Amasya, Basak; Fagerholm, Fabian

Designing experiment objects for Continuous Experimentation: An exploratory multiple-case study

Published in:
Proceedings of the 29th International Conference on Evaluation and Assessment in Software Engineering

DOI:
10.1145/3756681.3756944

Published: 01/01/2025

Document Version
Publisher's PDF, also known as Version of record

RESEARCH-ARTICLE

Designing experiment objects for Continuous Experimentation: An exploratory multiple-case study

BAŞAK AMASYA, Aalto University, Espoo, Uusimaa, Finland

FABIAN FAGERHOLM, Aalto University, Espoo, Uusimaa, Finland

Published: 17 June 2025

Citation in BibTeX format

EASE '25: Evaluation and Assessment in Software Engineering
June 17 - 20, 2025
Istanbul, Turkiye

EASE '25: Proceedings of the 29th International Conference on Evaluation and Assessment in Software Engineering (June 2025)
https://doi.org/10.1145/3756681.3756944
ISBN: 9798400713859

Designing experiment objects for Continuous Experimentation: An exploratory multiple-case study

Başak Amasya
Aalto University
Espoo, Finland
basak.amasya@alumni.aalto.fi

Fabian Fagerholm
Aalto University
Espoo, Finland
fabian.fagerholm@aalto.fi

Abstract

Continuous experimentation relies on the design and implementation of experiment objects: artefacts that are used as the treatment or stimulus in a software product experiment. To be used in practice, experiment objects must have traits that allow drawing correct inferences regarding a new product, while simultaneously being resource-efficient to produce. In this paper, we investigate what kinds of experiment objects companies use, how they are selected, what factors influence the selection, and how they evolve over time. We present a multiple-case study based on semi-structured interviews with representatives from eight companies. Four kinds of experiment objects were found: prototypes, minimum viable products, proofs-of-concept, and wireframes. The results show that deciding what kind of experiment object to develop is a collaborative and customer-oriented process. As experimentation proceeds, experiment objects can be thrown away, modified for new experiments, or even become products. Experiment objects were understood not only as artefacts in product experiments, but blended with purposes familiar from prototyping and user interface and experience design, and they were used as vehicles for customer communication. The study shows that experiment objects are created for a variety of purposes and understandings of experimentation. Guidelines for creating and managing effective experiment objects for different variations of continuous experimentation are called for.

CCS Concepts

• **Software and its engineering → Software design engineering**.

Keywords

experiment object, minimum viable product, continuous experimentation, case study

ACM Reference Format:
Başak Amasya and Fabian Fagerholm. 2025. Designing experiment objects for Continuous Experimentation: An exploratory multiple-case study. In *Evaluation and Assessment in Software Engineering (EASE '25), June 17–20, 2025, Istanbul, Turkiye*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756681.3756944

ACM ISBN 979-8-4007-1385-9/25/06
https://doi.org/10.1145/3756681.3756944

1 Introduction

Continuous experimentation (CE) – the practice of using field experiments with users to inform software product development – has become commonplace in many companies [2], especially well-known large enterprises such as Microsoft and Netflix [23]. CE seeks to improve software product development decisions by basing them on information about how users react to changes in the product [2, 14, 21]. It crucially relies on the creation of different kinds of experimental artefacts [14] which we call *experiment objects*. These are chosen or designed for the experiment and presented to users as the experimental treatment or stimulus. Experiment objects are not necessarily the same as the final product. For example, experimenters can attempt to save resources, creating an experiment object that they believe has the essential traits to stand in for some part of an imagined final product.

There are many open questions regarding the design, use, and management of experiment objects. Prior literature has proposed models and processes for conducting CE which contain the notion of experiment objects (e.g., [14, 21, 22]) and explored the conceptual and empirical aspects of some kinds of experiment objects (e.g., [1, 13, 24, 31, 39]), but how practitioners understand the concept and what range of concrete artefacts they use as experiment objects, has not been extensively studied.

The goal of this study is to explore and understand what artefacts are used as experiment objects in the industry, how they are used to support experimentation, and what happens with them after an experiment is finished. Specifically, we investigate the following research questions:

- RQ 1: What kinds of experiment objects do practitioners identify and utilise for their experiments?
- RQ 2: How do experiment objects evolve as the continuous experimentation process proceeds?

To address the research questions, we conducted a multiple-case study with eight companies. The case study is exploratory and seeks to learn from the experiences of practitioners. We interviewed company representatives in diverse roles who are involved in the creation and management of experiment objects in different forms. The study is based on results of the first author's master's thesis.

The study contributes, based on real-life cases, an understanding of how practitioners concretely see experiment objects and utilise them in software product development. It thus helps to improve the conceptualisation of experiment objects based on practice. It also contributes further knowledge of the considerations that go into the choice and creation of experiment objects, and thus develops a better understanding of factors that influence decisions about experiment objects and their use for different purposes. Finally, the study relates practitioners' understanding of how experiment objects are used at different stages of development, pointing towards opportunities for future guidelines and methods to assist in managing experiment objects over their life-cycle.

The paper is organised as follows. In Section 2, we briefly review the literature on CE, minimum viable products (MVPs), minimum viable features (MVFs), and prototyping. Section 3 explains the study research method. We describe the study results in Section 4, discuss findings, possible implications, and recommendations in Section 5, and present final remarks and conclusions in Section 6.

2 Background

Continuous Experimentation can be seen as part of Continuous Software Engineering, the industry trend of Agile software development and other iterative practices which aim to establish a feedback loop between customer demand and fast delivery of the software product [17]. In this section, we first discuss CE models and processes, showing how experiment objects appear in them. We also consider the Minimum Viable Product (MVP) concept since it is frequently discussed in relation to CE. We then discuss relevant concepts from the literature on prototyping. Together, the discussion on CE, MVPs, and prototypes helps to establish an initial conceptual understanding of experiment objects that frames our study.

2.1 Continuous Experimentation and Experiment Objects

CE structures the software product development process in terms of iterative, data-driven decisions based on observing the consequences of product changes [2, 14, 17]. By establishing how users react to the changes, product development may proceed in different directions. The most popular experimental method is considered to be A/B testing, where variants of a product are compared in terms of their impact on a chosen dependent variable, often related to user behaviour and business goals [2]. For example, a company wishing to know if a search engine feature increases ad revenue, may conduct an experiment comparing the ad revenue of a sample of customers receiving a search engine variant with the feature to that of customers receiving a baseline variant without the feature.

Experiments have the potential to decrease the uncertainties that arise from making assumptions about the viability of business models, strategies, and product features [14]. In the literature, *experimentation* is used both as a broad term for an unstructured, trial-and-error approach [6] and for systematic, structured, controlled experimentation [2, 26]. The extent to which experimentation should be structured and systematic can be debated and may be situationally dependent. However, CE intends to bring more structure to the experimentation process through various models [2, 14].

Many CE models, such as HYPEX [21] and RIGHT [14], imply or assume that the experiment cycle involves creating or modifying a software artefact that constitutes or is part of the experimental treatment. The RIGHT model [14] specifies that Minimum Viable Products (MVPs) or Minimum Viable Features (MVFs) are deployed in field experiments to rapidly test product hypotheses. However, Fagerholm et al. [14] also mention that in certain cases such as initial stages of development, experiments can still be conducted even without an MVP, implying that a different kind of artefact can be used instead. The HYPEX model [21] also uses the term MVP, but does not extensively detail what this concept means or how to create an MVP.

To our knowledge, no CE model to date provides further detail on what these experimental artefacts are, or how they should be designed and managed. In this paper, we use the term *experiment object* to denote the artefact in abstract terms, without locking the concept to any particular concrete instance of such an artefact. Experiment objects could be variants of an existing product, but they could also be artefacts created specifically and only for the experiment. They could be instantiated in many forms, such as a paper prototype, or a full-fledged interactive software implementation.

2.2 Minimum Viable Products and Minimum Viable Features

The MVP concept originates in the practitioner literature and was popularised particularly by Ries [34] and Blank [4]. Its definition is continuously evolving [24], but it has been defined as a version of a product that allows learning about a company's customers with minimal effort [34]. An MVF is the same idea applied to an individual feature of the product [14]: a variant of a product feature that can be built with minimal effort and used to learn about how customers respond to it.

Lenarduzzi and Taibi [24] explored the definitions of MVP in a systematic mapping study using both scientific and grey literature. They found that the notion of minimality could refer to several things, such as minimum effort to develop, having minimum functionality, minimum value organisation, minimum requirements, or smallest possible implementation, and that all these varying definitions could be traced to Ries' and Blank's original concepts. The most recurring notion of minimality was "minimum features".

Beyond the notion of minimality, MVPs can be understood in terms of their purpose. Lenarduzzi and Taibi [24] found various ideas of the purpose, e.g. allowing the product to be deployed, creating a viable product for the customer, gathering customer feedback, and collecting the maximum amount of validated learning about customers. In Lean Startup methodology, MVPs allow gathering customer knowledge with minimal effort but they are not necessarily meant to be sold as the final product [34]. In contrast, others define MVPs more in terms of a target: a basic version of the product that is viable in the sense that customers are willing to buy it (c.f. [5]). In the build-measure-learn cycle of Lean Startup, the MVP can be seen as an artefact to be used for customer validation [13]. Using the same perspective of validation cycles, an MVP is a product version that allows evaluating a cycle's hypothesis [39].

MVPs can also be examined empirically by asking how they are used by practitioners in reality. Duc and Abrahamsson [13] explored five early-stage startups to understand how they used MVPs, basing the classification of MVPs on Ries' list of MVP types [34]. They found that MVPs supported validated learning, facilitated product design, bridged communication gaps, and facilitated cost-effective product development activities, and that they acted as boundary-spanning artefacts between the entrepreneur team and external stakeholders. However, they noted that practitioners were not fully aware of the possible roles MVPs could take. Alonso et al. [1] found that various practices are used to develop MVPs, such as design thinking and user-centred design, and that end-user validation is emphasised for validating MVPs. Similarly, Matturro et al. [27] found that practitioners used landing pages and mock-ups as MVPs, and validated them using interviews and user comments. In both studies, a wide variety of artefacts were used, and not all validations were online controlled experiments (such as A/B tests) but workshops, interviews, surveys, and other user-centred design methods were also frequently used.

While the MVP concept has become popular in practice and is used in the scientific literature, its many meanings and uses make it problematic when investigating the artefacts that are used in CE. As demonstrated by the many meanings discussed above, the practitioner terminology is inconsistent and the MVP term may be used both as a broad concept encompassing the business and product experimentation aspects of CE, but also as a synonym for various types of artefacts. Not all artefacts that act as part of the experimental treatment in CE can be conceptually considered equivalent to MVPs or MVFs, because the conceptual meaning of the latter extends beyond their potential role in the experimental treatment. We therefore consider MVPs, and the smaller variant, MVFs, in the role of an experimental treatment, to be an instance of the more abstract notion of an experiment object. When looking for experiment objects in practice, we expect practitioners to speak frequently of MVPs, but we also expect that other terms and concepts may be used. Furthermore, based on the existing literature, we should expect that experiment objects play a role in many modes of experimentation, including A/B testing and qualitative approaches.

2.3 Minimum Viable Products and Prototypes

Prototypes and prototyping have been extensively studied in software engineering and human-computer interaction research. Their definition and usage seem to overlap with those of MVPs. Although defining prototyping in software development is more challenging than in other areas of engineering [18], a prototype can be seen as a functioning model for developing and evaluating design ideas [41]. Prototypes can differ in purpose, media, scope, use, and exploration strategy [3]. In terms of these differing aspects, some prototypes may be close to or equivalent to some of the definitions of MVPs discussed above. For example, the *purpose* of a prototype may be to *validate* market desirability and its *use* may be *with* potential customers in their real *environment* – the emphasised aspects correspond to the model by Bjarnason et al. [3]. User interface (UI) prototyping, in particular, represents a process of creating mock-ups of the final product's UI, providing a way for designers to investigate different design concepts and receive feedback from relevant stakeholders before finalising design decisions [32]. In CE, full implementation can be avoided when the prototype reveals flaws in the design [2].

There appears to be an ambiguity in the literature regarding the difference between MVPs and prototypes. The terms are sometimes used interchangeably [13, 39] and some consider that prototypes can serve as an MVP or vice versa, depending on how the concepts are defined (c.f. [3, 31]). It has been claimed that MVP development can be seen as similar to prototyping [39]. The prototype concept fits both the MVP aspect of user assessment and MVP cycle, and MVPs can cover aspects and possibilities of throwaway and evolutionary prototypes: an MVP can be reused if the experiment supports the product hypothesis, otherwise it should be thrown away [39]. As can be observed, the MVP and prototype concepts are mutually overlapping. From the perspective of CE, neither exactly equivalent with the notion of an experimental artefact used as a treatment, but in practice, both can be expected to denote such an artefact in practitioners' everyday speech. For this reason, the delineation of our search for experiment objects considers the definition of prototypes and the varied purposes, media, scopes, uses, and exploration strategies related to prototypes discussed above.

2.4 Prototype Fidelity and Implications for Experiment Objects

Prototype fidelity refers to how close a prototype is to the final product, e.g., a UI prototype compared to the final UI [10, 40]. The question of how closely the prototype should represent the final product has been debated in the literature [35]. If the fidelity of the prototype is too low or too high, it may not be suitable for the desired kind of user feedback [9, 25, 35]. Low-fidelity prototypes are cheaper and quicker to make, but leave more details open for implementation [35]. High-fidelity prototypes require more resources to create, and may lead to premature design decisions being made before validating them [32, 35]. On the other hand, their functionality can be more complete, allowing the user to explore the prototype as if it were the final product. High-fidelity prototypes can also be useful for marketing and sales purposes.

Similarly, experimenters in CE must decide what fidelity their experiment object should have. To draw correct conclusions from an experiment, the experiment object must be representative of the traits in the final product that influences the user behaviour of interest. However, the desired final version of the product is not always known beforehand [39], making the choice and design of experiment objects challenging. Customers may also form expectations of the final product based on the experiment object, similarly to how they may perceive a high-fidelity prototype as being ready for delivery [35], potentially altering their behaviour towards the final product in unexpected ways.

3 Research Method

To address our research questions, we used an exploratory multiple-case study approach with semi-structured interviews which were analysed qualitatively. We examined eight companies through interviews, yielding insights into the role of experiment objects in each, and allowing the identification of commonalities and differences. In this section, we describe the study design, sampling strategy, data collection and analysis methods, and the included cases. The study data are not publicly available due to anonymity reasons.

3.1 Case Study Design

A case study is a thorough explanation and evaluation of a real-life, contemporary bounded system – the case [11, 30]. Multiple-case studies examine various instances of a class of bounded systems. Studying multiple cases may be motivated by the desire to generalise the results [30, 43] but also to gain more varied knowledge through the differences present in the cases [11]. The latter is the primary motivation for our selection of multiple cases for this study.

Table 1: Semi-structured interview guide.

ID	Question
1	Can you give me an example of how you use experiments in your development process?
2	How did this start? Did you prepare any materials for the experiment? Is this the typical way you make them?
3	What kinds or types of materials do you use? Can you give examples?
4	How does your company choose which kind of material you use for experiments?
5	If someone would ask advice on choosing different types of materials, what advice would you give?
6	Can you describe some of the challenges that you have faced while designing experiment objects?
7	How do you overcome these challenges?
8	How do you understand that the material you are bringing to the experiment will lead to the right conclusions?
9	What do you do with the experiment object after you are done?

Table 2: Study participants and case company characteristics.

Case	Role	Tenure	Business Domain	Size	Main Location	Customer Market
A	Junior software developer	6 months	Marine and energy	Large	Finland	Global
B	Co-founder	3 years	Freight	Small	Turkey	Turkey
C	Senior software engineer	2 years	Software engineering	Large	USA	Global
D	Managing director	3 years	Healthcare	Medium	Netherlands	Europe
E	CIO	3 years	Insurance and banking	Small	Austria	Europe
F	Senior technology architect	11 years	Consultancy	Large	Ireland	Global
G	UI/UX designer	2 years	E commerce	Large	Turkey	Europe and Mid. East
H	Product manager	1 year	Software engineering	Medium	Turkey	USA and Europe

Tenure: length of employment in the company at the time of the interview. Size: Small: ≤ 10, Medium: ≤ 200, Large: ≥ 5000 employees. Main location: location of the company's head office.

A case study is defined by the unit of analysis, not the explored subject [11, 30]. The unit of analysis in this study is companies where experiment objects are used in continuous software product experiments. In addition to identifying the unit of analysis, we identify the use of experiment objects as the "quintain" or "object or phenomenon to be studied", as suggested by Stake [38], providing a guide for analysis across the cases. An initial understanding of the concepts used in defining the unit of analysis and the quintain – such as the experiment object and its role in CE – is given in Section 2. This understanding helps us focus the case study on those aspects of the unit of analysis that we are interested in.

The intent of conducting this case study is *instrumental* (c.f. [37]) because we seek to understand a specific issue. Thus, we are not primarily interested in the cases themselves. Instead, they act as vehicle to understand the concept of experiment objects better.

3.2 Sampling Strategy

A convenience sample with traits of purposeful sampling was used. Purposeful sampling involves selecting "archetypal" cases that serve the theoretical goals and research questions [33]. We used a combined strategy of typical and maximum variation cases [19]. In other words, we aimed to select companies with both commonalities and differences, not focusing on special or extreme cases. Our goal was to explore patterns in a diverse set of understandings of experiment objects rather than to generalise findings regarding their utility, such as successful or unsuccessful use in experiments.

We identified companies and participants by searching in startup incubators' member lists, through web searches, LinkedIn, and our contacts. We sought companies in different business domains, of different sizes, and with customer markets spanning different parts of the world. We also sought interview participants with a variety of roles and different lengths of experience to achieve further variation. This way, the experimentation processes could be understood from a variety of perspectives, broadening the exploration. We note that the intent was not to enable comparison of experiences from participants in the same roles in different companies, since our study is instrumental and focuses on the experiment object phenomenon rather than examining individual differences among practitioners.

There is no single answer to how many cases a multiple-case study should have [11]. An approximate guideline could be between four and ten cases [20]. Creswell and Poth [11] suggests that no more than four or five cases are typically chosen, and points out that the inclusion of too many cases may dilute the overall analysis. To determine the number of cases, researchers should consider the research goal, the complexity of the subject, the notion of saturation, and practical constraints [11, 30]. Saturation is a key consideration and it is reached when new interviews no longer provide additional information or perspectives [20, 36]. We did not define a precise number of desired cases upfront. Instead, we followed the variation in cases as data collection proceeded and applied the principle of saturation to determine the number of cases.

3.3 Data Collection

We used semi-structured interviews, as they fit the exploratory study goals. Semi-structured interviews cover chosen themes and include prepared questions [7], but allow for improvisation [36]. We formulated interview questions based on the research questions and our conceptual understanding of the subject. Pilot sessions were conducted with company contacts to help clarify and focus the questions for easier understanding by practitioners. The sessions also tested the flow and order of the questions and helped estimate the length of the interviews. The questions are shown in Table 1.

The first author conducted the interviews online in March 2024 and they were recorded. In the beginning of the interview, informed consent was obtained, the objectives of the interview and the study were explained briefly, and background questions were asked. Notes were taken during the interviews.

3.4 Data Analysis

Data analysis was carried out concurrently with the data collection, which is a common characteristic of qualitative research [36]. New perspectives can be discovered and inform subsequent data collection and analysis. Immediately after each interview, the first author wrote summaries and memos for the purpose of data analysis. In addition to assisting the analytical thinking process, these summaries and memos were especially useful in highlighting key points and enabling later recollection of contextual details. The recorded interviews were then transcribed. New insights can be obtained during the transcribing phase [36]. The transcripts were coded in a bottom-up manner [16] by identifying important statements which were subsequently grouped into higher-level, named themes. This coding process was guided by the research questions. Insights from the interviews and the ongoing coding were continually discussed with the second author as a means to validate and expand the analysis. The final coding and resulting interpretations were reviewed by both authors together.

3.5 Case Descriptions

The participants usually discussed more than one project in which they had previously taken or were currently taking part. They also reported on their overall insights based on their experience from many projects. Case company characteristics and study participants are described in Table 2. All companies have existed for more than 14 years except three, which were founded approximately 3 years ago. In total, eight interviews were conducted at eight different companies.

4 Results

In this section, we present the case themes that emerged from analysing the interviews. We discuss the participants' understanding of experiment objects, the types and uses of experiment objects in the companies, the factors influencing the choice of experiment objects, and their evolution in the development process. The results presented here are the synthesised cross-case results, but we also show case-specific examples to illustrate details.

4.1 Understanding Experiment Objects

The participants displayed widely different understandings and terms for many tools, methods, and artefacts they used for experimentation. The inconsistencies in terminology were also visible in the discussion around experiment objects. The experiment objects that were mentioned during the interviews were *MVPs*, *proofs-of-concept (PoCs)*, *prototypes*, and *wireframes*. Note that these terms are used here as words chosen by practitioners, and are not necessarily congruent with how the same terms are used in the scientific literature. They occurred in the interviews when the participants spoke about the materials they used for experiments. Table 3 shows the experiment objects discussed by each participant.

Table 3: Experiment objects mentioned for each case.

Case	Prototype	MVP	PoC	Wireframe
A			×	
B			×	×
C	×	×	×	×
D		×	×	
E	×			
F	×	×	×	
G	×			
H		×		

4.1.1 Prototype. The distinction between prototypes as experiment objects and prototyping as an approach was not always clear in the interviews, and we discuss them together under the term prototype. The participants tended to use these meanings interchangeably. They also introduced various nuances specific to their ways of using prototypes. For instance, one of the participants defined prototyping as designing the experiment "without any code" (Co-founder, Company B).

One participant mentioned that they used a third-party agency for prototyping that resembled a user flow demonstration approach.

> "I think prototyping is a great tool to bridge different expectations between different parties. I think that is the great value of tools like Figma, to have a common language and common base for developing applications." (CIO, Company E)

The same participant stated that there were valuable benefits of this process and that they were considering using prototyping in further phases of the development.

Prototype testing was another nuance of the concept that came up during one of the interviews. This approach involves observing users interacting with prototypes of new screens, including both redesigns and screens designed from scratch. By letting the user explore the prototype and closely monitoring their interactions, the goal is to identify potential issues and ensure that new features are intuitive to use.

> "For example, we made a new screen, but [...] We are not sure if they can find this new button. We are writing a scenario [...] to see where the user clicks on the screen." (UI/UX designer, Company G)

For Company G, the number of users to be included in prototype testing was decided according to saturation of the results – the point where experiment results do not change significantly with added participants and new insights are unlikely to be uncovered. By reaching saturation, the design team would be sure that they have seen most possible issues and user behaviours.

4.1.2 MVP. Multiple definitions for MVPs were voiced in the interviews, similarly to the situation described in several prior studies, as discussed in Section 2. Nevertheless, the definitions have commonalities and match with many of the definitions found in the literature. Practitioners seem to view MVPs as software artefacts already functioning to some extent, evolving to be a product, and as bigger investments. One of the participants gave an example by analogy to explain what defines an MVP and how it can evolve into a product.

> "What does the car need? Four wheels, a working engine, and a steering wheel, right? So these are the 'sine qua non' of a car. [...] Afterwards, we are adding features on the basis of a little more competitor analysis. [...] It is known that the product develops after considering the things that can make a difference with the competitor analysis and the supported budget." (Product manager, Company H)

Another participant viewed these kinds of experiment objects from the point of view of their relationship with the final product.

> "MVP is on its way to be a product. So it should from the start be considered as a product, all the product management activities should be processed over the MVP." (Managing director, Company D)

This indicates that developing an MVP can often require as much consideration as developing a complete product. The interviews raising this point implied that the effort required to create an MVP can be larger than for some other kinds of experiment objects.

Some participants mentioned that MVPs may be eligible for grants from governments or other institutions supporting development of new products. Such grants can aid companies in their investments, but some of our participants also considered MVPs not awarded grants as investments in their own right.

4.1.3 PoC. The term PoC came up in the interviews when some participants described the materials they used in experimentation. To our knowledge, the term has not been extensively studied in the literature from the perspective of how practitioners use it. We observed a wide range of viewpoints on PoCs in the interviews.

Nevertheless, we found a common conceptual core: PoCs can be understood as preliminary implementations that aim to demonstrate the feasibility of a concept or idea.

Participants differed in terms of considering PoCs as internal or external and whether they are implemented from scratch or as a feature extension on an existing product. Internal PoCs were considered to be used locally in the company, typically in a specific organisational unit such as a team. Their purpose was most often to experiment with a technical concept. External PoCs were considered to have a role in validating the feasibility of a solution to a party external to the company, such as a customer. Some of the participants explained this terminology from the point of view that there needs to be a concept to be proven or there needs to be a customer's concept to be rehearsed with. It was also emphasised that the idea for these types of experiment objects is more likely to come from the customer than in the case of other experiment object types.

Technical feasibility seems like the main aspect to test in PoCs. Some interviewees indicated that PoCs are mainly suitable for tailor-made or customer-specific software solutions whereas MVPs are more suitable for product-based solutions aimed towards consumers.

> "PoC can be a showcase for a tailor-made requirement coming from the customer." (Managing director, Company D)

Doing a PoC can also mean trying out a new update on the development environment first before putting it into production. Especially if the company is trying out an innovative idea or if the idea or the product is crucial for the business, they might like to keep this private and hidden from market competitors. In these cases, PoCs would not be shared with external stakeholders. This can be compared with dark launches [15], but with even further limited visibility.

Although one of the participants mentioned that PoCs are internal "by default" (Senior technology architect, Company F), there were some cases where the participants gave examples of external PoCs. Moreover, PoCs can also be considered as a marketing or presales tool. It can be used to acquire new customers or enter a new market.

PoCs can also be examined from the perspective of whether they are created entirely from scratch or built as an experiment on top of an existing solution. One participant acknowledged that it is possible to do a PoC from scratch but they have always done it by trying out new features on top of the existing product. Based on how the participants used the term, both scenarios seem possible.

4.1.4 Wireframe. Although wireframes seem to take part in all case companies' software development practices, they do not always serve as an experiment object, and were mentioned as such in only two cases (see Table 3). According to the interviews, they are sometimes used as a layout guide for the development team. Sometimes, however, they can also be shown to the customers in demo sessions to gather feedback and then act as an experiment object. Creating wireframes can also be seen as one of the development phases.

Table 4: Factors that influence the choice of experiment object for each case.

	Case							
Factor	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Business Value	×				–	×		
Current Situation		×		×	–	×		
Customer		×	×	×	–	×	×	×
Effort		×	×		–		×	
Legal				×	–	×		
Market					–			×
Purpose					–		×	
Technological Attractiveness		×			–			
Feasibility		×	×		–	×		
Resources	×	×	×		–			×

Case E did not yield any factors.

> "So initially, wireframe is to basically help us see the layout UX, and then the functionalities, how it will fit the screen. So after the prototype, we might create a wireframe." (Senior software engineer, Company C)

The participant added that the functionality of the product should be defined clearly while creating wireframes with all necessary functions. This implies that they are using the prototypes as experiment objects to get more clarity on specific product details.

4.2 Factors Influencing the Choice of Experiment Object

Many factors were discussed while talking about how companies choose between different experiment objects. It should be noted that participants may have forgotten to mention some factors or have not realised that they play a role in the decision-making process. The participants were not explicitly prompted to name factors. Many participants mentioned that their choice would depend on the specifics of the project.

Factors that play a role in choosing experiment objects are shown in Table 4, as discussed by each participant. Note that in Case E, no factors occurred that would influence the choice of experiment object. Only prototypes were mentioned as an experiment object in this case (see Table 3).

4.2.1 Purpose of Experiment Objects. As a factor on its own, the purpose of the experiment object – what it will be used for – was mentioned explicitly only once. However, purposes of experiment objects were discussed implicitly throughout the interviews. Some examples were already mentioned among the quotes in the previous section, so we do not repeat them here. They include communication internally and with customers, marketing, turning the object into a final product, and testing technological feasibility, but also importantly, experiment objects were created to learn about customer preferences and the viability of product development ideas and features.

In some cases, experiment objects were used for purposes outside the experiment process. For instance, wireframe diagrams which eventually are not used in the design could still be kept as an idea pool for the future. Furthermore, even if an MVP does not evolve into a product, the company can still learn from it, especially from the development team's and software engineers' point of view.

> "Some MVPs were not able to become a product in the market. But we learned a lot. We applied new technologies, we applied them to our solution backbone. So it is never a loss from my perspective." (Managing director, Company D)

The participants' comments indicated that knowledge gained in these kinds of situations can also be applied to the process of designing other experiments.

4.2.2 *Customer ("for whom the experiment is done").*

This was the most commonly mentioned factor playing a role in picking the type of experiment object and developing it (see Table 4). Some participants stated that in determining the direction of product development, the client's wishes come first, whereas some mentioned customer requests as a factor with lower priority. We observed in the interviews that participants considered it to be important to take into account whether the idea is coming from the customer or not, since it should be kept in mind that there can be some impacts or side effects that the customers cannot foresee when they are requesting something.

An important concept that was uncovered in this study was the concept of "right customer" which was raised by some of the participants. According to these participants, such customers play a crucial role in guiding the new product development process. Participants emphasised the importance of finding the right customers to gather feedback from: those who will provide the "right" feedback and lead product development in the right direction.

> "If it is the right customer, they direct you." (Co-founder, Company B)

These customers are more likely to tolerate mistakes that can occur in the early phases of development. Accordingly, one participant also highlighted the risk of going for big-name clients too early, as their intolerance for initial flaws can have negative impacts on the company's reputation (Product manager, Case H). We underline that the participants discussing this concept work at startups. This suggests that for startups, finding the right customer may be more critical than in more established companies.

The interviews indicated that there seems to be more trust towards the feedback that comes from the right customer than from that coming from just any customer. In this context, the right customer can also mean the end-users who have been using the product for a long time.

> "So if you're in trust with the people who use your systems, it could be like years long program or project, then you can actually see that, [the feedback] would be emerging from daily work and daily life of the client as well. So that they will be just like stating immediately, this [the features that being experimented on] saves us so much time." (Senior technology architect, Company F)

In summary, the customer participating in the experiment influences the choice of experiment object through the feedback they give in different stages of the experimentation process and through their relationship with the company.

4.2.3 *Market and value.*

Two factors that practitioners take into account in the decision-making processes related to experiment objects concern market research and business value.

Market research: Can be conducted to identify the gap in the sector and the experiment object can be chosen and the experiment process can be driven based on that information. Market research can also help determine the potential business value of the experiment object.

Business value: Is considered especially if the decision is being made by higher management. They tend to focus on whether an experiment will actually bring value to the company and also pick the experiment object accordingly.

4.2.4 *Technical factors.*

Technological attractiveness: Even a technologically simple innovation can sometimes bring value to the company. This effect appears to be particularly noticeable for technologically attractive solutions. A certain feature or extension of the product may be difficult or costly to implement, but it might still be worth doing if it will attract customers. The most common example of state-of-the-art technology in the interviews was generative AI-based enhancements to existing products. This aspect was frequently mentioned in connection with business value (see Section 4.2.3).

Technological feasibility: One purpose of experiment objects appears to be determining whether something is technologically feasible. Technical factors are then evaluated to see if a proposed solution has potential. The most common choice of experiment object mentioned for this was PoCs.

4.2.5 *Limiting factors.*

Some limiting factors were also mentioned, including available resources, the estimated effort needed for designing and developing the experiment objects, and compliance with legal regulations.

The estimated effort: The effort for developing an experiment object is taken into consideration together with its estimated impact. The aim is to optimise the trade-off between the effort put in and the impact it will have later.

The development team's current situation (the current backlog): In the cases we examined, experiment objects are created by the same teams and persons doing other development tasks. The same resources must then be allocated both to regular development and experimental activities. Participants mentioned also taking this into consideration while choosing experiment objects and prioritising experiments.

Resources: This factor is related to the team's current situation, but goes beyond it. It includes human, time and financial resources. It should be kept in mind that not all human resources are equal in the sense that jobs may require specialised skills. The choice of experiment object may thus depend on available skills.

Legal regulations: These can play a decisive role. In certain cases, experiment objects can be used to work around some regulations. For instance, according to one participant, in some countries an

MVP is not liable to the same legal constraints related to certification as a final product, which can allow the company to put their products in the hands of customers early.

4.3 Evolution of Experiment Objects

We now turn to analysing reported experiences related to how experiment objects evolve over time. The participants gave examples of experiment objects being modified according to customer feedback or kept as is or thrown away after customer feedback. Many stated that development continues according to customer feedback received in demo sessions. Such continuations could be on the level of individual features or feature sets, in terms of the full product, a simple change in the UI design, or in terms of core functionality.

Practitioners indicated that they pay special attention to whether a feature is being experimented on in the live product and the feedback is not positive – in this case, they remove the feature. However, some participants admitted that this can sometimes be forgotten. In fact, during the interviews, some of them remembered features they forgot to remove and said that they would do it after the interview.

When speaking about MVPs, participants indicated that these were intended to be developed into real products.

> "Hopefully, MVPs are becoming products. [. . .] We start promoting the product in the MVP stage so it is very important. In the PoC part, we store them. We move forward with the MVP if we really believe in that solution [explored in the PoC]." (Managing director, Company D)

Participants had slightly different views when they were explaining the development processes and experiment object choices in different development stages. Some companies seemed to be very clear on which type of experiment object to use and whether a particular type of object is needed at all in their case. Accordingly, some participants viewed some of the experiment object types as development stages.

> "If the idea is new to the market, [. . .] we need to build a PoC. [If not], and if we want to visualise how that product should function, then we create a prototype. [If not], and if we want to test the market, then we create an MVP." (Senior software engineer, Company C)

However, this process seems to be more abstract in the other participating companies where it mostly relies on the personal experience of team members in addition to the particular project and customer situation.

The senior technology architect from Company F also discussed the experiment objects across development phases beginning with prototyping, PoC and then MVP, indicating a progression in scale, scope and complexity of types of experiment objects. Not all these objects are needed in all projects, according to the architect. The company and the practitioners would choose which one or ones to pick according to the specifics of the project.

5 Discussion

The results indicate that the case companies usually use experimentation in a broad sense, and occasionally do more focused, controlled experiments. Some participants viewed experiment objects in terms of clearly defined development stages, with different terms being used for objects of different scale, scope, and complexity, while others did not consider them in terms of the evolution of the product. Next, we turn to address the research questions and discuss the results in more detail in relation to them.

5.1 Choosing Experiment Objects

RQ1: *What kinds of experiment objects do practitioners identify and utilise for their experiments?*

Our participants mentioned four kinds of experiment objects: prototypes, MVPs, PoCs, and wireframes, as shown in Table 3. The interviews revealed that multiple factors are considered when companies choose between different kinds of experiment objects as shown in Table 4. This decision appears to be a collaborative and customer-oriented process, where both customers and practitioners bring several constraints and desires that lead to choosing certain experiment objects.

The choice of experiment object depends on the specifics of the project. However, abstracting from the details of the cases, we can observe that practitioners considered PoCs in relation to tailor-made or customer-specific products and when the purpose was to test technical feasibility, while MVPs were associated with mass-market products and the purpose was to test product ideas in the market. Prototypes and wireframes were less clearly connected to this continuum. The interviews also showed that outward visibility was considered. Experiment objects were sometimes kept secret in internal experimentation and sometimes allowed to be visible to customers or the public in external experimentation. Thus, one consideration is how trade secrets or the company's product strategy can be protected while designing an experiment object.

The notion of a "right customer" was mentioned as a critical consideration especially for startups: practitioners emphasised the importance of finding customers from whom to gather feedback and who will lead them in the right direction in the project development while tolerating mistakes in the initial phases. This bidirectional, actively communicating relationship stands in stark difference to the relatively remote customers whose behaviour is observed from afar which is usually considered in later stages of CE [2, 14], and indicates that early-stage experimentation may be more collaborative and interactive than what is often assumed or reported in academic CE literature. If this is true more generally, it raises questions about how to develop experiment objects that are amenable to very interactive experimentation while still allowing for a systematic experimentation process where reliable and generalisable causal relationships between product changes and customer behaviour can be established.

We observe that the participants mentioned varied ways of conducting experimentation and choosing and naming experiment objects. To some extent, this may be due to allowing a broad definition of CE in our study, which goes beyond the strictly controlled online experiments which have popularised the practice. However, uncovering such differences was a purpose of the study. The variation we found may reflect the reality of a broader range of companies than those typically studied, echoing earlier findings on the viability of CE for a wide audience of organisations [26]. Others have found similar variation in the terminology (e.g., [13, 24, 39]), and Melegati et al. [29] have even argued that the term MVP should be replaced since it turns startups' attention to quickly releasing a product rather than experimenting and learning.

We further notice that only few companies in our study reported a systematic approach to choosing experiment objects. To expand on the overall factors shown in Table 4, some companies map object types to product development stages, others consider the fidelity of the object, the business value of the experiment, the effort to prepare the object, and available resources, while yet others seem to pick whatever seems to be most suitable in the moment. The lack of systematic approaches may mean that companies are currently navigating these choices through their specific innovation cultures – the attitudes, behaviours, practices, and other forces driving value creation in an organisation [8, 12, 28, 42].

Our findings illustrate the real-life complexities of designing and managing experiment objects. The academic literature lacks an understanding of the factors that influence experiment object design and management decisions in practice. Our exploration also points towards a need for further understanding of what outcomes can be expected when choosing a specific kind of experiment object in a specific experimentation context. Most CE models do not provide details on these topics and from that perspective, there is a risk that practitioners may choose to create experiment objects that do not function as intended in the experiment design or that waste resources because they require large effort to create, such as the MVPs mentioned in Section 4.1.2, but are then thrown away.

5.2 Evolution of Experiment Objects

RQ2: *How do experiment objects evolve as the continuous experimentation process proceeds?*

Our results indicate that the evolution of experiment objects is influenced heavily by customer feedback and iterative development practices (see Section 4.3). As the experimentation process proceeds, experiment objects can be (1) thrown away, (2) modified and used in another experiment, or (3) become products. The participants underlined that throwing away the experiment object is the last resort and remarked that an experimented feature or feature set should be removed if unsuccessful, especially if the experiment was done on a live product. However, they also acknowledged that occasionally this is forgotten.

Participants considered that the ultimate objective for MVPs is to deliver them as real products. In other words, the next stage for a successful MVP is to become a product. This is less straightforward for PoCs, which may or may not evolve into products. If the PoC was made explicitly for a tailor-made solution, then it seems more likely to be delivered to customers. However, depending on how the company views their product development stages, PoCs can evolve to be MVPs first. Our findings regarding MVPs stand in contrast to those of Terho et al. [39], which found that MVPs will be used again if the experiment has a favourable outcome. We found no instances of MVPs being used again in an experiment; they can instead become products if the outcome is favourable.

We propose that experiment objects can be viewed on a continuum in terms of their fidelity level. Practitioners seem to perceive MVPs as closest to the final product whereas wireframes and prototypes are further away. The lowest-fidelity experiment object mentioned was a Figma-based wireframe. The fidelity of a PoC can vary widely depending on its purpose. PoCs can also be made in response to customer requests. The perception of fidelity between these experiment objects overlapped, most significantly so that prototypes and wireframes were perceived as one overlapping pair, and MVPs and PoCs as another.

Reflecting on our analysis, we noticed that the case companies did not have systematic routines in place to manage experiment objects: how to reuse them either as such in new experiments, or evolve them from one type to another; how to separate the object from the real product in case of experiments on live products; and how to store the objects for potential future uses. Still, as can be seen in some of the remarks in Section 4, they do consider some management questions in isolation, for example that MVPs must be subjected to the full software development process including quality assurance, because it is intended to become a full product. However, these considerations on the companies' part do not seem to fully articulate how to determine when it is worth spending development resources on a fully verified and validated experiment object and when a light-weight, throwaway version would be enough.

Practitioners did not report on substantial activities to make sure that experiment objects can be retrieved again in case they would be needed for something, such as re-validating a conclusion that was obtained through a prior experiment. It seems that many of the experiment objects discussed in the interviews are transient in the sense that the environment required to re-enact the experiment is very hard to recreate, not to mention that there would be any mechanism by which an experiment object can be reapplied onto a future version of the product. It seems that if an experiment object is needed again, it has to be largely or completely re-implemented onto the current version of the product.

5.3 Implications and Recommendations

Although the use of experiment objects seems widespread in the case companies, their design and use relies on the intuition and experience of individuals, and they are understood in varied ways. To make the creation and use of experiment objects more systematic, we believe that conceptual clarity is needed for terms such as MVP and PoC. The envisioned purpose and utility of such experiment objects should be explained, articulating the desired outcomes of creating the object and providing a rationale for why to create one. Our participants reported on their own local rationale for their decisions. Collecting evidence of the outcomes was outside the scope of our study and should be investigated further.

Companies must carefully consider whether their goal is to test the technical aspects of a product-related concept or to assess its potential in the marketplace while deciding whether to use a particular kind of experiment object, such as those exemplified by the simpler concept of PoC or the more close-to-the-final-product concept of MVP. Additionally, it can be important to consider if the product is intended as a tailor-made solution for a single customer or as a mass-market product.

Experiment objects should be viewed as special kinds of artefact, each with its own life-cycle. Companies could benefit from guidance for the choice, design, use, and management of experiment objects. We hypothesise that experiment objects should be chosen based on what fidelity is required to learn within the company's means, i.e., to gain high-quality information for product decision-making at the respective product development stage, taking into account the available resources. We believe that it would be beneficial for companies to consider these factors when they choose between experiment object types. Companies should also consider experiment objects as first-class software artefacts whose lifecycle needs to be managed, potentially in slightly different ways than for other software artefacts such as the product's main source code. Based on our results, there is, for example, a need for feature removal mechanisms to make sure that unsuccessful experiments are removed from the active code base in cases of live product experimentation, but also for the possibility of greater reuse of experiment objects both in new and replicated experiments.

5.4 Limitations

This study is limited in terms of generalisability, given the limited number of participants and companies. It could be strengthened and the generalisability enhanced by investigating additional companies. However, the goal was exploratory, and thus we consider internal validity to be a more important criterion. We also note that the logic of an exploratory case study does not rely on the number of cases [19] and that we chose cases to give breadth to the exploration (see Section 3.2). The participants were chosen among a variety of roles and company domains and use the described approaches on a regular basis. We therefore consider the trustworthiness of their accounts to be high.

Given the nature of qualitative research, researcher and participant biases may have influenced the results. We aimed to provide details on the study design that makes it possible to understand the perspectives of the researchers. We based the interview design on the theoretical background provided in Section 2. The interview questions were open and the interviewer verified interpretations with participants directly during the interview. Participants were encouraged to elaborate on their answers to ensure that their perspectives were well represented in the data. During analysis, we sought alternative explanations in multiple analysis iterations, ensuring that results were corroborated by multiple participants. Thus, our conclusions are grounded in the data.

Social desirability bias may cause participants to answer according to their beliefs of expected answers. We structured the interview questions to avoid leading and to encourage participants to describe their experiences in their own words. Their answers were candid and rich with real examples, indicating that they reported in accordance with their own knowledge.

The fact that the experiment object concept was understood differently by participants may be interpreted as a construct validity concern. However, understanding the different kinds of artefacts they use as experiment objects and the meaning of terms used to denote these artefacts was one of the goals of the study. Moreover, the interview protocol was designed not to mention the term initially, but elicit the terminology the participants preferred.

6 Conclusions

We investigated the choice and evolution of different types of experiment objects for CE in eight companies. We conducted an exploratory multiple-case study using semi-structured interviews to uncover how the companies choose between experiment object types and how the objects evolve over the experimentation process.

The findings highlight the frequent use of experimentation and experiment objects in the case companies. Our participants reported using MVPs, PoCs, prototypes, and wireframes as experiment objects, but it is evident that there are different understandings of the concepts between the literature and the industry, and within different companies and practitioners in the industry as well. This warrants further investigation, since terminological and conceptual clarity would benefit both researchers and practitioners.

We found that choosing a particular kind of experiment object (RQ1) is a collaborative and customer-oriented process that depends on the specifics of the project. Several factors come into play, including the customer's desires, technological feasibility, business value of the experiment, estimated effort for experiment object development, resource availability, and market competition. We did not find a systematic selection process. Rather, the selection as described by the participants happened through the means typical for each company's innovation culture.

As the experimentation process proceeds, experiment objects can evolve (RQ2) in a variety of ways as they can be modified and utilised in other experiments, thrown away, or turned into real products. Participants did not report a systematic process for managing experiment objects, but objects denoted as MVPs were most often seen as destined for release as products, while objects denoted by other terms were seen as more preliminary. Companies differed in how they viewed the mapping between different experiment object types and stages of product development.

Our work opens several avenues for further research. We did not investigate how to take into account fidelity in the choice and design of experiment objects. Advances in this area could lead to better guidelines for creating experiment objects for specific experimentation purposes. The interaction between experimenter, experiment object, and customer would also warrant further research. Especially for startups, identifying the right customers to gather feedback and data from is crucial to ensure that it can support product decision-making but also that customers are tolerant towards the experimentation process. Not everyone wants their software to constantly change, and there are justifiable concerns with being constantly experimented on. We suggest that future work could explore how different kinds of experiment objects can be selected and combined for different experimentation purposes in different domains, and how experiment objects can be systematically managed over their life-cycle.

Acknowledgments

We extend our sincere thanks to all the participants who were so generous with their time, knowledge and experience. This work was supported by the Research Council of Finland [project 349574].

References

[1] Silvio Alonso, Marcos Kalinowski, Bruna Ferreira, Simone D.J. Barbosa, and Hélio Lopes. 2023. A systematic mapping study and practitioner insights on the use of software engineering practices to develop MVPs. *Information and Software Technology* 156 (2023), 107144.

[2] Florian Auer, Rasmus Ros, Lukas Kaltenbrunner, Per Runeson, and Michael Felderer. 2021. Controlled experimentation in continuous experimentation: Knowledge and challenges. *Information and Software Technology* 134 (2021), 106551.

[3] Elizabeth Bjarnason, Franz Lang, and Alexander Mjöberg. 2023. An empirically based model of software prototyping: a mapping study and a multi-case study. *Empirical software engineering* 28, 115 (2023).

[4] Steve Blank. 2013. Why the Lean Start-Up Changes Everything. *Harvard Business Review* (May 2013). https://hbr.org/2013/05/why-the-lean-start-up-changes-everything Accessed: 2025-01-17.

[5] Steve Blank and Shawn Carolan. 2021. A Path to the Minimum Viable Product. (2021). https://steveblank.com/2021/04/20/the-secret-to-the-minimum-viable-product Accessed: 2025-01-17.

[6] Petra Bosch-Sijtsema and Jan Bosch. 2015. User involvement throughout the innovation process in high-tech industries. *Journal of Product Innovation Management* 32, 5 (2015), 793–807.

[7] Svend Brinkmann and Steinar Kvale. 2018. *Doing interviews.* SAGE Publications Ltd.

[8] Sheldon A. Buckler. 1997. The spiritual nature of innovation. *Research Technology Management* 40, 2 (1997), 43–47.

[9] Euan Ross Coutts, Andrew Wodehouse, and Jason Robertson. 2019. A Comparison of Contemporary Prototyping Methods. *Proceedings of the Design Society: International Conference on Engineering Design* 1, 1 (2019), 1313–1322.

[10] Adrien Coyette, Suzanne Kieffer, and Jean Vanderdonckt. 2007. Multi-fidelity prototyping of user interfaces. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT '07).* Springer, 150–164.

[11] John W. Creswell and Cheryl N. Poth. 2018. *Qualitative inquiry & research design: choosing among five approaches* (4 ed.). SAGE Publications.

[12] C. Brooke. Dobni. 2008. Measuring innovation culture in organizations: The development of a generalized innovation culture construct using exploratory factor analysis. *European Journal of Innovation Management* 11, 4 (2008), 539–559.

[13] Anh Nguyen Duc and Pekka Abrahamsson. 2016. Minimum viable product or multiple facet product? The role of MVP in software startups. In *Agile Processes, in Software Engineering, and Extreme Programming: 17th International Conference, XP 2016.* Springer, 118–130.

[14] Fabian Fagerholm, Alejandro Sanchez Guinea, Hanna Mäenpää, and Jürgen Münch. 2017. The RIGHT model for Continuous Experimentation. *Journal of Systems and Software* 123 (2017), 292–305.

[15] Dror G. Feitelson, Eitan Frachtenberg, and Kent L. Beck. 2013. Development and Deployment at Facebook. *IEEE Internet Computing* 17, 4 (2013), 8–17.

[16] Arlene Fink. 2003. *The survey handbook* (2nd ed.). SAGE, Thousand Oaks, Calif.

[17] Brian Fitzgerald and Klaas-Jan Stol. 2017. Continuous software engineering: A roadmap and agenda. *Journal of Systems and Software* 123 (2017), 176–189.

[18] Christiane Floyd. 1984. A systematic look at prototyping. In *Approaches to prototyping.* Springer, 1–18.

[19] Bent Flyvbjerg. 2006. Five misunderstandings about case-study research. *Qualitative inquiry* 12, 2 (2006), 219–245.

[20] Stephen J. Gentles, Cathy Charles, Jenny Ploeg, and K. Ann McKibbon. 2015. Sampling in Qualitative Research: Insights from an Overview of the Methods Literature. *The Qualitative Report* 20, 11 (2015), 1772–1789.

[21] Helena Holmström Olsson and Jan Bosch. 2014. The HYPEX Model: From Opinions to Data-Driven Software Development. In *Continuous software engineering.* Springer, 155–164.

[22] Helena Holmström Olsson and Jan Bosch. 2015. Towards Continuous Customer Validation: A Conceptual Model for Combining Qualitative Customer Feedback with Quantitative Customer Observation. In *Software Business: 6th International Conference, ICSOB 2015.* Springer, 154–166.

[23] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. 2013. Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 1168–1176.

[24] Valentina Lenarduzzi and Davide Taibi. 2016. MVP Explained: A Systematic Mapping Study on the Definitions of Minimal Viable Product. In *2016 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA).* 112–119.

[25] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Trans. Comput.-Hum. Interact.* 15, 2 (2008).

[26] Vihtori Mäntylä, Bettina Lehtelä, and Fabian Fagerholm. 2022. The Viability of Continuous Experimentation in Early-Stage Software Startups: A Descriptive Multiple-Case Study. In *International Conference on Product-Focused Software Process Improvement.* Springer, 141–156.

[27] Gerardo Matturro, Gonzalo Nieto, Alfonso González, and Martín Solari. 2021. Minimum Viable Product Creation and Validation in Software Startups. In *2021 XLVII Latin American Computing Conference (CLEI).* 1–8.

[28] Patrick McLaughlin, John Bessant, and Palie Smart. 2008. Developing an organisation culture to facilitate radical innovation. *International Journal of Technology Management* 44, 3-4 (2008), 298–323.

[29] Jorge Melegati, Rafael Chanin, Afonso Sales, Rafael Prikladnicki, and Xiaofeng Wang. 2020. MVP and experimentation in software startups: a qualitative survey. In *46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA).* 322–325.

[30] Sharan B. Merriam and Elizabeth J. Tisdell. 2015. *Qualitative research: A guide to design and implementation.* John Wiley & Sons.

[31] Jürgen Münch, Fabian Fagerholm, Patrik Johnson, Janne Pirttilahti, Juha Torkkel, and Janne Järvinen. 2013. Creating Minimum Viable Products in Industry-Academia Collaborations. In *Lean Enterprise Software and Systems: 4th International Conference, LESS 2013.* Springer, 137–151.

[32] Jennifer N. Petrie and Kevin A. Schneider. 2007. Mixed-fidelity prototyping of user interfaces. In *Interactive Systems. Design, Specification, and Verification: 13th International Workshop, DSVIS 2006.* Springer, 199–212.

[33] Konstantinos Poulis, Efthimios Poulis, and Emmanuella Plakoyiannaki. 2013. The role of context in case study selection: An international business perspective. *International Business Review* 22, 1 (2013), 304–314.

[34] Eric Ries. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses.* Crown Publishing Group.

[35] Jim Rudd, Ken Stern, and Scott Isensee. 1996. Low vs. high-fidelity prototyping debate. *Interactions* 3, 1 (1996), 76–85.

[36] Per Runeson and Martin Höst. 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering* 14 (2009), 131–164.

[37] Robert E. Stake. 1995. *The art of case study research.* SAGE.

[38] Robert E. Stake. 2006. *Multiple case study analysis.* The Guilford Press.

[39] Henri Terho, Sampo Suonsyrjä, and Kari Systä. 2016. The developers dilemma: Perfect product development or fast business validation?. In *Product-Focused Software Process Improvement: 17th International Conference, PROFES 2016.* Springer, 571–579.

[40] Tom S. Tullis. 1990. High-fidelity prototyping throughout the design process. In *Proceedings of the Human Factors Society 34th Annual Meeting.* Human Factors Society, 266.

[41] Miriam Walker, Leila Takayama, and James A. Landay. 2002. High-Fidelity or Low-Fidelity, Paper or Computer? Choosing Attributes when Testing Web Prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 46, 5 (2002), 661–665.

[42] Xuemei Xie, Yonghui Wu, Peihong Xie, Xiaoyu Yu, and Hongwei Wang. 2021. Organizational innovation culture and firms' new product performance in two emerging markets: The moderating effects of institutional environments and organizational cohesion. *Journal of Management & Organization* 27, 5 (2021), 972–991.

[43] Robert K. Yin. 2012. *Applications of Case Study Research.* SAGE Publications.