings them closer to the problem they are solving. And then I think this induces more empathy, they’re not just the guys who deliver features*".

5.5.4 Challenges of Implementing CE

We identified multiple common challenges faced by teams while implementing CE on the LR. These findings are thoroughly described in Section 4.8; they served as the base of our coding process, which followed the integrated approach [9].

Such challenges were arranged into 3 groups (organizational, technical, and customer), following the organization originally proposed by Rissanen et al. [53]. Section 6.3 identifies and compares which items were similarly found on the literature review and which ones were exclusively found on the interview-based study.

[9] In this context, Churn is the measurement of customers abandoning the service

Organizational Challenges

- Low maturity of processes and low education in experimentation

Some items in the questionnaire were targeted to guide participants to describe their procedure while conducting experimentation. We identified a significant lack of maturity in these processes while analyzing some transcriptions, especially P1, P2, P8, and P16.

In some cases, this chaotic scene was even assumed by the participants themselves. In the following excerpt, P2 states that "*by the lack of self-organization, I believe that we make more multivariate tests than the A/B tests. We did not isolate variables, and when you have several deploys on a single day... who knows what you're measuring. We were not certain about the variables [being tested]. It was not a very scientific process*".

Similarly, when it was asked to P8 describe the processes his team used to guide their CE, the participant depicted as "*without technique, based on guesswork*". Participants P5 and P13 declared to have started working with technology-based projects very recently, and their academic background did not cover such a subject. Their businesses depend on technology to grow sustainably; however, this area is mostly outsourced to external partners.

Participant P5 described needing a "*dictionary*" still to understand most of the terms regarding experimentation. When the interview was about to finish, P13 said to "*feel motivated to start studying further*" the research topic, but he was unaware of all the mentioned Continuous Experimentation tools (described in Table 2.1).

- Prioritization of customers requests over experimentation

The most frequent issue reported by interviewees regards organization leadership prioritizing the development of customer requests instead of implementing and executing 'experiments'. This topic was similarly mentioned by 5 participants (P1, P3, P6, P7, and P9). According to P1, "*the biggest challenge is to deallocate the team focused on feature delivery to try to implement a hypothesis that needs to be validated. [...] There is a lack of team members to deliver what is in the backlog and implement tests for the hypotheses*".

On the report of P9, to successfully implement CE, his team should "*be generating numerous hypotheses easily and in a short period of time. Instead of generating one hypothesis, we should be generating 20, 30 and then developing and testing it [...]. Still, we have to define priorities, and we can not develop all the features and 'experiments'. We have to prioritize these features [customer requests] because we do not have time*".

- Creating a innovative culture within the company

Cultural barriers may negatively impact the adoption of CE in various ways. Some examples mentioned by the participants are misaligned expectations (P1, P3, P12), business segment mentality (P3, P12), and acceptance of failure (P15).

Cultural background may contribute to create diverging expectations while creating innovative products. P12 reported that their colleagues, which had low experience in software development, frequently proposed hypothesis and expected test them immediately but "*people sometimes does not comprehend that technology cannot be done overnight. It's not as simple as it sounds*".

Furthermore, participants P3 and P12 declared that experienced leaders with a background in traditional business segments (real state and agriculture, respectively) offered resistance while embracing a workflow based on experimentation. As P3 declared, "*we faced some challenges, especially with our business manager, who was much older than us and always worked in a traditional company, that must strictly obey contracts and follow plans. He had this mindset that 'when you're going to shoot, it must be a bullseye', which is hard to replicate when innovating in a startup*".

Participant P15 reported that his team needed to develop a 'test mentality' to implement CE better. The participant describes this as an antagonistic situation because they "*want to be super fast and agile, but at the same time it needs to test the product a thousand times*", and even after that, the results may reveal frustrating.

Technical Challenges

- Trustworthiness and statistical significance of data

Participants reported that their team did not conduct as many 'experiments' as expected due to the low number of users on their platforms. P1 added that their team even designed A/B tests but never ran it on a production environment "*afraid of the low statistical volume*"; P2, P3, and P4 provided a similar testimonial.

Participant P4 highlighted that the infrastructure required to execute some 'experiments' might be a challenge, but the low statistical relevance was an impeditive criterion for running such A/B tests.

Customer Challenges

- Complex stakeholder structures

Rissanen and Münch [53] proposed that CE faces particular challenges when implemented in a Business-to-Business (B2B) domain. Participant P3 supported the author's proposal by reporting that running experimentation may become especially challenging because positive impacts perceived by one type of customer may generate negative ones to the other end. P10 contributed with similar reports.

P4 reported that their customers are geographically distant and thus, "*'in loco' experiments become frequently impracticable*". Furthermore, the participant reports that additional attention must be addressed when trying to increase empathy with the customer's routine (rural activities), being it much different from their urbane routine in the office.

In the opposite scenario, participant P10 states that too frequent interaction with the customer may highlight misleading hypothesis. According to the interviewee: "*the customer has direct contact with us. I receive quality complaints directly. And you need to know how to deal with it because there are a lot of people with negligible complaints [...] which may distract us from real issues*".

- Interruption of service during deployment of experiments

Participant P15 revealed that their team have fully adopted the "*mindset of experimentation*", and due to a lack of rigorous Quality Assurance processes, they needed to slowdown the deployment of new 'experiments'.

According to the participant: "*when we go out doing things very quickly and want to deploy it fast, it becomes our Achilles' heel. [...] We are really working to ensure the quality of our deliveries, trying to foresee corner cases that might happen along the way*".

6. ANSWERING THE RESEARCH QUESTIONS

This chapter presents answers to the RQs introduced in Section 1.3. Such answers were achieved by combining and contrasting the findings of the literature review and the interview-based study. Each of the following sections formulates answers by enumerating findings sorted by the frequency in which they were found.

Figure 6.1 exhibits a summary of the answers found in the result analysis in a mind map representation. The orange boxes represent the answer to RQ1, while the green box represents answers to RQ2 and the blue boxes represent answers to RQ3. Appendix F summarizes our findings and identifies which ones were found on the LR and which emerged from the interviews. Additionally, Figures 6.2, 6.3, and 6.4 present zoomed-in versions of Figure 6.1, focusing on each respective RQ's answers for improved readability.

6.1 CE Adoption in Technology-Based Startups (RQ1)

This section seeks to answer the previously mentioned RQ1:

RQ1 How do technology-based startups implement Continuous Experimentation?

Figure 6.2 presents the items on the 'Data Collection Tools' column as found only during the interviews. Such a subject was not included during the LR activity, resulting in no appearances in the literature.

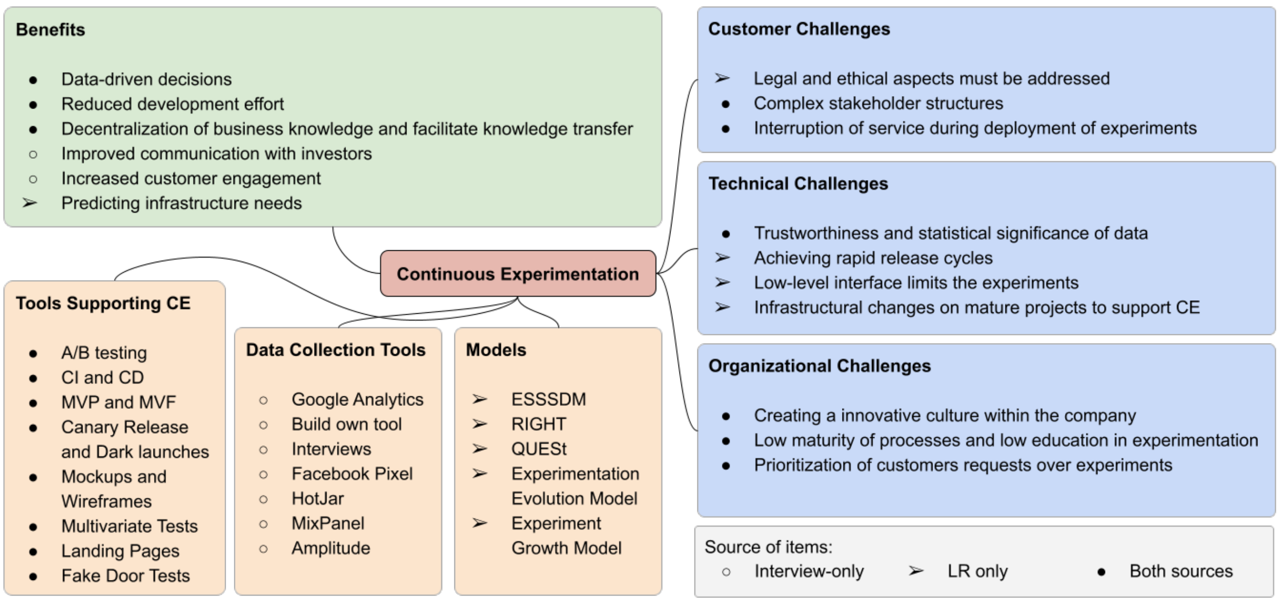

Figure 6.1 – Mind Map of Aspects of CE

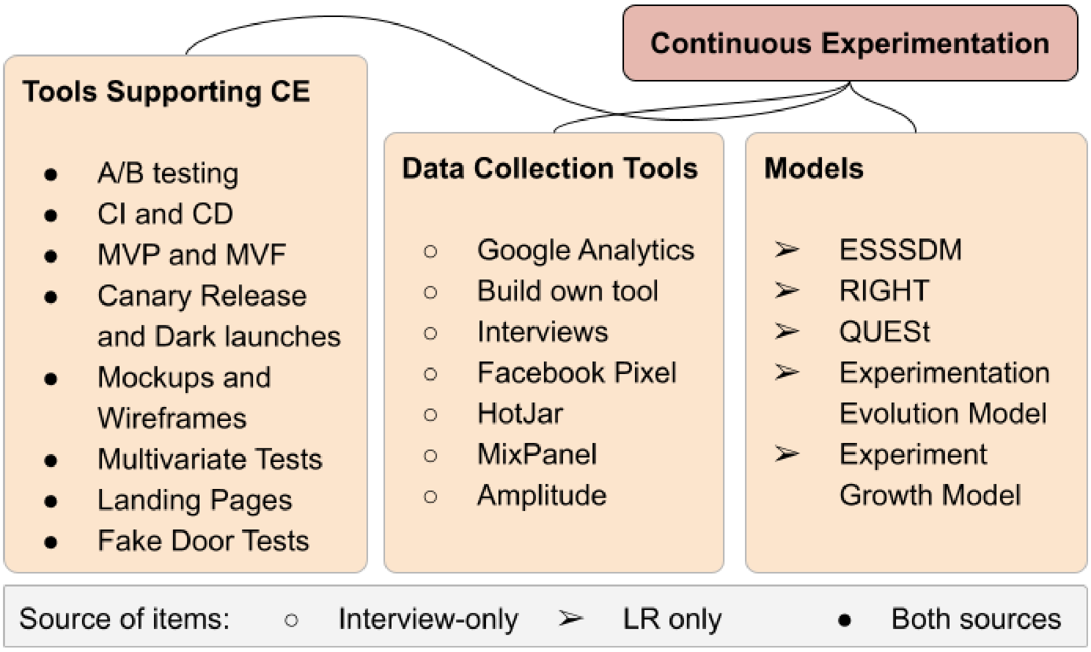

Figure 6.2 – Mind Map of Aspects of CE - Zoomed in RQ1

The LR activity highlighted various tools used to support the execution of experimentation while implementing CE (see Table 4.4). The literature's most popular technique is 'A/B Testing' however, this item figured in third among interviewees, being adopted by 8 participants (50%). We identified that the subject is well-known among the participants, but technical challenges prevent vast adoption. According to P6, "*even though it is promising, preparing ourselves to run A/B tests takes time, and we can not afford that effort right now*".

The second most cited technology on the LR is 'CI & CD', which are not specifically meant to conduct experimentation. Still, literature endorses automatical deployment as a measurement to reduce human interference in the results [26]. The interviews showcased that 6 participants are already using a fully automated deployment process, 3 participants have some level of automation; and most of the remaining participants agreed with the importance of the subject and declared that CD is planned to be adopted soon. In this scenario, we can assume that at least one of the technical barriers (i.e., CD) is being addressed by the participants, which can later translate to more CE adoption.

We identified 'MVP & MVF' and 'Mockups & Wireframes' as the first and second most used tools in the interviews analysis, respectively. We interpreted this result considering that using the MVP approach to guide the product development does not aggregate extra technical challenges (especially when compared with A/B testing, for instance).

Similarly, 'Mockups & Wireframes' are already commonly used during development for communication among designers and developers. Eventually, such artifacts reach the customers as non-functional prototypes during the experimentation process. The 'Fake door test' figured among the least mentioned both in the interviews and in the LR, suggesting that such practice is neither commonly adopted in the academy nor by the industry.

The LR also highlighted 5 models of implementing CE. Still, the interviewees did not cite any of such models. We suggest that such detachment among the literature and industry can be justified because such models were proposed and tested among well-established startups, and the companies we interviewed are still maturing such practices.

Besides, no participant mentioned any academic publication in their testimonials. Regarding grey literature, 8 additional books were suggested as referential material to our research topic. This result indicates no clear consensus among participants about the literature on the topic, considering only the first 2 items [13, 62] have appeared more than once, except 'The Lean Startup', which was considerably popular among our sample.

It is also noticeable that the P11's suggestion seems to not correlate with the research topic. When combined with other inputs from this participant, it becomes clear that the interviewee confuses the definition of experimentation with project management.

Fagerholm et al. [20] highlight that the whole CE process should be well-drawn and coordinated when running multiple 'experiments' simultaneously only that statistical interactions among executions do not interfere in the trustworthiness of the results achieved. However, we identified that 7 out of 12 participants assumed to be executing multiple experiments at once. By combining such finding with the testimonials of low maturity of processes, we can assume that a significant portion of the experiments generate untrustworthy results.

6.2 Perceived Benefits While Using CE (RQ2)

This section seeks to answer the previously mentioned RQ2:

RQ2 What are the perceived benefits while adopting Continuous Experimentation for technology-based startups?

The initial four topics were found in the literature and summarized startups' perceived benefits after implementing CE. The last two topics were originated from the interview-based study, without precedent in the literature. All items are sorted according to their popularity in our LR results.

- Data-driven decisions

Observing the literature and the testimonial of participants [P3, P8, P10, P12, P13], we identified that this is one of the most significant perceived benefits. Participant P3 highlighted that by adopting CE their team could "*abandon decisions based on 'gut-feeling' and start to make informed decisions*". P8 suggests that their team started using gathered data

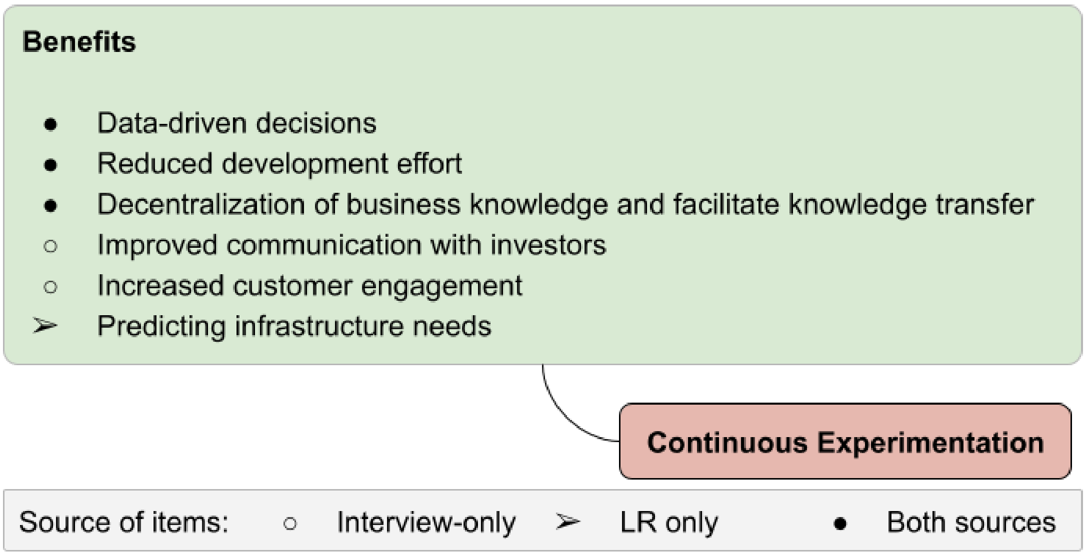

Figure 6.3 – Mind Map of Aspects of CE - Zoomed in RQ2

during board meetings, which facilitates and speeds up the business decision process. P12 provided a very impactful policy that guides his team: "*what we don't know [not measured], we can't improve*".

- Reduced development effort

One of the main reasons startups end their activities is 'Shortage of capital or lack of planning when using resources' [8]. Participants reported that by adopting CE, their startups could extend their existence by reducing the waste of resources invested in developing unnecessary solutions, thus validating more business hypotheses.

We understand that for that reason, this benefit was even more popular in the interviews than the previous one, being reported by 7 participants [P1–P4, P12, P15, P16]. P4 defines this benefit as "*good for risk management*", because it converts the decision of 'what to develop' more accurate. This topic also was considerably popular on the LR findings, being supported by multiple authors (e.g., [53, 64, 71].

- Decentralization of business knowledge and facilitate knowledge transfer

Only participant P4 reported this benefit. We understand that such a result could be foreseen, considering that most of the startups which participated in the interviews has a small number of team members, ranging from 3 to 20 people. We assume that concerns regarding knowledge transfer would possibly be more popular among bigger-sized teams.

- Predicting infrastructure needs

This benefit was not reported by any participant in our interview-based study. Once again, we understand that such concern was found in the literature among considerably larger companies' sizes, which does not represent our interviewed startups.

- Increased customer engagement

This finding was not identified on the LR but was reported by P5, P9, and P10. These participants mentioned that CE adoption has positively impacted business metrics such as NPS[1] and reduced Churn rate. Such business metrics are beyond the scope of this research, and thus we will not deepen on this topic.

- Improved communication with investors

This item was only reported by P5 on the interviews, without precedent on the LR activity. According to this participant, startup companies that are funded by investors need to report business progress to such stakeholders frequently. The health and trustworthiness of such relationships are fundamental in order to receive future investments. We considered that such finding was only reported by P5 either because the other participant's companies are self-funded or their investors already exited the business.

6.3 Perceived Challenges While implementing CE (RQ3)

This section seeks to answer the previously mentioned RQ3:

RQ3 What are the challenges while implementing Continuous Experimentation for technology-based startups?

This section presents all the challenges of implementing CE found in the literature organized into three groups. These groups were originally proposed by Rissanen et al. [53] and later expanded with the participants' contributions during the interview-based study.

6.3.1 Technical Challenges

- Trustworthiness and statistical significance of data

This topic was found four times [P1-P4] on the interviews and was the most common technical challenge on the LR with 8 occurrences. Participants reported difficulty when trying to execute 'experiments' with newly launched services, which still do not have enough users to generate statistical relevance.

Participant P4 also added that implementing proper technical infrastructure to conduct A/B testing, for example, is too costly, and their team cannot afford it at the stage their

[1] Net Promoter Score is a business metric typically used to measure one customer's likelihood to recommend a given product or service.

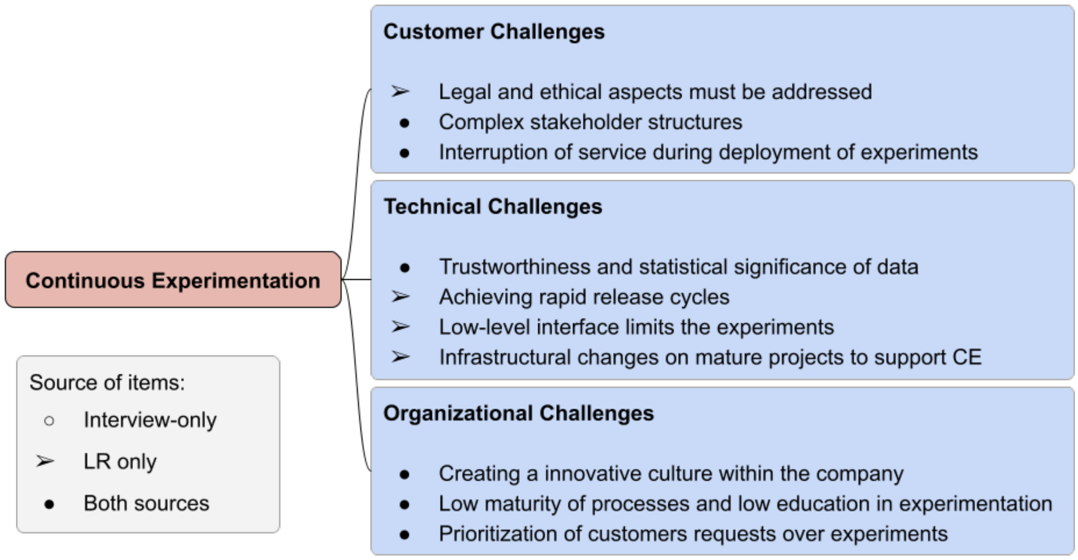

Figure 6.4 – Mind Map of Aspects of CE - Zoomed in RQ3

company is found at. That report also suggests that P4 favors developing features without experimentation instead of investing resources into preparing the proper infrastructure.

- Achieving rapid release cycles

Lindgren et al. [35] highlights that automated release cycles become very important when integrating CE to the product development workflow. Longer development cycles may lead the team to lose track of running ‘experiments’ or lose valuable competitive advantage [22]. However, this finding was not found in the interviews. We understand that such a scenario is associated with the low-maturity level faced by our group of participants.

- Low-level interface limits the experiments

This finding was only reported by Rissanen and Munch [53] on the LR when considering software products that offer a low-level interface to interact with. Due to the nature of the businesses that we interviewed, this issue was not mentioned by any participant.

- Infrastructural changes on mature projects to support CE

Such finding was also only reported by Rissanen and Munch [53] on the LR when arguing that companies with mature products face additional barriers for adopting CE due to the overhead imposed by customizing a software running with several customers.

Most of our interview participants do not have a technical background or are not directly responsible for engineering decisions. We understand that such a scenario directly translates to the fact that this topic was not mentioned during any interview.

6.3.2 Organizational Challenges

- Creating a innovative culture within the company

During the interviews, we identified multiple cultural reasons which can negatively impact CE adoption. Participants P3 and P12 contributed by stating that divergent expectations with older team members generated friction when their teams started conducting experimentation. According to these participants, team members with no previous experience in innovative environments may find it difficult to accept that the hypothesis may prove itself wrong and be later removed because it did not add value to the final solution.

P15 mentioned that their team cultivated a "*test mentality*" to embrace better the situation that can emerge from a fast-paced development and experimenting environment, which may reveal itself frustrating when the solution faces unexpected behavior.

- Low maturity of processes and low education in experimentation

We identified multiple mentions of poorly designed or unstructured conducting of ‘experiments’ when analyzing the transcriptions. Participants P5 and P13 attributed the low maturity of their processes due to the lack of proper education on the subject and no previous experience with technology-based services.

Participants P1, P2, P8, and P16 reported several misconducts, such as running multiple ‘experiments’ simultaneously without concerning about isolating each result; or conducting ‘experiments’ until positive results are achieved instead of executing them until predefined conditions are met, leading to false-positive results.

- Prioritization of customers requests over experiments

This item was only mentioned once [53] in the LR but found 5 times in the interview-based study [P1, P3, P6, P7, P9]. These participants reported that leadership constantly interferes in the execution of ‘experiments’ by prioritizing the development of customer requests instead. P1 stated that their biggest challenge was to get the team fully focused on validating the hypothesis instead of delivering feature requests.

6.3.3 Customer Challenges

- Legal and ethical aspects must be addressed

The LR highlighted that teams conducting CE must address the legal and ethical aspects involved [16]. However, no participants have raised concerns about legal and ethical aspects during the interviews.

Several countries have expressed concern about users' privacy while using technology services, and laws are being enforced to protect these individuals. For example, Brazilian companies must follow the LGPD[2] under severe financial penalty when unobserved. This finding suggests that our participants may be running 'experiments' unaware of such associated legal aspects.

- Complex stakeholder structures

Participant P3 revealed that their marketplace service offers additional challenges while testing the hypotheses because positive impacts perceived by one type of customer may generate negative ones to the other end. Participant P10 offered similar feedback. P4 added that their office is geographically distant from their customers, which hinders the execution of "*in loco* experiments".

- Interruption of service during deployment of experiments

This challenge was only mentioned by participant P15, who revealed previous experiences in which experimentation degraded customers' experience. The participant stated that such an episode taught their team to improve the quality assurance stage before deploying poorly tested changes.

[2] *'Lei Geral de Proteção de Dados Pessoais'* is the Brazilian legislation similar to the General Data Protection Regulation (GDPR) found in Europe.

7. FINAL CONSIDERATIONS

This chapter presents the final considerations of this study. Section 7.2 presents well-known limitations that threaten the validity of the conclusions achieved. Section 7.1 highlights a summary of the findings presented in Chapter 6. Finally, Section 7.3 provides insights regarding the continuation of this research.

7.1 Conclusions

As Yaman et al. [71] point out, the highly dynamic scenario faced by technology organizations imposes the risk that the product developed may add little or no value to customers, wasting time and resources in the process. CE emerges as an experiment-driven development approach that intends to mitigate such risks by iteratively testing assumptions critical to an initiative's success.

This approach has become very popular after the lean movement won the attention of many entrepreneurs and authors on the subject. The increasing number of publications on the subject supports this observation, as seen in Chapter 4. Despite the topic's popularity, alarming numbers of startup bankruptcy are still seen nowadays [2].

In this scenario, this research aimed to characterize how technology-based startups currently adopt CE. The results presented here were achieved through the combination of a literature review activity and a semi-structured interview-based study.

We analyzed 33 papers during the literature review and conducted 16 interviews with members of the entrepreneurial ecosystem. Such results were later compiled in order to formulate answers to our Research Questions (as seen in Section 1.3), which covered in short: how our focal group currently adopts CE; what are the benefits perceived; and what are the challenges that need to be addressed in order to conduct CE.

The findings presented in Chapter 6 suggest that this research topic is significantly relevant on both academic and entrepreneurial grounds. Appendix F summarizes our findings by visually comparing occurrences originated on the LR and on the interviews.

We described six benefits experienced by technology-based startups when conducting CE. The 'Data-driven decisions' and 'Reduced development effort' were commonly evidenced in both the literature and by interviewees. Such finding suggests that practitioners consider CE as a valid alternative to cope with one of the main reasons that lead to startup bankruptcy, 'waste of resources' [8].

The results of the LR showed that the most popular challenge among such practitioners is 'Trustworthiness and statistical significance of data'. However, 'Low competence and education in experimentation' was the most popular challenge in our focal group in the interviews. This outcome highlights a significant difference between the CE's maturity level among startups at a national level (interview's participants) compared to foreign companies (LR participants). Many findings were presented in Section 6 that evidenced 'experiments' being poorly conducted and possibly leading to untrustworthy results.

We understand that trustworthiness of results is a fundamental challenge that still needs to be addressed by our focal group. Only once such milestone is surpassed will participants be able to concern about more complex challenges.

We expect that future practitioners can use the results presented in this thesis in order to plan and act ahead to overcome the listed challenges and enhance the possible benefits of adopting CE. Also, we expect that the results achieved here can contribute to future academic studies regarding CE and its impacts on the SE processes, considering such research area still needs to be deepened.

7.2 Limitations

Possible measurements were planned ahead and taken during the execution of this research to minimize the impacts of well-known risks.

For example, during the literature review activity, we followed predefined processes (i.e., Snowballing approach [65]), which already had previous positive results in our research group. We also peer-reviewed the results with research colleagues, and the whole process is documented in this publication, which facilitates later reproduction to compare results achieved. Although, we acknowledge that some steps of such activity are passive to a certain level of interpretation of the agent.

We focused on reaching a diverse population of technology-based startup members during the interviews following our predefined focal group's profile. Table 5.1 testify that the participants have multiple academic backgrounds, years of experience in the industry, current roles, and geographic locations. Characteristics regarding gender, ethnicity, age, and other aspects were not considered during the participants' invitation and may pose as threats to validity, considering different backgrounds may influence the answers provided.

The approach chosen for the interview-based study (i.e., semi-structured interviews) is directly influenced by the interviewees' contributions. Participants' misunderstanding concepts and personal believes are likely to affect our conclusions. To mitigate such risk, we explicitly reported the number of times each result was obtained during the interviews so that it is possible to assess the validity of such contributions.

Only the main researcher executed the coding process, which can lead to biased results and influence the conclusions drawn. To reduce such threat, once again, the outcomes were presented to research colleagues to share our conclusions and reassure the correctness of the process. Also, we consulted publications of former members of our research group that used the same procedures in order to mitigate flaws in our execution.

7.3 Future Work

By evaluating the findings of this research, we identified opportunities for the continuity of this research, as follows:

1. Re-conduct the interviews

The results presented in this study demonstrate a snapshot of the participants' perspectives on the research topic. Multiple participants noticed the low level of maturity their projects faced while conducting 'experiments' and reported interest in learning more and refining their processes by the end of the interview.

We propose to repeat the interviews, especially with the same participants listened during this study so that we will be able to compare current and future answers and elaborate on the evolution of the adoption of CE among technology-based startups.

2. Apply literature's CE models in the industry

The interviews revealed that no participant was familiarized with the CE models present in the literature (present in Section 5.5.2). We propose a field study in which such CE models are thoroughly presented to participants and applied in real projects in order to be further analyzed according to their impacts on the most challenges identified in the present publication. We can take advantage of the proximity our research group has with the startups present at the TECNOPUC Park to obtain volunteers to execute such a study.

3. Propose a new CE model focused on reported challenges

The Brazilian entrepreneurial ecosystem may impose additional challenges (e.g., labor market and tax regulation, juridical insecurity) while running a startup company. Assuming the CE models found in the LR may not consider such context faced by our focal group, we suggest a research project to develop a new CE model that addresses the challenges reported explicitly by our participants.

REFERENCES

[1] Alexander, I. “Engineering as a Co-operative Inquiry: A Framework”, *Requirements Engineering*, vol. 3–2, 6 1998, pp. 130–137.

[2] ARRUDA, C.; Nogueira, V.; Cozzi, A.; Costa, V. “Causas da Mortalidade de Startups Brasileiras”, *Núcleo de Inovação e Empreendedorismo, Fundação Dom Cabral*, Nov 2014, pp. 33.

[3] Auer, F.; Felderer, M. “Current State of Research on Continuous Experimentation: A Systematic Mapping Study”. In: Proceedings of the Euromicro Conference on Software Engineering and Advanced Applications, Prague, Czech Republic: IEEE, 2018, pp. 335–344.

[4] Bank, W. “Doing Business 2020”. WebPage, Source: https://www.doingbusiness.org/en/reports/global-reports/doing-business-2020, Jun 2020.

[5] Björk, J.; Ljungblad, J.; Bosch, J. “Lean Product Development in Early Stage Startups”. In: Proceedings of the From Start-ups to SaaS Conglomerate - Life Cycles of Software Products Workshop, vol. 1095, Potsdam, Germany: CEUR-WS, 2013, pp. 19–32.

[6] Blank, S. “The Four Steps to the Epiphany: Successful Strategies for Products That Win”. New Jersey, USA: John Wiley & Sons, 2020, 384p.

[7] Brooks, F. P. “The Mythical Man-month”. Boston, USA: Addison-Wesley Professional, 1975, vol. 20, 320p.

[8] CBInsights. “Top 20 Reasons Startup Fail”. WebPage, Source: https://bit.ly/folha__startup_failure, Jun 2020.

[9] Cruzes, D. S.; Dyba, T. “Recommended Steps for Thematic Synthesis in Software Engineering”. In: Proceedings of the International Symposium on Empirical Software Engineering and Measurement, Banff, Canada: IEEE, 2011, pp. 275–284.

[10] Dalio, R. “Principles: Life and Work”. New York, USA: Simon and Schuster, 2017, 592p.

[11] Dingsøyr, T.; Lassenius, C. “Emerging Themes in Agile Software Development: Introduction to the Special Section on Continuous Value Delivery”, *Information and Software Technology*, vol. 77, 9 2016, pp. 56 – 60.

[12] Dmitriev, P.; Gupta, S.; Kim, D. W.; Vaz, G. “A Dirty Dozen: Twelve Common Metric Interpretation Pitfalls in Online Controlled Experiments”. In: Proceedings of the International Conference on Knowledge Discovery and Data Mining, Halifax, Canada: ACM, 2017, pp. 1427–1436.

[13] Eyal, N. “Hooked: How to Build Habit-forming Products”. New York, USA: Penguin, 2014, 256p.

[14] Fabijan, A.; Dmitriev, P.; McFarland, C.; Vermeer, L.; Holmström Olsson, H.; Bosch, J. “Experimentation Growth: Evolving Trustworthy A/B Testing Capabilities in Online Software Companies”, *Journal of Software: Evolution and Process*, vol. 30–12, Nov 2018, pp. e2113.

[15] Fabijan, A.; Dmitriev, P.; Olsson, H. H.; Bosch, J. “The Benefits of Controlled Experimentation at Scale”. In: Proceedings of the Conference on Software Engineering and Advanced Applications, IEEE, Vienna, Austria: IEEE, 2017, pp. 18–26.

[16] Fabijan, A.; Dmitriev, P.; Olsson, H. H.; Bosch, J. “The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-driven Organization at Scale”. In: Proceedings of the International Conference on Software Engineering, Buenos Aires, Argentina: IEEE, 2017, pp. 770–780.

[17] Fabijan, A.; Dmitriev, P.; Olsson, H. H.; Bosch, J. “The Online Controlled Experiment Lifecycle”, *IEEE Software*, vol. PP, Out 2018, pp. 1–1.

[18] Fabijan, A.; Dmitriev, P.; Olsson, H. H.; Bosch, J. “Online Controlled Experimentation at Scale: An Empirical Survey on the Current State of A/B Testing”. In: Proceedings of the Euromicro Conference on Software Engineering and Advanced Applications, Prague, Czech Republic: IEEE, 2018, pp. 68–72.

[19] Fabijan, A.; Dmitriev, P.; Olsson, H. H.; Bosch, J.; Vermeer, L.; Lewis, D. “Three Key Checklists and Remedies for Trustworthy Analysis of Online Controlled Experiments at Scale”. In: Proceedings of the International Conference on Software Engineering: Software Engineering in Practice, Montreal, Canada: IEEE, 2019, pp. 1–10.

[20] Fagerholm, F.; Guinea, A. S.; Maenpaa, H.; Munch, J. “Building Blocks for Continuous Experimentation”. In: Proceedings of the International Workshop on Rapid Continuous Software Engineering, New York, USA: ACM, 2014, pp. 26–35.

[21] Fagerholm, F.; Sanchez Guinea, A.; Mäenpää, H.; Münch, J. “The RIGHT Model for Continuous Experimentation”, *Journal of Systems and Software*, vol. 123, 1 2017, pp. 292 – 305.

[22] Fitzgerald, B.; Stol, K.-J. “Continuous Software Engineering: A Roadmap and Agenda”, *Journal of Systems and Software*, vol. 123, 1 2017, pp. 176 – 189.

[23] Gupta, S.; Ulanova, L.; Bhardwaj, S.; Dmitriev, P.; Raff, P.; Fabijan, A. “The Anatomy of a Large-Scale Experimentation Platform”. In: Proceedings of the International Conference on Software Architecture, Seattle, USA: IEEE, 2018, pp. 1–109.

[24] Jardim, M. “On The Understanding Of The Role Of Experimentation In technology-based Startups”, Technical Report, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil, 2021, 42p.

[25] Keller, G. W. “The One Thing”. Portland, USA: Bard Press, 2013, 240p.

[26] Kevic, K.; Murphy, B.; Williams, L.; Beckmann, J. “Characterizing Experimentation in Continuous Deployment: A Case Study on Bing”. In: Proceedings of the International Conference on Software Engineering: Software Engineering in Practice Track, Buenos Aires, Argentina: IEEE, 2017, pp. 123–132.

[27] Khanna, D.; Nguyen-Duc, A.; Wang, X.; Khanna, D.; Nguyen-Duc, A.; Wang, X. “From MVPs to Pivots: A Hypothesis-Driven Journey of Two Software Startups”. In: Proceedings of the International Conference of Software Business, Xi’an, China: Springer, 2018, pp. 172–186.

[28] Kitchenham, B.; Charters, S. “Guidelines for Performing Systematic Literature Reviews in Software Engineering”, *EBSE Technical Report*, vol. 2–EBSE 2007-001, 7 2007, pp. 65.

[29] Klement, A. “When Coffee and Kale Compete”. Scotts Valley, USA: CreateSpace Independent Publishing Platform, 2016, 228p.

[30] Klepper, S.; Bruegge, B. “Impact of Hypothesis-Driven Development on Effectiveness, Quality, and Efficiency in Innovation Projects”. In: Proceedings of the Software Engineering and Software Management, Bonn, Germany: German Informatics Society, 2018, pp. 181–188.

[31] Kohavi, R.; Deng, A.; Frasca, B.; Walker, T.; Xu, Y.; Pohlmann, N. “Online Controlled Experiments at Large Scale”. In: Proceedings of the International Conference on Knowledge Discovery and Data Mining, New York, USA: ACM, 2013, pp. 1168–1176.

[32] Kohavi, R.; Deng, A.; Longbotham, R.; Xu, Y. “Seven Rules of Thumb for Web Site Experimenters”. In: Proceedings of the International Conference on Knowledge Discovery and Data Mining, New York, USA: ACM, 2014, pp. 1857–1866.

[33] Kohavi, R.; Longbotham, R. “Online Controlled Experiments and A/B Testing”, *Encyclopedia of Machine Learning and Data Mining*, vol. 7–8, 1 2017, pp. 922–929.

[34] Kujala, S.; Kauppinen, M.; Lehtola, L.; Kojo, T. “The Role of User Involvement in Requirements Quality and Project Success”. In: Proceedings of the International Conference on Requirements Engineering, Paris, France: IEEE, 2005, pp. 75–84.

[35] Lindgren, E.; Munch, J. “Raising the Odds of Success: The Current State of Experimentation in Product Development”, *Information and Software Technology*, vol. 77, 9 2016, pp. 80–91.

[36] Loureiro, R. “Brasil Tem Mais de 8 Mil Investidores em Startups”. WebPage, Source: https://exame.com/especiais/brasil-ja-tem-mais-de-8-mil-investidores-em-startups/, Jan 2021.

[37] Mattos, D. I.; Bosch, J.; Olsson, H. H. “Challenges and Strategies for Undertaking Continuous Experimentation to Embedded Systems: Industry and Research Perspectives”, *Agile Processes in Software Engineering and Extreme Programming*, vol. 314, 5 2018, pp. 277–292.

[38] Mattos, D. I.; Dakkak, A.; Bosch, J.; Olsson, H. H. “Experimentation for Business-to-Business Mission-Critical Systems: A Case Study”. In: Proceedings of the International Conference on Software and System Processes, New York, USA: ACM, 2020, pp. 95–104.

[39] Melegati, J. “Improving Requirements Engineering Practices to Support Experimentation in Software Startups”. In: Proceedings of the Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, New York, USA: ACM, 2019, pp. 1172–1175.

[40] Melegati, J.; Wang, X. “QUESt: New Practices to Represent Hypotheses in Experiment-driven Software Development”. In: Proceedings of the International Workshop on Software-Intensive Business: Start-ups, Platforms, and Ecosystems, Tallinn, Estonia: ACM, 2019, pp. 13–18.

[41] Melegati, J.; Wang, X.; Abrahamsson, P. “Hypotheses Engineering: First Essential Steps of Experiment-driven Software Development”. In: Proceedings of the International Workshop on Rapid Continuous Software Engineering, Montreal, Canada: IEEE, 2019, pp. 16–19.

[42] Mindtools. “The Build-Measure-Learn Feedback Loop: Creating Real Value by Testing Ideas”. WebPage, Source: https://www.mindtools.com/pages/article/build-measure-learn.htm, Jun 2020.

[43] Nuseibeh, B.; Easterbrook, S. “Requirements Engineering: A Roadmap”. In: Proceedings of the Conference on the Future of Software Engineering, Limerick, Ireland: ACM, 2000, pp. 35–46.

[44] Ohno, T. “Toyota Production System: Beyond Large-scale Production”. Boca Raton, USA: CRC Press, 1988, 152p.

[45] Olsson, H. H.; Bosch, J. “From Opinions to Data-driven Software R&D: A Multi-case Study on How to Close the ’Open Loop’ Problem”. In: Proceedings of the Conference on Software Engineering and Advanced Applications, Verona, Italy: IEEE, 2014, pp. 9–16.

[46] Olsson, H. H.; Bosch, J. “The HYPEX Model: From Opinions to Data-Driven Software Development”. In: Proceedings of the Continuous Software Engineering, Gothenburg, Sweden: Springer International Publishing, 2014, pp. 155–164.

[47] Osterwalder, A.; Pigneur, Y. “Business Model Generation: Inovação em Modelos de Negócios”. Rio de Janeiro, Brazil: Alta Books, 2020, 288p.

[48] Poppendieck, M.; Poppendieck, T. “Lean Software Development: An Agile Toolkit”. Boston, USA: Addison-Wesley, 2003, 240p.

[49] Preece, J.; Sharp, H.; Rogers, Y. “Interaction Design: Beyond Human-computer Interaction”. Hoboken, USA: John Wiley & Sons, 2015, 585p.

[50] Pressman, R. “Software Engineering: A Practitioner’s Approach”. New York, USA: McGraw-Hill, 2009, 928p.

[51] Prestes, M.; Parizi, R.; Marczak, S.; Conte, T. “On the Use of Design Thinking: A Survey of the Brazilian Agile Software Development Community”. In: Proceedings of the Agile Processes in Software Engineering and Extreme Programming, Copenhagen, Denmark: Springer, 2020, pp. 73–86.

[52] Ries, E. “The Lean Startup: How Today’s Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses”. New York, USA: Crown, 2011, 289p.

[53] Rissanen, O.; Munch, J. “Continuous Experimentation in the B2B Domain: A Case Study”. In: Proceedings of the International Workshop on Rapid Continuous Software Engineering, Florence, Italy: IEEE, 2015, pp. 12–18.

[54] Ros, R.; Bjarnason, E. “Continuous Experimentation Scenarios: A Case Study in E-commerce”. In: Proceedings of the Euromicro Conference on Software Engineering and Advanced Applications, Prague, Czech Republic: IEEE, 2018, pp. 353–356.

[55] Ros, R.; Runeson, P. “Continuous Experimentation and A/B Testing: A Mapping Study”. In: Proceedings of the International Workshop on Rapid Continuous Software Engineering, Gothenburg, Sweden: IEEE, 2018, pp. 35–41.

[56] Schermann, G.; Cito, J.; Leitner, P. “Continuous Experimentation: Challenges, Implementation Techniques, and Current Research”, *IEEE Software*, vol. 35–2, 1 2018, pp. 26–31.

[57] Schermann, G.; Cito, J.; Leitner, P.; Zdun, U.; Gall, H. C. “We’re Doing it Live: A Multi-method Empirical Study on Continuous Experimentation”, *Information and Software Technology*, vol. 99, 7 2018, pp. 41–57.

[58] Seaman, C. B. “Qualitative Methods in Empirical Studies of Software Engineering”, *IEEE Transactions on Software Engineering*, vol. 25–4, 8 1999, pp. 557–572.

[59] Sommerville, I. “Software Engineering: Pearson New International Edition”. London, England: Pearson Education Limited, 2013, 768p.

[60] StartupBase. “StartupBase - A Base de Dados do Ecossistema de Startups”. WebPage, Source: https://startupbase.com.br/home, Jul 2021.

[61] Stavru, S. “A Critical Examination of Recent Industrial Surveys on Agile Method Usage”, *Journal of Systems and Software*, vol. 94, 8 2014, pp. 87 – 97.

[62] Sutherland, J. “Scrum: A Arte de Fazer o Dobro do Trabalho na Metade do Tempo”. Rio de Janeiro, Brazil: Sextante, 2016, 293p.

[63] Tripathi, N.; Klotins, E.; Prikladnicki, R.; Oivo, M.; Pompermaier, L. B.; Kudakacheril, A. S.; Unterkalmsteiner, M.; Liukkunen, K.; Gorschek, T. “An Anatomy of Requirements Engineering in Software Startups Using Multi-vocal Literature and Case Survey”, *Journal of Systems and Software*, vol. 146, 8 2018, pp. 130 – 151.

[64] Vargas, B. P.; Signoretti, I.; Zorzetti, M.; Marczak, S.; Bastos, R. “On the Understanding of Experimentation Usage in Light of Lean Startup in Software Development Context”. In: Proceedings of the Evaluation and Assessment in Software Engineering, New York, USA: ACM, 2020, pp. 330–335.

[65] Wohlin, C. “Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering”. In: Proceedings of the International Conference on Evaluation and Assessment in Software Engineering, New York, USA: ACM, 2014.

[66] Wohlin, C. “Second-Generation Systematic Literature Studies Using Snowballing”. In: Proceedings of the International Conference on Evaluation and Assessment in Software Engineering, New York, USA: ACM, 2016, pp. 1–6.

[67] Womack, J.; Jones, D. “Lean Thinking : Banish Waste and Create Wealth in Your Corporation”. Abingdon, England: Free Press, 1996, vol. 48, 393p.

[68] Yaman, S.; Fagerholm, F.; Munezero, M.; Mannisto, T.; Mikkonen, T. “Patterns of User Involvement in Experiment-driven Software Development”, *Information and Software Technology*, vol. 120, 4 2020, pp. 106244.

[69] Yaman, S. G. “User Involvement in Continuous Experimentation”, *SIGSOFT Software Engineering Notes*, vol. 43–1, March 2018, pp. 1–4.

[70] Yaman, S. G. “Initiating the Transition Towards Continuous Experimentation: Empirical Studies with Software Development Teams and Practitioners”, *Journal of Systems and Software*, vol. 151, 10 2019, pp. 95.

[71] Yaman, S. G.; Munezero, M.; Munch, J.; Fagerholm, F.; Syd, O.; Aaltola, M.; Palmu, C.; Mannisto, T. “Introducing Continuous Experimentation in Large Software-intensive Product and Service Organisations”, *Journal of Systems and Software*, vol. 133, 7 2017, pp. 195–211.

APPENDIX A – ETHICS COMMITTEE APPROVAL

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL - PUC/RS

Plataforma Brasil

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: On the understanding of the role of experimentation in technology-based startups

Pesquisador: SABRINA DOS SANTOS MARCZAK
Área Temática:
Versão: 2
CAAE: 41004420.8.0000.5336
Instituição Proponente: UNIAO BRASILEIRA DE EDUCACAO E ASSISTENCIA
Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 4.503.244

Situação do Parecer:
Aprovado
Necessita Apreciação da CONEP:
Não

Endereço: Av.Ipiranga, 6681, prédio 50, sala 703
Bairro: Partenon **CEP:** 90.619-900
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3320-3345 **Fax:** (51)3320-3345 **E-mail:** cep@pucrs.br

APPENDIX B – CONSENT FORM

Project members' telephone numbers were censored in the following version of the consent form, unlike the original version.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Nós, Matheus Jardim Bernardes (aluno de Mestrado) e Sabrina Dos Santos Marczak (professora orientadora), responsáveis pela pesquisa *On the understanding of the role of experimentation in technology-based startups* (em português, Compreendendo o papel da experimentação em startups de base tecnológica), estamos fazendo um convite para você participar como voluntário neste estudo.

Esta pesquisa pretende caracterizar como a experimentação contínua está sendo utilizada no ambiente de startups de tecnologia: identificando quais os principais benefícios e desafios percebidos por praticantes, além de listar quais as ferramentas e práticas mais comuns neste meio.

Acreditamos que ela seja importante porque irá nos ajudar a estudar o impacto da prática de experimentação contínua como fator de sucesso para startups de base tecnológica.

Para sua realização será feito o seguinte: iremos coletar dados dos participantes através de entrevistas seguindo roteiro semiestruturado. Cada entrevista terá duração aproximada de 45 minutos.

Sua participação constará de forma anônima nas publicações que surgirão com base nas entrevistas concedidas. Reforçamos que o objetivo deste estudo não é avaliar o participante, mas, sim, analisar os processos de trabalho relacionados ao tópico da pesquisa. O uso que se faz dos registros efetuados durante o teste é estritamente limitado a atividades acadêmicas e buscaremos garantir seu anonimato e confidencialidade.

É possível que aconteçam os seguintes desconfortos ou riscos: divulgação de dados confidenciais (quebra de sigilo) e desconforto ou constrangimento durante gravações de áudio e/ou vídeo. Você tem o direito de pedir uma indenização por qualquer dano que, comprovadamente, resulte da sua participação no estudo.

Os benefícios que esperamos do estudo são um conjunto de problemáticas identificadas na execução de experimentação contínua além de práticas para evitá-las, ajudando equipes de desenvolvimento de software a atingir o sucesso na condução de experimentos.

Durante todo o período da pesquisa você tem o direito de esclarecer qualquer dúvida ou pedir qualquer informação sobre o estudo, bastando para isso entrar em contato, com Sabrina Dos Santos Marczak no telefone ([redacted]) [redacted], ou Matheus Jardim Bernardes no telefone ([redacted]) [redacted], ambos a qualquer hora.

Em caso de algum problema relacionado com a pesquisa você terá direito à assistência gratuita que será prestada pelos responsáveis da pesquisa, através dos contatos acima citados.

Você tem garantido o seu direito de não aceitar participar ou de retirar sua permissão, a qualquer momento, sem nenhum tipo de prejuízo ou retaliação, pela sua decisão.

Rubrica do participante   Rubrica do pesquisador resp.

Se por algum motivo você tiver despesas decorrentes da sua participação neste estudo com transporte e/ou alimentação, você será reembolsado adequadamente pelos pesquisadores.

As informações desta pesquisa serão confidenciais, e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos participantes, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre sua participação.

Caso você tenha qualquer dúvida quanto aos seus direitos como participante de pesquisa, entre em contato com o Comitê de Ética em Pesquisa da Pontifícia Universidade Católica do Rio Grande do Sul (CEP-PUCRS) em (51) 33203345, Av. Ipiranga, 6681/prédio 50 sala 703, CEP: 90619-900, Bairro Partenon, Porto Alegre – RS, e-mail: cep@pucrs.br, de segunda a sexta-feira das 8h às 12h e das 13h30 às 17h. O Comitê de Ética é um órgão independente constituído por profissionais das diferentes áreas do conhecimento e membros da comunidade. Sua responsabilidade é garantir a proteção dos direitos, a segurança e o bem-estar dos participantes por meio da revisão e da aprovação do estudo, entre outras ações.

Ao assinar este termo de consentimento, você não abre mão de nenhum direito legal que teria de outra forma.

Não assine este termo de consentimento a menos que tenha tido a oportunidade de fazer perguntas e tenha recebido respostas satisfatórias para todas as suas dúvidas.

Se você concordar em participar deste estudo, você rubricará todas as páginas e assinará e datará duas vias originais deste termo de consentimento. Ao assinar e rubricar todas as páginas deste documento, você de forma voluntária e esclarecida, nos autoriza a utilizar todas as informações de natureza pessoal que constam em seu prontuário de atendimento, imagens, resultados de exames e diagnóstico, material biológico se for o caso, para finalidade de pesquisa e realização deste estudo. Você receberá uma das vias para seus registros e a outra será arquivada pelo responsável pelo estudo.

Eu,____, após a leitura deste documento e de ter tido a oportunidade de conversar com o pesquisador responsável, para esclarecer todas as minhas dúvidas, acredito estar suficientemente informado, ficando claro para mim que minha participação é voluntária e que posso retirar este consentimento a qualquer momento sem penalidades ou perda de qualquer benefício. Estou ciente também dos objetivos da pesquisa, dos procedimentos aos quais serei submetido, dos possíveis danos ou riscos deles provenientes e da garantia de confidencialidade e esclarecimentos sempre que desejar.

Rubrica do participante Rubrica do pesquisador resp.

Diante do exposto expresso minha concordância de espontânea vontade em participar deste estudo, autorizando o uso, compartilhamento e publicação dos meus dados e informações de natureza pessoal para essa finalidade específica.

Assinatura do participante da pesquisa ou de seu representante legal

Assinatura de uma testemunha

DECLARAÇÃO DO PROFISSIONAL QUE OBTEVE O CONSENTIMENTO

Expliquei integralmente este estudo clínico ao participante ou ao seu cuidador. Na minha opinião e na opinião do participante e do cuidador, houve acesso suficiente às informações, incluindo riscos e benefícios, para que uma decisão consciente seja tomada.

Data:___________________ ____

Assinatura do Investigador

____

Nome do Investigador (letras de forma)

APPENDIX C – INTERVIEW SCRIPT

This questionnaire is an adaptation of the original one used during the interviews, which was written in Portuguese.

Table C.1 – Interview Questionnaire Script

Questions about the participant's profile	
1	Fullname:
2	Academic background (level and course):
3	Current position:
4	How many years of experience in the technology industry?
5	How many years of work in this company?
6	Did you read any of the books and authors below?
6.a	The Lean Startup, by Eric Ries
6.b	Running Lean, by Ash Maurya
6.c	Experimentation in Software Engineering, by Wohlin et al.
6.d	Lean Inception, by Paulo Caroli
6.e	The Four Steps to Epiphany, Steve Blank
6.f	Other books on Lean and Experimentation?
6.g	What is your perception of the subject?
7	Do you have any education/training on the subject of Experimentation and Lean Startup?
Questions about the company's profile	
8	What is the name of the company?
9	How old is the company (approx.)?
10	How many people work at the company?
10.a	1-5 people
10.b	6 to 10 people
10.c	11-25 people
10.d	26-50 people
10.e	more than 50 people.
11	Provide a brief description of the business model:
12	Where are the company's customers?
12.a	within the state of headquarters
12.b	more than one Brazilian state
12.c	in more than one country;
Questions about the project's profile	
13	Project name:

14	Briefly describe the project's business model.
15	Number of employees:
16	How long have you been working on this project?
17	Briefly describe the profile and role of the project members:
18	How is the team organized?
18.a	Locally, in a single office
18.b	Distributed in national territory
18.c	Globally distributed
18.d	If other, describe.
19	Is the project already online?
If so:	
19.a	How big are the users and customers today?
19.b	Do you collect usage metrics (Analytics)? How?
19.c	Is there anyone responsible for analyzing these metrics? What is the process involved?
If not:	
19.d	Do you intend to collect access metrics?
19.e	Did you elaborate a work plan to deal with this data?
20	What management methodology is used?
20.a	Scrum
20.b	XP
20.c	Kanban
20.d	Waterfall
20.e	RUP
20.f	If other, describe.
21	How are projects published? (CD or manual?)
Questions about experimentation tools	
22	Which of the tools below does your project use? Detail how they fit into the work routine.
22.a	A/B test
22.b	Canary release
22.c	Dark spear
22.d	False door tests
22.e	Landing pages;
22.f	Mockups and wireframes
22.g	MVP and MVF
22.h	Multivariate tests
22.i	If other, describe.

22.j	None of the previous;
23	Does your project have a hypothesis backlog?
If so	
23.a	Who manages it?
23.b	How do the hypotheses arise?
If no	
23.c	(no question)
24	Does your team conduct experiments?
If so:	
24.a	Describe the procedure for the experiments.
24.b	Simultaneous or isolated?
24.c	How long does an experiment last?
24.d	How is the implementation of these experiments prioritized?
24.e	Which metrics are measured in the experiments? And how are the results analyzed?
24.f	What are the perceived benefits of adopting this practice?
24.g	And what are the challenges?
24.h	What factors helped in the adoption of this practice?
24.i	Do you think the experiments are being carried out well?
	Are there well-defined success criteria for running an experiment?
24.j	Was there a cultural barrier to the adoption of this experimentation practice?
	Are they still there? How were these barriers overcome?
If not:	
24.k	Is the team interested in the subject?
24.l	What is your definition of experimentation?
24.m	What prevents adoption? For example, cultural or organizational barriers, technology.
Conclusions	
25	Any other comments on the subject that you would like to report?

APPENDIX D – DEBRIEFING QUESTIONNAIRE

Courtesy of Prof Cleidson de Souza, UFPA, Brazil

Debriefing Questions

Local:
Date:
Interviewer:
Interviewee:

1. What were the key points observed about the focus?

2. What did you find to be most surprising about this observation?

3. What did you see or hear that was pretty much what you expected (or like other sites that you have seen)?

4. What did you learn about the problem and "fixes" that you didn't know before? That you did?

5. What would you ask if we could go back? Would you ask the next participant this as well?

6. What worked really well?

7. What didn't work so well or what should be changed?

8. Other comments?

APPENDIX E – THEMATIC SYNTHESIS MIND-MAP REPORT

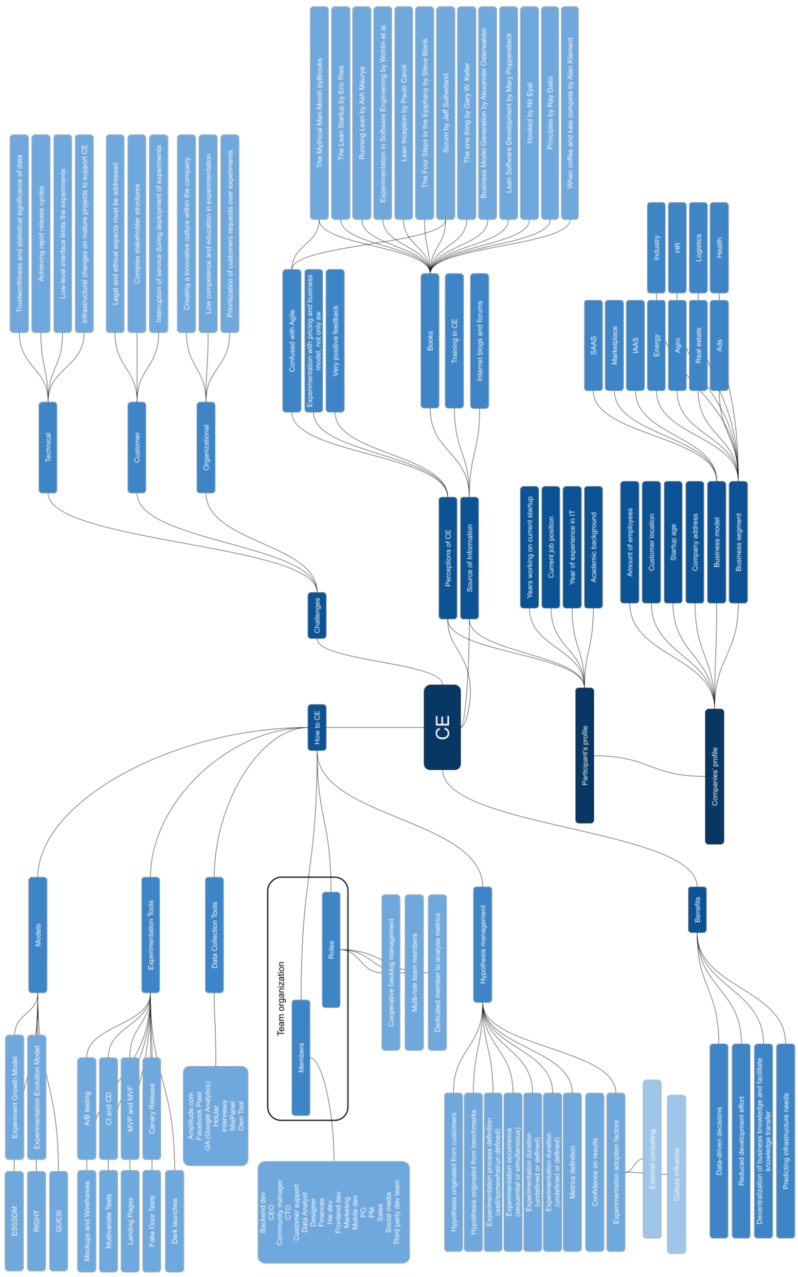

Figure E.1 – Mind Map of Interviews' Codes

APPENDIX F – SUMMARY OF FINDINGS ACCORDING TO SOURCE

Table F.1 – Comparison of Findings according to Source

	Literature	Participants
Benefits		
Data-driven decisions	[14, 20, 45, 53, 69, 71]	[P3, P8, P10, P12, P13]
Reduced development effort	[20, 38, 45, 53, 64, 71]	[P1–P4, P12, P15, P16]
Decentralization of business knowledge and facilitate knowledge transfer	[14, 15, 38, 64, 71]	[P4]
Predicting infrastructure needs	[14, 15]	
Increased customer engagement		[P5, P9, P10]
Improved communication with investors		[P5]
Technical Challenges		
Trustworthiness and statistical significance of data	[14, 20, 21, 31, 35, 53, 55, 69]	[P1–P4]
Achieving rapid release cycles	[20–22, 35, 56, 57]	
Low-level interface limits the ‘experiments’	[53]	
Infrastructural changes on mature projects to support CE	[53]	
Organizational Challenges		
Creating a innovative culture within the company	[15, 31, 53, 55, 64, 68, 71]	[P1, P3, P12, P15]
Low competence and education in experimentation	[20, 21, 35, 53, 56, 71]	[P1, P2, P5, P8, P13, P16]
Prioritization of customers requests over experiments	[53]	[P1, P3, P6, P7, P9]
Customer Challenges		
Legal and ethical aspects must be addressed	[16, 53, 68]	
Complex stakeholder structures	[53, 69, 71]	[P3, P10]
Interruption of service during deployment of experiments	[38, 53]	[P15]

Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br