

Aalto University
School of Science
Master's Programme in Computer, Communication and Information Sciences

Juho Pekkinen

Applying Lean Startup Methodology to Develop an Electronic Examination System

Master's Thesis
Espoo, May 13th, 2019

Supervisor: Professor Marjo Kauppinen
Advisor: Ari Korhonen D.Sc. (Tech.)

Author:	Juho Pekkinen	
Title:	Applying Lean Startup Methodology to Develop an Electronic Examination System	
Date:	May 13th, 2019	Pages: viii + 58
Major:	Computer Science	Code: SCI3042
Supervisor:	Professor Marjo Kauppinen	
Advisor:	Ari Korhonen D.Sc. (Tech.)	
<p>The digitalization has radically transformed higher education by enabling new learning activities for example on the form of online courses. Programming and simulation assignments have been corner stones of the education for a long time in the field of computer science. However these new teaching methods have not affected exam and to this day the most of the exams in universities are taken in a traditional way, with pen and paper, sitting in a lecture hall.</p> <p>The goal of this thesis was to study how lean startup methodology can support the development of electronic examination system. Action research approach was utilised in this study. In the current state analysis phase, four teachers were interviewed to understand their process of organising exams. Lean startup methods, tried in this thesis, were setting hypothesis and Minimum Viable Product (MVP). Additionally workshops were organised to elicit input for presented MVPs from teachers. Finally the experience from applied methods were reported.</p> <p>The MVP was useful throughout the study to communicate complex ideas and to elicit feedback. Setting hypotheses helped when building the second MVP to recognise key assumptions from technical requirements and thus helped to focus on the most important topics. Based on the results of the study the solution for electronic examination system is the exam taken in a lecture hall with student owned devices. The system can be built on four distinct components: students user interface, possibly implemented on top of A+ Learning Management System (LMS), grading interface, interface for teachers to construct exams and OS and other infrastructure to run on laptops.</p> <p>The results of the study indicate that lean startup methods can be useful in designing electronic examination system. The iterative process and used practices helped in conceptualising and evaluating the complex domain area.</p>		
Keywords:	lean startup, electronic examination, computer science education, action research	
Language:	English	

Tekijä:	Juho Pekkinen		
Työn nimi:	Lean startup -metodologian soveltaminen sähköisen tenttijärjestelmän kehittämässä		
Päiväys:	13. toukokuuta 2019	Sivumäärä:	viii + 58
Pääaine:	Tietotekniikka	Koodi:	SCI3042
Valvoja:	Professori Marjo Kauppinen		
Ohjaaja:	Tohtori Ari Korhonen		
<p>Digitalisaatio on mullistanut korkeakoulutuksen mahdollistamalla täysin uudenlaisia opintosisältöjä muun muassa verkkokurssien muodossa. Tietotekniikan alalla erilaiset ohjelmointiharjoitukset ja simulaatiot ovat pitkään olleet opetuksen kulmakiviä. Kuitenkin nämä uudet opetusmuodot eivät ole vaikuttaneet tentteihin ja suurin osa yliopistotenteistä suoritetaan perinteisin menetelmin, kynällä ja paperilla, kaikki yhtä aikaa luentosalissa istuen.</p> <p>Tämän diplomityön tarkoituksena oli tutkia kuinka lean startup metodologiaa voisi soveltaa sähköisen tenttijärjestelmän kehittämiseen. Toimintatutkimuksena toteutetussa tutkimuksessa aluksi haastateltiin opettajia ja muodostettiin kuva kuinka he järjestävät tenttejä. Interventio-vaiheessa sovellettiin lean startupista omaksuttujen käytänteiden, hypoteesien asettamista sekä pienintä mahdollista tuotetta (MVP). Lisäksi työpajatyöskentelyä hyödynnettiin tiedon keräämisessä. Tutkimuksen jälkeen tärkeimmät opit listattiin.</p> <p>Lean startup menetelmistä MVP auttoi ongelman jäsentämisessä ja viestimässä sekä palautteen keräämisessä kohderyhmältä. Hypoteesien asettaminen auttoi järjestelmävaatimusten muotoilussa nimenomaan käyttäjien näkökulmasta sekä keskittymään kaikista keskeisimpiin aiheisiin. Tutkimuksen perusteella luentosalissa järjestetty tentti, jossa opiskelijat käyttävät omia tietokoneitaan on yksinkertaisin tapa tuoda nykyaikaiset oppimismenetelmät tenttitilanteeseen. Eri käyttäjäryhmät huomioon ottaen tenttijärjestelmä koostuu neljästä erillisestä alijärjestelmästä: oppilaan näkymä tenttiä tehdessä, tentinlaatumistyökalut, tarkastuksenhallintajärjestelmä sekä laitteisto, käyttöjärjestelmä ja infrastruktuuri.</p> <p>Tutkimuksen tulokset viittaavat, että lean startupin menetelmät voivat auttaa sähköisen tenttijärjestelmän suunnittelussa. Jatkuva ja vähittäinen kehitysmalli auttoi jäsentämään ja arvioimaan monitahoiselle sovellusalalle suunnattua ratkaisua.</p>			
Asiasanat:	lean startup, sähköinen tenttiminen, tietotekniikan opetus, toimintatutkimus		
Kieli:	Englanti		

Acknowledgements

First I would like to thank my supervisor, professor Marjo Kauppinen for her relentless support throughout this thesis. I am grateful for support she provided with the fresh and unconventional research perspective I had chosen. Especially her expertise with the research method was valuable for whole the thesis process.

I would like to thank also my instructor, university lecturer Ari Korhonen for providing me with interesting and ambitious research topic and for guiding me through the vast literature bodies of computer science and online education.

The university education has indeed been a journey for me and I am grateful for my parents on their unlimited support throughout my studies. During my time studying, I have learned a lot also outside the classroom. Being part of this awesome community makes me proud and I want to thank Sähköinsinöörikieltilta, Joutomiehet, Otaniemen jalostusseura and Kampusjaosto for all crazy projects that I have had a privilege to part of. Along this journey I have made friends with many magnificent people and there are far too many to name them all. Thank you to all of you.

I am truly grateful for my fiancée Jette for believing in me the whole time even when I did not believe in me myself. You really have made this possible. Last I thank you Reima, my tail-wiggling Australian terrier. Your never-ending playfulness and unconditional love was able to lift my mood anytime I was feeling down.

Helsinki, May 13th, 2019

Juho Pekkinen

Abbreviations and Acronyms

HTML	Hyper Text Markup Language. A script language used to define web pages.
LMS	Learning Management System. A software system that administrates and delivers educational resources and tools.
MVP	Minimum Viable Product. A functional version of the product that implements features required to test the hypothesis. A fundamental part of the lean startup methodology. A close synonym to a prototype.
MCQ	Multiple Choice Questions
MOOC	Massive Open Online Course. A course that utilises online learning activities to provided education for anyone with the internet access.
OS	Operating System. The low-level software that supports a computer's basic functions such as scheduling tasks and controlling peripherals.
RAM	Ramdom Access Memory.
RE	Requirements Engineering. One of the four fundamental activities of SE. Iterative process of eliciting, analysing, presenting and validating user and business needs for the software.
RST	reStructured Text. A file format for textual data. Used in A+ LMS for describing how documents are to be compiled into HTML web pages.
SE	Software Engineering. An engineering discipline that is concerned with all aspects of the software production.
TUT	Tampere University of Technology. A polytechnic university in Finland that as of 1.1.2019 has been incorporated into the new Tampere University

UCD	User-Centered Design. A framework of processes to address usability, user characteristics, environment, tasks and workflow when designing products or services.
USB	Universal Serial Bus. Connector type and protocol to connect wide range of peripherals to a PC.
UX	User Experience. Encompasses all aspects of the end-user's interaction with the company, its services and its products.
XP	Extreme programming. A collection of tried agile software development best-practices.

Contents

Abbreviations and Acronyms	v
1 Introduction	1
1.1 Motivation and background	1
1.2 Research problem and questions	2
1.3 Scope of the thesis	3
1.4 Structure of the thesis	3
2 Background	4
2.1 Stakeholders	4
2.2 Motivation for going towards electronic examinations	5
2.3 Current solutions	6
2.3.1 WebToTest	6
2.3.2 ViLLe	7
2.3.3 EXAM Consortium	7
2.3.4 Tentti Plussa	8
2.3.5 Abitti	9
2.4 Criteria for electronic examination systems	9
2.4.1 Automatic marking	10
2.4.2 Supervision of exams	11
2.4.3 Online learning	13
2.4.4 Exam content and creating exams	15
2.4.5 Integration	16
2.5 Analysis of existing systems	18
2.5.1 Summary	21
3 Research methods	22
3.1 Literature review	22
3.2 Empirical study	22
3.2.1 Research Process	22
3.2.2 Data Collection and Analysis	23

4	Results of the literature review	25
4.1	Lean startup	25
4.1.1	Overview of the Lean Startup	25
4.1.2	Principles of the Lean Startup	27
4.1.3	Practices of the Lean Startup	28
4.1.4	Studies about applying Lean Startup	30
4.2	The applied process model	31
5	Results of the empiric study	33
5.1	Current State	33
5.2	Action Intervention	35
5.3	Reflective Learning	42
5.3.1	Setting hypothesis	42
5.3.2	MVP	42
5.3.3	Workshops	43
6	Discussion	45
6.1	Requirements for the examination system	45
6.1.1	Role of the exam in universities	46
6.1.2	Electronic examination in regards with online learning	46
6.1.3	Subsystems of the electronic examination system	47
6.2	Good practices of the lean startup	48
6.3	The applicability of lean startup methods	49
6.4	Limitations of the study	50
7	Conclusions	51
7.1	Future work	51
A	Manuscript of the interview for current state analysis -phase (in Finnish)	57

Chapter 1

Introduction

This thesis studies existing electronic examination systems and teachers' work practices when arranging exams and tries to find match between these two. As a result, list of user requirements for examination system is produced and it is discussed how those can be achieved by the use of existing solutions.

1.1 Motivation and background

Universities are constantly developing new ways to utilise digitalization in education. For example Aalto University has a university wide initiative, Aalto Online Learning, that gathers together various pilots developing new ideas for online education [Aalto University, 2019]. While digital solutions provide ways to give more precise and timely feedback, established process especially related to organising exams, have clear merits. So it is the goal of this thesis to understand what kind of needs teachers have for examination system and how electronic solutions could meet those needs. Eventually a goal from teachers' perspective could be that assessment and control are integrated to teaching activities in such a way that examinations as separate activities can be discontinued. From the students point of view digitalization aims to develop education and assessment so that activities would not be detached from the application domain context.

The literature is somewhat divided on the terminology and terms like electronic, on-line [Arnow and Barshay, 1999] and computer aided examinations are used seemingly as synonyms for each other. In this thesis we suppose this to be true and use electronic examination to cover all these variations because the thesis is not focused on just online education context.

The Lean Startup is new, practically oriented approach to address customer and user needs when developing novel and radically innovative prod-

ucts and businesses drawing elements from Customer Development presented by Blank [2005] and from agile software development methods. It contains tools and practices to manage both new product development within a development team, and the business venture as a whole. As an emerging and business originated approach, there seems to be little critically reviewed information available about its effectiveness. The approach has some success stories that are reported mainly by author of the methodology. According to Ries [2011], the Lean Startup is about breaking business idea down to hypothesis about customer or user behaviour and then rigorously testing those hypothesis. The testing is advised to be conducted in iterative build-measure-learn loops using what Ries calls minimum viable product (MVP) [Ries, 2011]. In the essence the MVP can be of any form as long it allows developers to conduct the experimentation. It can be anything from low quality prototype to fully featured product but important aspect is that development is done continuously from former to latter and there are no milestones for beta testing or official release.

Ries [2011] presents several cases of successful new ventures using the Lean Startup but his claims of the applicability of the method in large companies are not supported as much. For this reason the study focuses on to research scientific literature and possible case studies to assess how Lean Startup is applied in large, established organizations. Also concept of internal startup is taken to focus of this study. In further detail, the focus is on the development of digital services and products from the perspective of the software engineering discipline.

In his article Blank [2013] calls lean startup as a methodology and that term will also be used in this research.

1.2 Research problem and questions

This thesis aims to understand **how can lean startup methodology support the development of electronic examination system**. Related research questions are listed below.

- RQ1:What kind of electronic examination systems suits the needs of the computer science education?
- RQ2:Which lean startup practices can be applied for the development of an electronic examination system?
- RQ3:What lessons can be learned from applying lean startup practices for electronic examination systems?

1.3 Scope of the thesis

The thesis is restricted to study electronic examination from the perspective of teachers and students. The scope is also restricted to target the education of computer science because the programming as an exercise is quite different from essay and multiple choice question (MCQ) questions. Also computer science has quite drastic contrast between environment where course assignments are done and how exam is taken. In course there are automatic marking [Ala-Mutka, 2005] and instant feedback available while programming with pen and paper is inherently very different to how it is done in practice.

Aalto Learning and Technology research group has conducted extensive research on topics such as automatic marking and has developed their own A+ LMS especially for computer science education [Karavirta et al., 2013], [A+ LMS, 2019]. Because we want to utilise this knowledge and prior work, we are interested in solutions that could use A+ as a basis for implementing solution.

1.4 Structure of the thesis

The rest of the thesis is built as following: in Chapter 2 the background for electronic examinations is presented. Chapter 3 introduces applied research methods. The Chapter 4 presents literature review of lean startup methodology. Results of the study are presented in the Chapter 5 and implications and limitations of the study are discussed in Chapter 6. Finally Chapter 7 draws conclusions of the study. The Table 1.1 presents where answer to each research question can be found.

Research question	Literature review	Empirical study
RQ1	2.4	5.2
RQ2	4.3	
RQ3		5.3

Table 1.1: Structure of the thesis.

Chapter 2

Background

In this chapter we present current state of the electronic examination domain. Especially the focus are objectives for electronic examination systems and criteria to assess the completion of these objectives. Later in the chapter current solutions that are in use or have been tested are analysed through the criteria. Due to scarce amount of publications, the analysed systems bias heavily towards those used in Finnish universities merely due to the fact that information of systems used there was available via shared projects and initiatives.

2.1 Stakeholders

Organising exams involves several stakeholders in Aalto University some that are integral part of the process and some others who have only limited interest in exams. In this section these stakeholders are introduced and scope is refined from the perspective of the stakeholders. The stakeholder groups are picked from book chapter by Korhonen and Multisilta [2017]. The stakeholders that they list are teachers, students, institution, researchers and developers. Additional stakeholder groups that needs to be addressed regarding electronic examinations are for example teaching assistants and lawyers.

The first one of two most important stakeholders, teachers are responsible designing and grading exams. Their main interests are how much effort the process takes and how well exams measure students' skills and knowledge. In addition they are concerned about students plagiarising or collaborating extensively and exams are a way to keep this in check. Teachers are the main stakeholder group for this thesis.

Students are the other one of the most important stakeholders. Relating to exams they are concerned with whether they can know in advance what

to expect in the exam and how well their study efforts will prepare them to solve exam questions. When considering electronic exams students are additionally concerned with the usability and familiarity of the system and tools provided. This thesis will focus also on students as stakeholders.

Teaching assistants may be involved in grading the exams especially with the most largest courses. They are not on the main focus of the thesis but their involvement is kept in consideration during the research.

As of current developer or admins do not have big role in organising exams but their impact would increase if development goes towards electronic examination systems. Their interests are likely related to reliability, maintainability and interoperability of the said systems on their own and in cooperation with existing systems.

In the context of the examinations, institution stakeholders are personnel that provide general services for teachers and students. This group is responsible for practicalities of exams like reserving adequate lecture rooms, collecting, copying and distributing exam questions sheets as well as recruiting and instructing invigilators. Their role in the process of organising exams is not on the focus of the thesis.

Lawyers are by no means key stakeholders when exams are organised and they are not in the scope of this study. They still provide certain guidelines to follow when considering equality and fairness. Additionally considering electronic system lawyers are likely to have some requirements regarding data protection and aforementioned aspects.

Of all stakeholders listed above, teachers are in the main focus of this study. Additionally students are considered in the later parts of the study.

2.2 Motivation for going towards electronic examinations

For teachers electronic examinations allow several benefits. Medley [1998] argue that practical approach of electronic examinations may be better indicator of achieving goals of introductory computer science course, to be able to write functioning programs. Additionally Mason and Voit [1998] note that because of being more similar with lab assignments, electronic examinations can show whether students have done all assignments. They also noted that computer science classes usually include both theoretical and practical content and they might need different testing approaches. The experiences of using electronic content in computer science courses have also shown that the workload of the teacher shifts so that most effort is required before course to

make assignments and graders but on the other hand time required to mark assignments during the course is reduced. Similar change could be expected in the case of electronic exams also.

According to Medley [1998], students find electronic examinations more fair. Mason and Voit [1998] add that students have higher motivation to study during the course when electronic examination is utilised at the end of the course. Additionally Stowell and Bennett [2010] found that online exam reduces test anxiety for those suffering stress in classroom environment. While their study found the effect to be reversed for those not suffering from anxiety in classrooms, having options available would be good approach because then students can choose best option that suits their study habits and other preferences.

Pedagogical literature classifies evaluation into two categories, formative and summative [Lindblom-Ylänne and Nevgi, 2009]. The former emphasizes interaction between student and teacher to give both parties feedback how they could learn better and also how to support learning better. The latter for of evaluation emphasizes control, equality and conformity to standards by making sure that every student is evaluated by same assignments, against same learning goals and ensuring that learning is demonstrated in standard environment and format.

Other aspect of classifying evaluation is how it is integrated with the application domain of the knowledge being evaluated. An example of detached evaluation is that computer science education is evaluated by output produced with pen and paper while the actual programming is usually done with computer in interaction with compilers and other aiding software.

2.3 Current solutions

In this section we present a selection of systems that we are aware that have been used for electronic examination and discuss their characteristics based on the objectives listed in previous section. First the WebToTest system used by Arnow and Barshay [1999] is discussed as a sort of baseline. After that selection of systems from Finnish Universities are discussed to show the state-of-the-art at least in Finnish education institutions.

2.3.1 WebToTest

In their study Arnow and Barshay [1999] analysed an electronic examination system that they had developed on top of their own LMS for computer science education, WebToTeach that is currently a commercial project under

a name of CodeLab[Arnow, 1999]. WebTotest was a system that was used via standard internet browser in the computer lab. The main view of the system presents form-based view on the exam and it is accompanied by the Language Workshop feature where students were able to write, compile and execute their own source code that was after copied to answer form.

Teachers prepare exams through a form-based interface. Automatic tests for programming assignments are done by defining source code blocks that precede and succeed student generated code which are then concatenated into one source code. This source code is run with predefined input and output is compared with the one defined by the teacher. After the examination teachers have options to override automatic marking and to mark assignments that require manual marking using again browser based interface.

2.3.2 ViLLe

Rajala et al. [2016] report about pilot at the university of Turku. Their system utilises multiple kinds of assignments related to programming and all the assignments are assessed automatically [Rajala et al., 2016]. Supported assignment types include traditional programming exercises; robot exercises where students program crane to perform operations and achieve goals as effectively as possible; organising code lines to right order; and connecting right item together. During the exam students were allowed to resubmit their solutions to try getting better grades. It was not discussed in the article how teachers create exams or graders or how security of the system is achieved.

2.3.3 EXAM Consortium

EXAM consortium consists of 27 Finnish universities and universities of applied science that work towards building shared infrastructure for organising electronic examinations [EXAM Consortium, 2019]. The system consists of web service to reserve and manage exam schedule, dedicated computer lab with computer-aided supervision and desktop devices with custom EXAM software for taking the exam [EXAM Consortium, 2018]. The web service allows students to browse all available exam slots offered either by searching for course code or name or via a direct link provided by course instructor. When reserving exam slot, students are assigned a specific device from the lab that they have to use for the test. The physical class room setup is up to the each organising institution but usually there is at least recording video surveillance and in addition possibly some kind of access control mechanism. The smaller EXAM room that has six computers at the Aalto University is presented in Figure 2.1. The desktop devices themselves control access:

only user account associated with student who has reserved exam slot is allowed to login on device. If student tries to log in on wrong device, they are instructed towards right seat in class room. The EXAM software on the computers allow students to only open web browser showing nothing but appropriate exam questions that might be randomly selected from larger pool of questions.



Figure 2.1: Picture of EXAM room at Aalto University.

2.3.4 Tentti Plussa

Tentti Plussa is a system developed at Tampere University of Technology (TUT) that builds on infrastructure provided by the EXAM system. Tentti Plussa also utilises A+ LMS, the same that is used at Aalto University department of computer science. At TUT A+ is branded as "Plussa" to alleviate connection with Aalto University. The exam lab is controlled with the EXAM infrastructure: students use EXAM to reserve test slot and to login to system. TUT uses student card as an additional safety measure and as an access token to exam lab. Students are assigned randomly exam

questions that are in fact secure links to separate exam instance of A+ LMS which is allowed to access only from specified computers in the examination room. Students return their answers to Plussa just like they did with the course exercises.

2.3.5 Abitti

Finnish high schools are using Abitti [2019] for their national matriculation tests. According to the public information material, the system consists of local servers storing exam questions; students' laptops that are booted in safe mode either from USB storage or from network drive; and of secure wired or wireless network dedicated for exam use. The exam system is also used in almost every exam in Finnish high schools and there have been trials to use it in introductory university courses offered for students still in high school. There is quite a little published information about the system, especially in the peer-reviewed sources. However the system should support different questions types because it is used in all subjects ranging from languages to mathematics and natural sciences. One university lecture was interviewed for using Abitti on aforementioned introductory university course. She stated that there was little difficulties, because the system supported only free text answers and not compiling or running students' source code.

2.4 Criteria for electronic examination systems

Arnou and Barshay [1999] presents four aspects how electronic exams can improve computer science education based on work by Mason and Voit [1998] and Medley [1998]. The improvements of education provided by electronic exams as presented in the publication are (1) automatic marking, (2) security, (3) distance learning and (4) flexibility in question variety. These four aspects are extended by Arnou and Barshay [1999] with three additional criteria that are (5) ease of setting up exam questions, (6) integration with existing LMS and (7) facilitating integration with homework problem sets. These seven aspects can act as a basis when formulating evaluation criteria for electronic examination systems. All seven principles are not elaborated in detail but their definitions have to be interpreted by how objectives are met with each system used in studies by Arnou and Barshay [1999], Mason and Voit [1998] and Medley [1998]. Because all three studies are about twenty years old, requirements can also be interpreted through the development of

education technology and especially how some requirements and goals came up in empirical part of this thesis. Also the objectives listed above lack explicit success criteria so we try to fill in the blanks by levying the accumulated knowledge from developing the A+ LMS. Exam construction **(5)** is criteria applicable only for teachers' perspective and as such it is incorporated as part of the preceding criteria. Similarly both criteria **(6)** and **(7)** involve integrating examination system with either existing systems or materials so they are combined into one criteria.

2.4.1 Automatic marking

Automatic marking can be used in many ways to measure different aspects of the student's submission. In her survey focusing on automatic marking of programming exercises the marking is divided into two main categories, dynamic and static assessment that are then divided into more detailed sub-categories based on what features of the program instructors like to test. Ala-Mutka [2005]

Mason and Voit [1998] expects on-line examinations to reduce time required to mark the assignments. Reasons why this would be beneficial are obvious: time saved from rather mechanical tasks means more time for example for interacting with students or developing the course. Automatic marking is one of the key features of the A+ LMS [Karavirta et al., 2013] and so its features are taken as the baseline for automatic marking in this thesis. According to Karavirta et al. [2013] A+ supports automatic marking of virtually any kind of exercises as long as the external grader can receive submission and send grades through standard HTTP connections. Automatic marking can be classified by whether grader is able to identify finite or infinite amount of right solutions. Marking of multiple choice questions or grading student provided code by its output or return value by string matching are examples of graders recognizing finite amount of right solutions. Other option is that grader compares the solution submitted by the student to the one provided by course instructor by giving them same inputs and then for example running unit tests to check the state of the software.

From students point of view the benefit of automatic marking is the immediate feedback it provides and chance to reiterate and resubmit their solution for better grade. A study by Malmi et al. [2005] shows that given a chance, students are willing to resubmit their answer to get highest possible grade. This behaviour can be also be expected in the exam situation leading to larger portion of highest grades. The feedback provided from the automatic marking can be either binary passed or failed or more detailed including partial points or even contain feedback how the solution could be improved.

Also when considering from the perspective of exams, this behaviour can be challenging because of the restricted time available for the exam.

2.4.2 Supervision of exams

For the context of this thesis, the security, or supervision, of exam has two distinct goals. The first is to guarantee that the student authenticated properly and the second is that students use only those resources that instructors have intended them to use during the exam. It is argued that plagiarism increases in transition from classroom education to online environment. Corrigan-Gibbs et al. [2015] found out that around up to 35 percent of students in their sample committed some sort of plagiarism. However the extent of plagiarism overall might vary from institution to other and even between courses. Contrary to study by Corrigan-Gibbs et al. [2015], Chen et al. [2017] found that there was no clear plagiarism in online exam environment similar to the EXAM system, which was hypothesized to be susceptible for collaboration and circulation of exam questions and solutions among students. When considering security, good comparison point is how security is enforced in traditional pen and paper exams. At Aalto University the guidelines for Aalto University [2017] arranging exams focus on two aspects of security: authenticating the student and preventing cheating. For example identity of student is verified by one of the invigilators when turning in the exam from official ID or student card [Aalto University, 2017]. Cheating and communication between students is inhibited for example by not allowing any excess items like bags and especially mobile phones near when exams are in progress and keeping students placed separate enough from each other.

Preventing cheating in online learning is difficult, because there are so many ways to try to gain unfair advantage as the Rowe [2004] demonstrates in his paper. Examples of possible ways to assault or exploit the system include acquiring access to instructor's computer one way or another and so retrieving exam questions before hand or changing grades or own answers after the exam. According to Rowe [2004] another broad category of attack vectors is unauthorized help during the exam. For example students could hire other students to take the exam on their behalf. These issues are not completely unique for online examinations. Also printed exams, student answers and graded exams need to be stored carefully throughout the process and student identity is needs to be checked during the traditional exams as well. Indeed, several of these countermeasures proposed by the Rowe [2004] are included in the examination guidelines of the Aalto University. However the online examinations provide new opportunities to compromise the security of exams because without careful planning the online exam servers can be accessible

from wherever in the internet possible vulnerable to software defects, hacking and other exploits while compromising paper exams requires physical access to the exams.

The countermeasures for plagiarism in online exams and learning activities vary a lot. Rowe [2004] proposes quite heavy measures including having the position of data security manager while Corrigan-Gibbs et al. [2015] suggests that even a sterns warning of consequences getting caught has significant impact on plagiarism rates in online education. Another approach is to just detect cheating after submissions and handle cases afterwards. One such solution is one studied by Simon [2005]. He implemented a solution to watermark exam templates so he could detect if there was forbidden cooperation between student. His solution focused on only one kind of plagiarism and he noted that preventing cheating especially in online education context is constant race between teachers and those who try to cheat. Additionally it is notable that this kind of plagiarism detection system is probably useful for getting a grasp how common unethical practices are among students but it does not necessarily scale well to handle cases in large courses.

Another method of detecting plagiarism was presented by Hellas et al. [2017]. They tracked the distance from the final submission as a function of time and tried to identify uncharacteristic patterns. For example big leaps towards the final submission were indicators that student might have copied the answer from somewhere else. In the study they additionally tried to classify different types of plagiarism and what motivations each kind of student had. The result of the study was that there are three types of unethical behaviour, help-seeking, collaboration and systematic cheating. These three types can have different motivations for cheating and can be told apart from statistical analysis. Based on their study, help-seekers might be struggling with the course and use unethical practices as last resort. Motivations for other types of plagiarism is not that clear based on the research. It might be that because collaboration is encouraged to some extent in programming courses, those who collaborate unethically simply are not aware of what practices are allowed. Systematic cheaters might seek out ways to cheat the system and find loopholes as some kind of practice or pastime. Because of different motivations for plagiarism, these three types might require different countermeasures. Help-seekers may require just different kind of support from instructors and low-barrier contact method for assistance and collaborators might require just clear communication to what extent collaboration is allowed and how it should be reported. On the other hand systematic cheaters may not be deterred even by warnings of consequences and there needs to be absolute minimal attack vectors available to leverage.

In overall plagiarism especially in online education calls for both technical

countermeasures and changes in course design to reduce motivation to cheat. There is also a place for pedagogic discussion what level of security starts to obstruct learning goals of the course because additional countermeasures limit for example the usability of the systems. Evaluation criteria for security aspect of electronic examination systems is divided into two subcategories, access control and cheat prevention. In both subcategories fully supported means human invigilators or equivalent technical supervision, for example advanced recording video surveillance like in the case of EXAM system. Partly supported authentication control is decided to mean single-factor authentication with education institution credentials and from the perspective cheat prevention it requires some kind of plagiarism detection after the exam.

2.4.3 Online learning

Online learning is one definition used to describe learning setting opposing the traditional class room setting Moore et al. [2011] found that there is no clear distinction between definitions of distance, online and e-learning. They noted that there might be geographical tendencies towards using the certain term. Last they noted that because concept of distance learning is connected with actual learning environments, definition of concept becomes also definition of characteristics for said environment. Taking this into consideration, definitions for distance learning are likely to vary from education institution to another because different learning environments have different features and different teachers use them differently.

In their conference paper about electronic examination Rytönen and Myyry [2014] present classification for the exam sessions. The classification has two dimensions: whether the place and time are fixed for the exam or not. To extend this classification third dimension is introduced: can students complete exams on any device or does it have to one provided by the education organisation. Applying the classification for the context of online learning, success criteria for electronic examination system supporting online learning is how many of three dimension of examination setup support taking the exam at distance. All dimensions of the examination are discussed in detail below but for the scope of this study these options are limited to binary yes or no values. Choosing the location of examination has several different levels of freedom. The opposite of every students taking the student in predefined lecture hall is that exam can be taken even in one's home. Location dimension has a connection with security aspect. If exams have to be taken on-campus at designated spaces there is option that human invigilators are used to enforce authentication and how rules are followed. In overall at university premises there are more options to enforce access control

such as physical access to exam lab is provided only for students with valid student card. On the other hand when taking exams off-campus technical solutions have to be utilised more to achieve sufficient level of security.

Being able to choose time when taking the exam brings a lot of flexibility for students thus being beneficial attribute. From teachers' perspective temporal distribution of exam sessions might bring additional work load if several versions of exam questions are required to avoid students taking exams earlier to share questions for their peers.

The device that students have available during the exam can has big impact on many other aspects of the exam. Computing power limits the scale of problems available for solving while operating system may limit software tools that can be installed on the device. Usability and familiarity should not be neglected either: exam situation is easily stressful and have to use specific computer for the first time increases the stress.

Online education where all learning activities and interactions between students and instructors happens online is a strict opposite for traditional classroom or contact education. Between these two extremes is wide spectrum of education arrangements where some of the activities or interactions take place online and other in the classroom face-to-face. This kind of education is called blended learning. Osguthorpe and Graham [2003] presents three elements that can be blended in education: activities, students and instructors. When learning activities are blended same group of students and instructors engage in both class room and online learning activities. Other options are that in addition to students engaging in classroom education there are some students who interact only online. In the same manner the last form of blended education would be that face-to-face course has several instructors, some interacting in classroom and others online. The courses that are part of this thesis study fall mostly to the first model of blended learning. Same instructors and students engage in lectures and exercise sessions face-to-face and students complete programming exercises online. However there may be elements from the second model as well because some courses have massive open online course (MOOC) versions of them that are offered totally as online course.

There are several studies of factors affecting the success of online or blended education. Many studies focused on what affects the perceived satisfaction of the students. For example Sun et al. [2008] summarised thirteen factors from eight studies that were compiled into six dimensions, learner, instructor, course, technology, design and environmental. Sun et al. [2008] conducted also a case study trying to find factors with the most impact. They found out that seven most important factors were learner computer anxiety, instructor attitude towards online learning, online course flexibility and on-

line course quality, perceived usefulness, perceived ease of use and diversity in assessment. These factors fall into different dimensions so that both factors in course and design dimensions were found to be important while no factors from technology dimension were found to be important. The presented six dimensional model has been tested out in further studies. For example Chen and Yao [2016] found the design dimension to be the the most important dimension. Volery and Lord [2000] studied critical factors that enable the success of online education. The most important factors they found were technology, the instructor and the previous use of technology from student's point of view. While these factors were originally intended for the context of pure online education, we extend them to the context of electronic examinations. These success factors do not strictly correlate with the three dimensions of the online learning. The most related dimension is clearly third one: applied device and software tools. It looks like that electronic examination system would support online learning best if the system is easy to use and if students are familiar with the system or are other ways capable to operate with the system.

The model proposed by Sun et al. [2008] could probably be used to evaluate also electronic examination systems, because examinations are learning activities as much as for example lectures are. Also most studies reported that technology or system related factors are among the most important ones, as Chen and Yao [2016] reported, so the framework is likely extendable to the context of electronic examinations. However it will not be feasible to apply six dimensional model in this thesis, one reason being that acquiring software systems to test them may be tedious. Also conducting proper usability evaluation for examination systems to reveal their compliance with design dimension would be worth its own thesis study and as such falls outside the scope of this research.

2.4.4 Exam content and creating exams

While the focus of the on-line examinations is to test practical programming skills of the students, Mason and Woit [1998] note that computer science, especially in polytechnic education, includes also theoretical knowledge that might be tested more easily in traditional exam. To test different skills and knowledge areas different exercises must be utilised. Programming exercises are the most basic exercise type that was used in research studies, for example by Arnow and Barshay [1999] and Rajala et al. [2016]. Other exercise types that are mentioned in researched studies are multiple choice questions (MCQs), free-text answers, algorithm simulations and context specific programming exercises. All exercise types have different relations with other

aspects of exam, especially how easy and thorough automatic marking is and what kind of tools they require to complete. The list of exercise types is not exclusive. Instead it is an advantage for a system if it supports arbitrary number of question types and adding them is easy.

The programming exercises test students' ability to read, modify and produce functioning source code [Mason and Voit, 1998]. Automatic marking programming exercises usually requires quite a lot infrastructure to allow robust tests and lightweight process of creating new exercises. This one reason why electronic examination systems designed by Arnow and Barshay [1999] and Rajala et al. [2016] used existing LMSs as infrastructure.

MQCs are lightweight problems that are suitable for testing basic concepts and terminology. They are also easy to mark because of clear division between right and wrong answers. Variations to MQCs include exercises where students need to pick all options that apply or pick exactly one correct answers while other options are wrong.

Free-text answers is broad question type ranging from questions that are answered with one sentence to ones that require several pages of discussion. All this kind of exercises are easy to construct, in minimal setup they require just one text box stating the question and one input box where the answer is typed. Also these exercises require minimal tools: it is usually sufficient to provide means to produce text while ability to format text is optional. Automatic marking of free-text answers is quite a mixed bag. Short text answers can be marked automatically using regular expressions for example or students can be instructed to format their answers into machine readable format like comma separated lists. However long text answers are much harder to mark automatically as the grader would need to be able to interpret natural language and the context of the answer.

Algorithm simulation problems test students' ability to follow algorithm execution by asking them to perform for example sorting algorithms step by step. The ones used with the A+ LMS have incorporated automatic marking and they randomize the problem for each attempt Karavirta et al. [2013].

Specific programming exercises, like one presented by Rajala et al. [2016] are assignments that test students ability to solve problems in specific context. They might require applying several different skills like programming and algorithm design to complete successfully.

2.4.5 Integration

Integration has several objectives based on the stakeholders. From teachers' point of view the most important goals are being able to easily calculate final grades for the course and as stated by Arnow and Barshay [1999] to reuse

same exercises during the course as well as in exams. Additionally data integration between systems may provide new opportunities for advanced learning analytics and primary data for research purposes.

According to Arnow and Barshay [1999], using homework problem sets in exam can be very effective in motivating students to pay more attention to homework because doing them well has more direct impact on performance in final exam. Additionally having same problems of same type during the course and in exams increase the familiarity of the exam situation and so might reduce stress factors.

Integration is important also from the perspective of developers and system administrators. They are interested in maintainability, extensibility and interoperability of the system to enable easy transfer of data from system of data and introduction of new data formats.

The table 2.1 presents electronic examination system evaluation criteria that is used in this study. The criteria are based on ones presented originally by Arnow and Barshay [1999]. The criteria have been adapted for this thesis by combining two evaluation categories and elaborating some categories by adding subcategories. The descriptions of each criteria are illustrative and based on studied systems. The goal of the criteria is not to strictly compare systems but to offer terminology to classify and describe different kinds of systems designed for different kinds of use. Next all evaluation criteria are described in detail and comparison to traditional examinations are provided when applicable.

Minimal level of automatic marking is comparable to optical marking of MCQs that are used in many standardised tests. On next level the student's output is compared to one defined by the instructor before hand, The string matching can ignore whitespaces in the output as WebToTest system [Arnow and Barshay, 1999] does or utilise regular expressions. On highest level system can use grader extensions to run tests provided by the instructor [Karavirta et al., 2013].

The exam system that has minimal security level resembles any course assignment blurring line between exam and continuous evaluation that is done during the course. It is typical for the course assignments that identifying student is based on their own report: written name and student number attached to report that is compared to email or LMS account used for submitting assignment. On medium level of cheat prevention measures are taken mostly after the exam situation: submission may be run into a plagiarism detection system or watermark system like described by Simon [2005] may be used. The highest level of security corresponds with current setup for identifying students and preventing cheating in pen and paper exams: human invigilators verifying identity from official document and also patrolling

exam hall to prevent cheating. Of reviewed systems state-of-art of technological enforcement of security and access control consists of multi-factor authentication (student card and account credentials) combined with recording video surveillance and standardised software environment as is the case with Tentti Plussa.

To fully support online learning, electronic examination system should totally support all three listed subcriteria, location, time and tools. All subcriteria range from being restricted to a single option to completely up for students consideration. In the middle there is the option where students can choose from limited set of options chosen for them. For example when considering time in traditional exams teacher and institution design the exam schedule and students have to follow it. This is equivalent with the option with minimal choice. The next option would be that students are provided rather wide time range, say two weeks for example, when students can book exam time whenever fits best for them. This is the case with the Tentti Plussa system. In the most free option students can take whenever they wish after the prerequisites are met.

Support for question types involves the sheer amount of different types the system supports readily but also to fully support different question types the system needs to be easily extensible with new question types. The complexity of supported questions does not matter

Integration has different criteria for two different stakeholder groups, teachers seek to integrate systems so that problem sets from the course can also be utilised in exams. When the integration is fully supported teachers can for example choose from shared repository which questions to include in course material and which in exam and both LMS and exam system would support same question format. The other end of the spectrum is that problem sets are incompatible between the systems and in the middle is the state where question sets are compatible but reusing them requires manual work. Developers and system administrators are interested in integration with existing systems so that the system would be as maintainable and reliable as possible while also allowing extensibility.

2.5 Analysis of existing systems

Results of analysing systems presented in this study are summarized in Table 2.2 the legend used to code the results are presented in Table 2.3.

WebToTest marked student submissions automatically by comparing their output with predefined input with the one defined by the teacher. [Arnou and Barshay, 1999] The security of the system is based on allowing only se-

	Minimal or no support	Partly supported	Fully supported
1. Automatic marking	Only for MCQs or similar	String matching and / or regular expressions	Unit testing
2. Security			
ā) Access control	No authentication	Weak authentication	Human invigilators or multi-stage authentication
ā) Cheat prevention	Student is trusted	Detection after exam	Video surveillance, technical restrictions or human invigilators
3. Online education education			
ā) Location	At predefined location	Selected from a list of possible locations	At anywhere
ā) Time	At predefined time	Selected from available slots	Anytime
ā) Tools	With predefined device	Any device provided by the organization	Any device
4. Support for several question types	Only one or two supported types	Supports wide set of question types but adding new question types is cumbersome	Abstracted interface that supports arbitrary number of different questions.
5. Integration			
ā) With existing systems			
ā) With exercise sets	Assignment format are incompatible between systems	Reusing question sets require manual work	Exercises are presented in same format in LMS and examination system

Table 2.1: Summary of success criteria for electronic examination systems.

lected connections to other computers to prevent communication or accessing third-party resources and human invigilators. Arnov and Barshay [1999] argue that their system is suitable for online education due to its web-based infrastructure, this comes with a compromise of a security because enforcing security measures is much more difficult in uncontrolled environments. Due to being built upon a LMS, WebToTest supports extendable amount of exercise types and integrates well with existing systems allowing especially reuse

	Automatic marking	Security	Online education	Problem types	Integration
WebToTest	O	X	O	X	X
ViLLe	X	X	-	O	X
EXAM Consortium	-	X	O	O	X
Tentti Plussa	X	X	O	X	X
Abitti	-	X	O	-	-

Table 2.2: Analysis of electronic examination systems.

Symbol	Description
-	Not supported
O	Partly supported
X	Fully supported

Table 2.3: The legend for Table 2.2

of exercises from the courses.

ViLLe supports automatic marking of programming exercises by running the source code. The security of the system is strong, human invigilators supervise the computer lab and also authenticate students. However the system is not well suited for online learning purposes. Rajala et al. [2016] demonstrated the exam only in specified computer class and so that all students were present at the same time. The ViLLe system integrates well with the existing course material because it is built on LMS carrying the same name.

The EXAM system meets one objective for electronic exam system well and two more somewhat. The security of EXAM system is rigorous and the system has been adopted for production use at many consortium member organisations. From computer science education point of view EXAM would need more supported question types and also automatic marking is not well supported. The only question types that are usable within EXAM system are free text answers and multiple choice questions.

Tentti Plussa supports all same problem types that A+ LMS: programming problems, multiple choice questions, algorithm simulation and free text answers. Automatic marking is supported for all of the question types, although longer essay answer are likely to require manual review. Also being just a new instance of the same system, it is easy to share exercises between course and exam and also between different courses. Also because of close

relation with A+ LMS, data from both systems can be fetched in same format and so integration with the existing LMS is easy. Tentti Plussa is not applicable to use in online learning because it is very strictly tied to a specified physical location. On the other hand exams can be at least in theory be distributed wherever it is needed and so it is technically possible to take exams at different campuses. Overall, Tentti Plussa achieves four of the five listed objectives totally and the fifth partly.

As the knowledge is available, the Abitti does not support automatic marking at all. The security grade is highest of all evaluated systems because high security standards of the matriculation examination. Human invigilators supervise the exam room, all items brought into room are examined thoroughly and so on. The Abitti system is made to be used in the local are network so it does not support online education very well. However the system is designed to use students' own laptops so it scales up quite well. For the context of computer science education selection of exercise types is narrow, while supporting different materials, answers are mostly free text or MCQ types.

2.5.1 Summary

There are several aspects how electronic examination systems can be evaluated and many of those aspects like security or online learning are deep research topics on their own so it is difficult to synthesize a complete framework for evaluation purposes. Especially the online aspect was discussed deeply, mostly because it relates closely to how well online learning activities and especially scale for courses that have several hundreds of students. Literature provided quite good overview of security and cheating in online learning. I suspect also based on common knowledge that security is contradicting aspect for usability of the system. Because teachers are in the focus of this thesis, security is not in main focus of the empiric part while the usability was not covered that well in the literature and will be studied more in practice.

The results aggregated to the Table 2.2 does not point out a system that would suit the needs of this study perfectly. While Tentti Plussa, ViLLE and Abitti all receive good evaluations, they all lack one or another critical requirement. Overall all these systems are too limiting in their use scenario, a criteria that was not captured in the evaluation template. Because none of the systems suffice our needs, the empiric study will start to build on existing LMS and other technology already in use at the Aalto University while picking up good parts of the analysed software as requirements.

Chapter 3

Research methods

3.1 Literature review

The literature body about lean startup was gathered by database searches to Google Scholar and Scopus. Initially criteria were either matching to text phrase "lean startup" or citations to Ries's original book "How today's entrepreneurs use continuous innovation to create radically successful businesses" introducing the lean startup methodology. Later on more material was searched by adding "case study" and "experience report" to search phrases. Search string "'lean startup' AND 'case study'" produced 967 hits from all 2799 papers citing to Ries's original work. During the search, journal article were preferred, but also conference and workshop papers were accepted. The relevance of the paper was determined first by the topic and second by the abstract. When analysing papers, answers to following three questions were looked after: What elements of the lean startup was applied in the study, was lean startup accompanied with any other framework or methodology, and writers conclusion about applicability of lean startup methodology. Based on this analysis, conclusions were drawn how the study under analysis supports lean startup methods presented by Ries [2011].

3.2 Empirical study

3.2.1 Research Process

The research was conducted as an action research as described by Avison et al. [1999]. A qualitative action research was chosen as it allows deep understanding of the organisation and testing out new practices [Avison et al., 1999]. Also iterative nature of the action research allowed us to adjust what

practices were applied and how. Due to time constraints for the empiric research, only few iterations of action research was performed.

The gathering of domain area specific knowledge was started by interviewing one key informant within parent organization of the client. Knowledge and connections gained from the interview was used to get accustomed with earlier and current proceedings on the topic in Finnish universities.

The action research process consisted of three phases: problem diagnosis, action intervention and reflective learning Avison et al. [1999]. The problem diagnosis utilised requirements engineering and UCD methods in collection, analysis and presentation of the data. The action intervention employed modified version of the lean startup methodology. The practices specifically under research were, vision, MVP and hypothesis setting. The applied research process is illustrated in Figure

The qualitative methods of RE and UCD used in problem diagnosis phase allowed us to delve deep into needs of the key user groups and elicit also tacit requirements for improvement. Semi-structured interviews were used to understand key user groups and the process was accelerated by the use of key informant who provided contacts around the organisation. Key informant also helped with understanding different goals for and approaches taken towards the electronic examination within Aalto University as well as in other Finnish universities. After the collection of information, affinity diagrams were used to analyse and organise the information. Eventually customer journey maps were used to document and communicate the current use, needs and point points of exams.

In active intervention phase we applied chosen best practices of lean startup methodology. The practices were vision, setting hypothesis and MVP. Vision was chosen for the action intervention because sharing and communicating the common goal has been shown to help teams reach better efficiency and Ries [2011] illustrates very little what he means with the vision and how it can be communicated. These methods were applied with university lecturers at the department of computer science.

3.2.2 Data Collection and Analysis

In problem diagnosis phase data was collected with semi-structured interviews. The method was chosen because it allowed elicit tacit knowledge and experience of the examination system while also following planned format. Two researchers analysed the questions twice before the interviews. The first researcher focused on the content of the questions and the second on the neutrality and clarity. All interviews were recorded and transcribed verbatim.

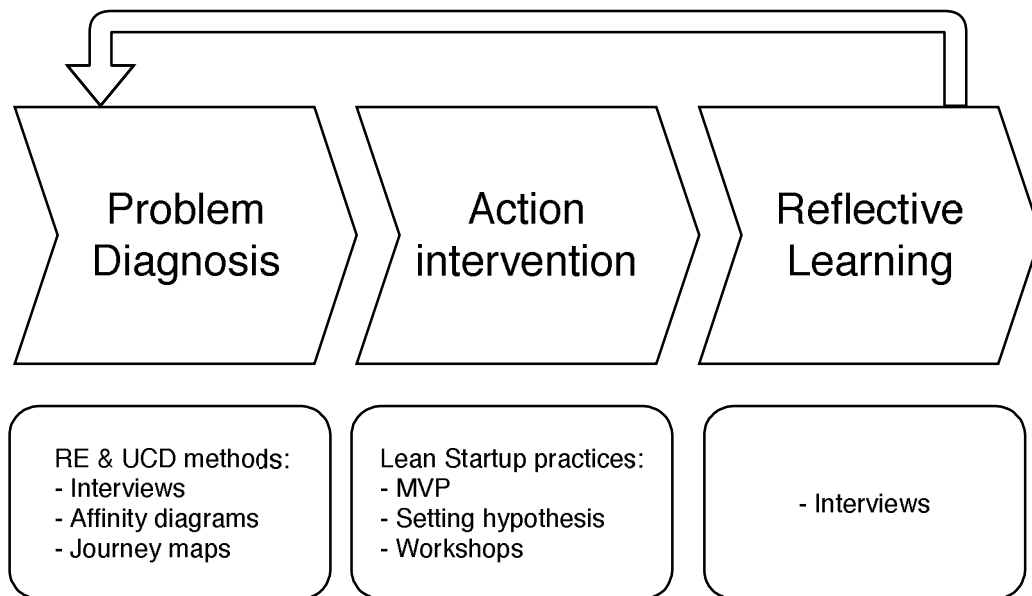


Figure 3.1: Action research process of the study.

When analysing the interviews the process models produced by interviewees were compared with each other. First the common activities in the process were identified and second the activities that were unique to a single interviewee. Special attention was paid to each description of the activity to understand whether interviewees had same meaning and content for activities with similar names. The results of the analysis were presented as a customer journey map describing the activities and motivations behind them.

Chapter 4

Results of the literature review

4.1 Lean startup

4.1.1 Overview of the Lean Startup

Lean Startup is an iterative process where assumptions about product idea, customer and market are tested with experiments using minimum viable product (MVP), smallest increment that allows experimentation. Lean Startup process model and related activities are pictured in Figure 4.1

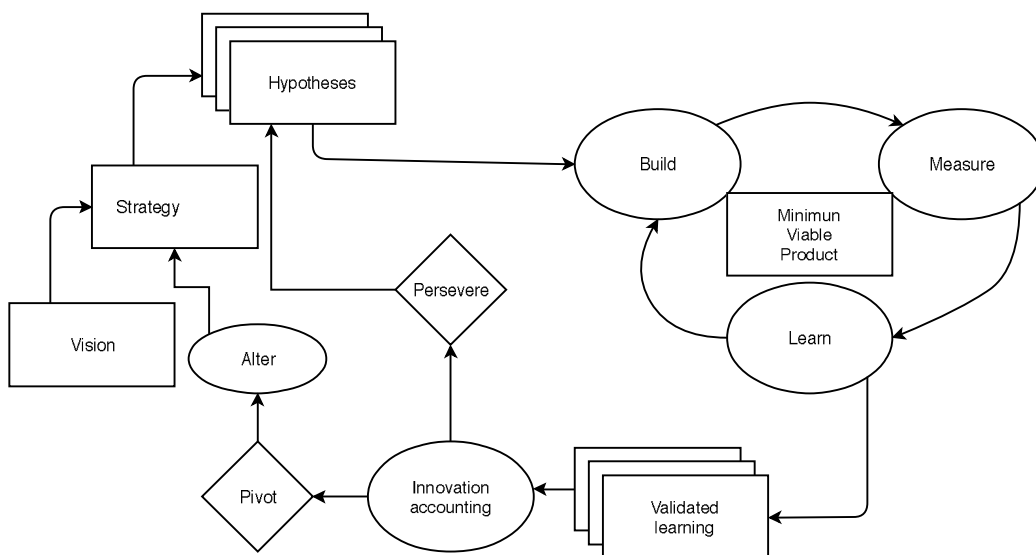


Figure 4.1: Lean startup process model.

Two fundamental paradigms that carry through the whole lean startup methodology are positivism and effectuation logic. Ries [2011, page 272]

shares the positivist management paradigm with F.W. Taylor and whole book is based on view that startup as an institution can be modelled as atomic activities. This paradigm is also visible on Ries's view that best way to obtain information from the customers is to observe them and that quantitative metrics give indisputable information about the interaction between customer and service. Interpretative and qualitative methods are used only as secondary way to obtain information. This view is demonstrated for example when Ries discusses about suitable metrics. He argues that metrics should be derived from user behaviour directly.

Effectuation logic within lean startup means taking small steps in quick iterations and adjusting frequently to achieve the ultimate vision instead of fixing for example release plans several months ahead. Effectuation logic also means using the existing resources (capabilities, knowledge, etc.) and gradually developing resources to match the needs of product and business development.

Blank [2013] offers another, very condensed overview of the lean startup methodology, as he calls it himself. According to him, lean startup has three underlying principles: business plans are initially just good guesses, the use of hand-on customer-centred development method Blank calls "customer development" and thirdly agile development. The first principle emphasizes the use of effectuation logic mentioned earlier: bringing product and business idea to contact with customers and partners as quickly as possible to get feedback for adjusting the business model.

Customer development is business design framework developed by Steve Blank himself. It has two parts: searching for appropriate business model and executing that discovered model [Blank, 2013]. The lean startup connects to that former part of the customer development: discovering new customer needs and adjusting business model based on customer feedback. The latter part is more related to traditional business management: executing and monitoring predefined plans and that phase could be seen as a continuum after business model has stabilized after applying lean startup long enough.

Blank [2013] describes third principle, agile development just as an iterative and incremental development of the product, starting from most important features to deliver minimum viable products as soon as possible. This kind of agile development is nothing new in software development, but as Blank [2013] notes, it goes well with customer development as one of the most important advantages of agile development is being able to respond to changing customer needs.

Blank has many similarities with Ries, which is very understandable as, at least according to Blank [2013], Ries coined lean startup after having his course of startup management combining customer development with agile

software development practices. While Ries [2011] has clear practitioner focus, Blank expands methodology by adding new tools and connecting with existing practices. One key addition is inclusion of Osterwalder's business model canvas [Blank, 2013]. This formalizes the vision behind lean startup, one important part that was neglected by Ries. The business model canvas has the ability to document key hypothesis about different elements of the vision.

4.1.2 Principles of the Lean Startup

In software engineering software development processes are built on principles, practices and tools. According to Oxford English Dictionary [2018] a principle is *"A fundamental truth or proposition on which others depend; a general statement or tenet forming the (or a) basis of a system of belief, etc.; a primary assumption forming the basis of a chain of reasoning."* and a practice is *The actual application or use of an idea, belief, or method, as opposed to the theory or principles of it; performance, execution, achievement; working, operation* and tool is an implementation of a practice.

Ries [2011] starts his book by laying his five principles that are used to build lean startup. Based on the above definition of principle, some items on the above list are disputed as principles. Also some of the principles relate only to either business or development perspectives of lean startup. Below the five principles are explained and discussed whether they are relevant principles for the scope of this thesis.

Entrepreneurs are everywhere. This sets foundation for the generalization of lean startup to be applicable for everyone who seeks to develop disruptively innovative products and who is facing big uncertainties. This principle is important from the perspective of the thesis because it implies the applicability of lean startup to problem domain and context.

Entrepreneurship is management refers to that lean startup has two perspectives: product and business development. This principle is clearly out of the scope of this thesis.

Validated learning is term Ries uses to measure the success of a startup and as such it is more related to business development perspective. Validated learning is compact and reader-friendly report of the conducted experiment and as such is practise of how results of the learning are represented in a coherent form.

Build-measure-learn as Ries describes it, is the main activity of a startup. An activity is in its total essence a way to apply ideas and principles and so a practice as defined above.

Innovation accounting is related to how investors can measure progress

and set goals for startups and is related more to the business development perspective. As the validated learning above, innovation accounting can be seen as external reporting practice for the product development team.

4.1.3 Practices of the Lean Startup

Ries [2011, page 22] describes **Vision** as startup's destination: creating a thriving and world-changing business. This vision rarely changes through the journey of the startup. Ries's definition for the vision lacks a theoretical and practical context: what elements does the vision contain and how it can be presented, analysed or communicated. Blank and Dorf [2012] sheds some light on definition of vision in his book Startup owner's manual. He says that startup's vision should be seen as a series of untested assumptions that need to be proved by bringing them into contact with customers. Thus vision could be documented, shared and communicated as a simple (prioritized) list of assumptions about the product idea.

Strategy is a plan of the entrepreneur how they can achieve their vision [Ries, 2011]. Strategy usually contains business model, product road map and strategic decisions like partnerships and customer segment. Strategy is not absolute plan how the vision is achieved, rather Ries [2011] states that strategy best guess at the moment how to do that and so strategy is subject of change if innovation accounting shows that current path of developed does not yield good enough results.

Hypotheses are key assumptions about customers, their needs and about how they would interact with the product idea under development. They are also corner stones that cause the whole product strategy or at least part of it to fall apart if proven inaccurate. Ries [2011] instructs that to guide business development process properly, hypothesis need to include clear quantifiable indicator how to measure whether they are valid or not.

Building means implementing a test for a chosen hypothesis by adding new features to the MVP. The fundamental idea of building, according to Ries [2011], is to do it with as little effort as possible and implement only the absolutely necessary features and architecture.

To assess the effectiveness of new features of MVP implemented in building phase, change has to be **measured** [Ries, 2011]. For this Ries recommends quantifiable metrics, "actionable metrics". Three characteristics, or triple A, for actionable metrics are actionable, accessible and audible. For a actionable metric it is clear which action causes the change in metric and how the change in metric could be replicated. Being accessible means for the metrics that they should be as easily understandable as possible and one key part of that is to make them very tangible. One example Ries [2011,

144] gives is a visit to website which is itself fuzzy term but unique user visiting the site is much more tangible. Accessibility also means that metrics should be available to as wide audience as possible within the startup to allow informed decisions and avoid making same mistakes again. Last attribute, audibility, means that metrics should be linked to real world and real customers so well that managers or other employees can check with the source that the facts are correct. Also audible metrics should be derived from master data directly and with very simple process to allow traceability. One tool Ries [2011] encourages to achieve the goal of good metrics is cohort analysis where each new set of users is treated as a separate unit of analysis so that progress of key indicators can tracked through the development of the product. In cohort analysis key metrics, like percentage of visitors of the website continuing to register as an user or purchasing product, is calculated separately for each cohort [Ries, 2011]. The idea is that these metrics should improve over the time as the development team learns to know their customer better and better.

The opposite of actionable metric is vanity metrics, which according to Ries [2011] are business metrics that show easily at least some progress. One example given of vanity metrics is total amount of registered users. If the amount grows steadily, is it enough or should the team be worried that they cannot convert visitors to registered users any better than before.

The last phase of the build-measure-learn -loop is to read collected metrics to **learn** whether hypothesis under test was validated or rejected. This knowledge is converted then to **validated learning**. Validated learning as a concept is tightly linked with appropriate metrics discussed just above and Ries [2011] uses same triple A attributes to describe how validated learning should be reported. Actually, it's the simple and traceable reporting that makes learning validated as the results can be easily checked afterwards.

Minimum Viable Product (MVP) is implementation of hypothesis aimed to allow customers to either validate or invalidate the hypothesis. According to Ries [2011] MVP should be smallest possible increment that is sufficient to conduct experiment at hand, other than that, there is little description what MVP should include or look like. Ries [2011] calls for innovative ways to implement MVPs especially in very early stages of startup life-cycle and to avoid any unnecessary infrastructure and in extreme cases to avoid coding altogether. One example Ries mentions is concierge MVP: startup personnel simulates visioned product idea by using personal effort. This allows startup companies to skip timely and expensive early product development and to quickly learn from close collaboration with potential customers.

Innovation accounting consists of three phases. First MVP is used to

establish a baseline how customer react to and interact with the product. In second phase development team implements a number of small changes to the MVP to find direction that improve the baseline, or “tune the engine”. Third phase is decision of pivoting or continuing with the current strategy. The decision is made based on the efficiency of the engine tuning -efforts. If the company is on the right track, accumulated validated learning should enable faster and faster improvement of the baseline. In the opposite case if the improvement of the baseline is stagnant, pivot is likely the right decision.

Pivoting means committing to a major shift in strategy to fulfill product vision [Ries, 2011]. Ries mentions ten different pivots in the book ranging from changing customer segment to choosing different monetizing method or new logic how new users and customers are acquired. According to Ries [2011], pivots are done to accelerate startup’s learning and growth. Actionable metrics are used to determine the need for a pivot: Ries [2011] says that accumulated learning should be exponential and for example flat rate of growth of customer base hints that a pivot might be necessary.

Experimenting is the fundamental approach of lean startup to both acquire information about customers and to develop the product, because the product is used as an tool to collect the customer information. Experimentation also dictates what kind of interaction is emphasized between customers and the startup. Ries [2011] emphasizes eliciting reactions from users and observing that behaviour to measure the success of the design instead of asking them directly what they think of the current MVP.

4.1.4 Studies about applying Lean Startup

Case studies and other researches of applying lean startup practices were slightly hard to come by. Especially difficult it was to find studies that would analyse the use of each practice separately.

May [2012] reports about her experiences of using lean startup to develop novel consumer application. She does not explicitly state what elements of the lean startup they used but mentions at least validating hypothesis and MVP as well as maximizing the efficiency of learning instead of for example development output. May [2012] paired lean startup with agile software development and use of User-Centered Design (UCD) methods for user experience (UX) design. In this case agile software development was a mix of scrum, an agile framework for software development, and extreme programming (XP): sprints, dedicated cross-functional teams, test-driven development and pair programming. The applied UCD methods included user stories, personas, ethnography and wireframes. May [2012] does not analyse the usefulness of applied lean startup elements directly, but many recom-

recommendations list by her can be addressed by the use of lean startup. Those include engaging users earlier to validate not only the product idea but also the business model and not committing too early to a single technological solution. In conclusion the experiences of May [2012] do not validate the use of specific lean startup practices but on the other hand it does not either discard lean startup as unnecessary or harmful. For lean startup it is promising that many issues noticed in the report could be remedied with it. It is also noteworthy that report calls for supporting or complementary practices and frameworks, like UCD methods in this case. It seems that lean startup gives good approach and framework to new product development but it requires other frameworks to detail the implementation of lean startup to address the current situation.

Münch et al. [2013] report about a case study of utilising MVP as described by Ries [2011] in a collaboration project between a established company and university. The target of the study was the whole development process which was manifested by a functioning prototype. While the perspective of the study was broader than just the technological artefact, at least some of the results were inherently linked with the characteristics of the artefact. For example they reported that regular weekly review sessions helped in communicating the results and next design decisions.

In their research Khanna et al. [2018] studied the relationship between hypotheses and MVPs. They found out that in their cases the relationships were non-linear and incomplete meaning that not all hypotheses were ever implemented into MVP and MVP had features that were never formulated as hypotheses. This may suggest that using the hypotheses to guide the development process may be trickier than suggested by Ries [2011].

4.2 The applied process model

It was discussed above that the academic year has five periods when examinations are mostly organised. This limits the possibilities for large-scale user-testing from the students perspective. Thus, the iteration length is increased and additional methods are utilised to validate smaller details of the system between the user-testings.

As the preceding section shows, the lean startup methodology requires rigorous following of the principles to work but also at the same time it leaves quite much room for different implementations as the methodology focuses more on practices and why they are important. Thus, we will use modified and more detailed version of the methodology in this study. As suggested by May [2012], we will apply user-centered methods to help with requirements

engineering and UX design. This choice is also affected by the application context. Because examinations are organised only few times a year, there is no easy way for extensive user-testing with the students. UCD methods are used to enable the quality of the MVP for large-scale user-testing.

To limit the scope of the thesis, only some of the lean startup practices are chosen to be applied and analysed in detail. These practices are MVP, setting hypothesis, experimenting and validated learning.

Chapter 5

Results of the empiric study

5.1 Current State

Four university lecturers were interviewed to understand their process for arranging exams. It was also studied what motives teachers had for having exams and what kind of requirements those motives set for possible electronic system. The customer journey map of arranging exams from teachers perspective is illustrated in Figure 5.1.

Teachers start planning exams by scheduling exam dates for the upcoming academic year. Usually this does not take much effort, because courses are established over specific periods and so there is rarely need to drastically adjust dates from year to year. After this task there is usually long gap before the process continues. Teachers start to design and think about possible exam questions when the exam date approaches. Most teachers had quite established structure for the exam questions based on the course content and they also had different ways of managing and utilizing exam questions used over several years. After a teacher has come up with the exam, they might need it translated, because bachelor students are eligible by law to have exam questions in either Finnish, Swedish or English based on their language of earlier education. Aalto University offers teachers translator services but they need to be scheduled at least a week or two in advance. There is also conflict in schedules within the process because students can request exam on their language just few days prior to exam.

At some point during this time students may contact teachers and request special arrangements for their reading impairment or for other reasons. Usually handling these requests does not require much effort because university and department have good procedures but they still disrupt work flow of the teacher and require immediate action. When there is about one week until

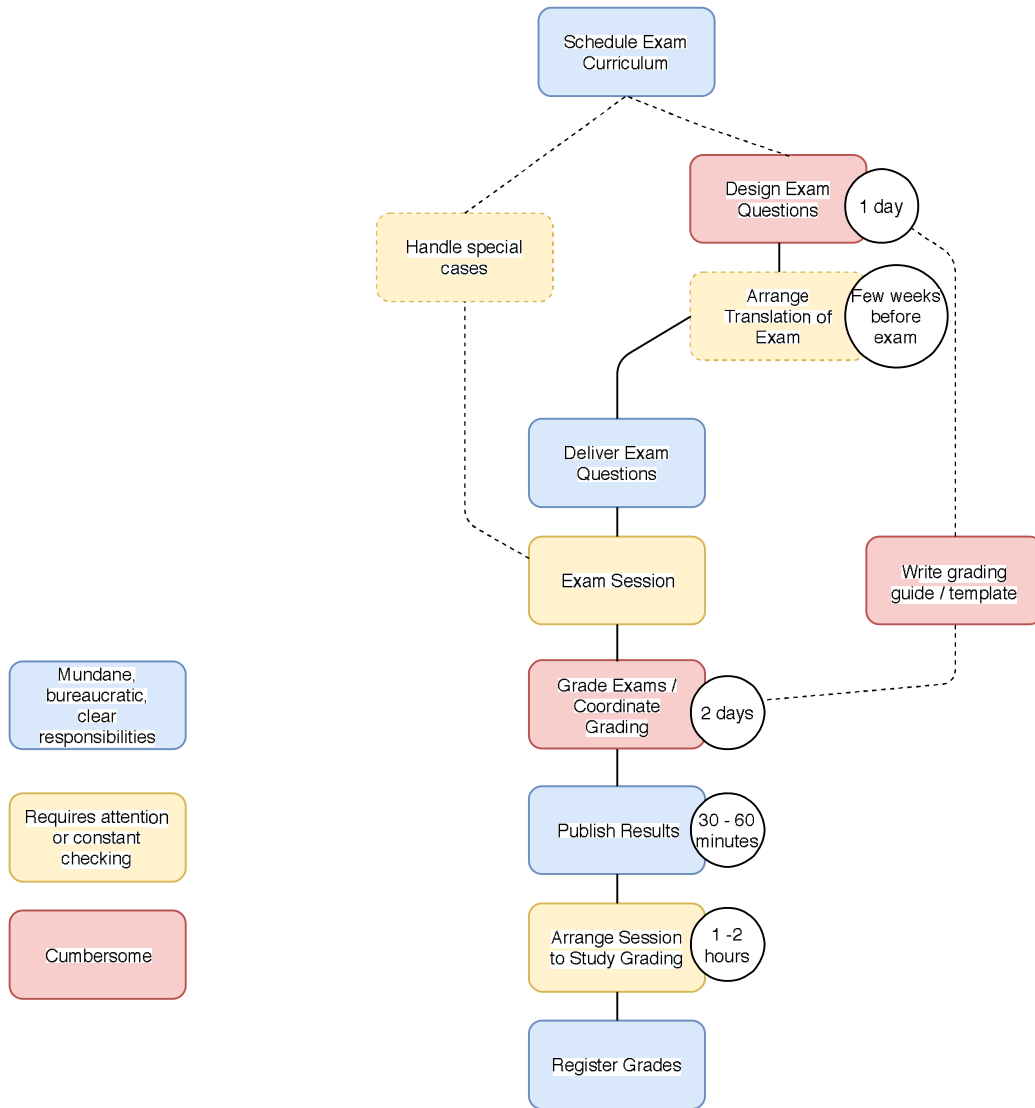


Figure 5.1: The customer journey map for teachers.

the exam, assignments are sent to the secretary of the department who prints adequate amount of copies and who has already taken care of booking space for the exam and hiring supervisors.

Even though every lecturer had almost every task listed in customer journey map in their process, the processes had also distinctive differences, one notable being how formalised the processes were. While some teachers performed tasks in clear sequential order and had automated scripts to do manual work, others had more emergent and intertwined processes.

Absolutely the most important reason for having exams was the need to

assure that students had completed the course by themselves. To ensure this kind of control, security would be critical requirement for the examination system. During the interviews several concerns for the security aspect of the electronic examination was brought into attention. Those concerns ranged from the circulation of answers to frequently asked exam questions to how students can cheat automatic identity checking systems if there are no invigilators present in electronic examination situations.

5.2 Action Intervention

The first MVP was a three-dimensional conceptual model of electronic examination systems illustrated in 5.2. The model is an extension of two-dimensional model presented by Rytönen and Myyry [2014]. Three dimensions are the components of online learning: space, time and tools. All components have wide spectrum but they were condensed to binary options to emphasize characteristics of options and to limit the amount of the combinations. It was decided that for this MVP restricted timeslot meant that all students must take the exam at the same limited time of couple of hours. In other extreme students can take exam whenever they have completed prerequisites for it but "free" option also includes situations where students can choose from previously selected set of times or have extended period to do exam at their own pace. The main distinction between the free and restricted for this component is whether all students take exam at the same, quite strict time frame. The restriction of space for exam was chosen to mean that in there is human control in the space enforcing identity checks and observing for use of prohibited tools and resources. Finally the restriction for allowed tools meant that all students take exams with standardised workstations or laptops provided by the university while the other option was that students can use their own devices. In the middle are cases where students use some form of sandboxing to get standardised work environment with for example virtual machines or live operating system booted from USB drive. These solutions could be used on either student or faculty owned devices.

The text description of each option tries to depict each of the eight combination. Some options have similarities with learning activities used already, while other combinations are more of theoretical discussion. Starting from the case where all three options are controlled, this resembles most the current setup how exams are taken at lecture halls, under human supervision and exactly at the same time. If the space option is relaxed, students would take university provided devices to their homes and take the exam at the same time with everyone else but wherever they wish. Compared to first

		TOOLS			
		Controlled		Free	
		SPACE		SPACE	
		Controlled	Free	Controlled	Free
TIME	Free	Exam aquarium	Home exercise taken on standardised device	Exam aquarium with own devices	Home exercise
	Controlled	Traditional exam	Exam taken at home on standardised device	Traditional exam session with own laptops	Home exam

Figure 5.2: The first MVP: Three dimensions of online learning and electronic examinations. Extended from two dimensional model by Rytönen and Myyry [2014]

case, the security would be bigger concern because it would rely totally on technical solutions. If also time is relaxed, the situation would be quite similar to one presented by Simon [2005] with the exception that students take devices from the university with them. In this setup students may also have extended period of time compared to traditional few hours to complete the exam in addition to being able to start it whenever they wish. Starting again from the situation where all three components are controlled, relaxing requirement for time would lead to setup similar to exam aquariums: exams are taken in restricted space with devices provided by the university but they may come as they please or at least reserve time suitable for their schedules.

Next we consider four options where students are not restricted to use university devices. If both time and space are restricted it is again very similar to traditional exam situation but this time students would bring their own laptops in alleviating the need for large amounts of devices possible dedicated solely for exam use. If the place is relaxed, students would take the exam at their homes with their own devices everyone at the same time. If also time is to be relaxed, would be quite indistinguishable from regular learning activities done during the course. Last, if space would be restricted

but time is free, students could come with their own device whenever they want to take exam. This would be an extension to exam aquarium situation. This would be beneficial if there would be very specific instruments at the exam space for example 3D printers that would be utilised in the exam.

The MVP was tested in a workshop organised for the four university lecturers interviewed in the current state analysis phase. The workshop begun with the presentation of the customer journey map and process depicted there was discussed together. All teachers agreed that customer journey map captured the essentials of the exam organizing process. Next teachers were presented with MVP and we discussed together what implications each of the options would have for their process using brainstorming. Options including standardised tools were quite quickly discarded as unfeasible because it was seen unpractical and taking too much resources to have enough dedicated hardware to enable nine hundred students to take exams in just few patches. It was also agreed on that while likely having positive effect on the experience of the students taking exams, open choice for time and space had too small benefits compared to costs that they would be feasible with the current technology. Especially the restriction of space depends heavily on the available technology because checking identity and enforcing other security policies is more difficult than in traditional environment. The restriction of time helps teachers because it allows them to have smaller set of exam questions and designing exam questions was one of the most time-consuming tasks identified in the interviews. Based on this discussion, the workshop converged on the solution concept that described electronic examinations as students taking exams on their own computers in designated and controlled room at specified time.

Figure 5.3 present electronic examination systems analysed in Chapter 2 along with the solution concept proposed at the workshop to the first MVP. The figure consists of three subplots, each depicting how systems relate to two of the three dimensions. In the Figure 5.3 EXAM system and Tentti Plussa are represented as one entity because Tentti Plussa utilises EXAM in class room management. Their features related to the dimension of the first MVP are identical and Tentti Plussa adds new features related to the content of the exam.

All systems are towards the upper right corners of the plots, meaning they emphasise control for all components of the online learning. Abitti is the only system using solely students' devices. The security is forced by strict control of the space exam is taken and also the device must always maintain connection with the server. The EXAM system on the other hand is the only system that clearly allows students to choose the time of the exam. For ViLLe and WebToTest this was not specified exactly but implication was that

all students are in the computer lab at the same time [Arnou and Barshay, 1999], [Rajala et al., 2016]. Our proposed solution tries to relax restrictions for these dimensions step-by-step starting from devices.

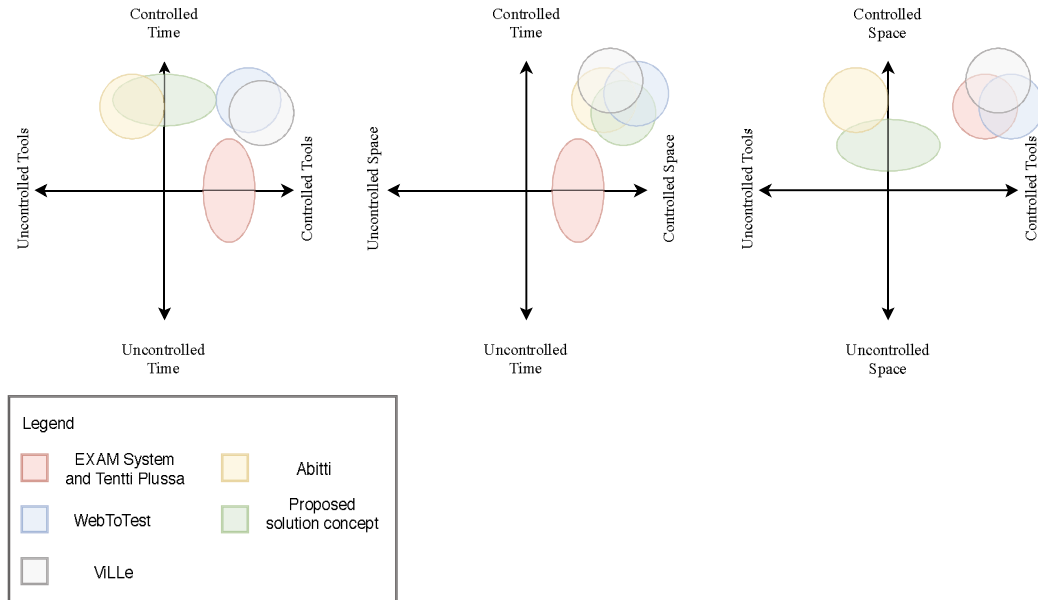


Figure 5.3: Plot of all analysed examination Tools systems.

In the following research cycle hypothesis were set to guide building the next prototype. These hypotheses are listed in Table 5.1 and they are designed to test solution concept proposed at the workshop earlier. While building the second MVP, it was realised that the electronic exam system would be quite large system consisting of a few components that are used by different stakeholders and in different phases of examination process. It was also clear that implementing even a mock-up of these components wouldn't be feasible for the next MVP, so it was decided that the most important part to test was how students view using laptops with standardised system in the exam and also how they react overall to electronic examinations. The other components were listed to the end of the hypothesis list because they would need further validation. The original hypothesis derived from preliminary solution concept are identified by the prefix "O" and additional hypothesis discovered during building the second MVP are identified with the letter "A"

In Table 5.1 the first column assigns a hypothesis an unique identifier while the second column presents one sentence description and the third one describes how the hypothesis is thought to be implemented or otherwise tested. The last column gives a rough estimate how crucial the assumption

Exercise	Category	Submissions	Points	Course staff
1.1 Exam instructions				
1.1.1 {'en': 'Feedback', 'fi': 'Palaute'}	feedback	0 / 5	0 / 0	Submitters: 0 0% View submissions
1.2 Algorithm analysis (10 + 1 points)				
1.2.1 {'en': 'Algorithm analysis', 'fi': 'Algorithm analysis'}	questionnaire	0 / 5	A 0 / 10	Submitters: 0 0% View submissions
1.3 Terminology (4 x (1 + 1) points)				
1.3.1 {'en': 'Feedback', 'fi': 'Palaute'}	feedback	0 / 5	0 / 0	Submitters: 0 0% View submissions
1.4 Colouring a tree (2 + 6 + 4 points)				
1.4.1 {'en': 'Balanced Search Trees', 'fi': 'Balanced Search Trees'}	questionnaire	0 / 5	0 / 0	Submitters: 0 0% View submissions
1.4.2 {'en': 'Exercise A6', 'fi': 'Tehtävä A6'}	questionnaire	0 / 5	0 / 0	Submitters: 0 0% View submissions
1.4.3 {'en': 'Coloring Red-black Tree', 'fi': 'Punamustan puun väritys'}	visualization	0 / 5	A 0 / 4	Submitters: 0 0% View submissions
1.5 Connected Subgraphs (7 + 3 points)				
1.5.1 {'en': 'Network', 'fi': 'Network'}	submit	0 / 5	A 0 / 10	Submitters: 0 0% View submissions
1.6 Give feedback				
1.6.1 {'en': 'Feedback', 'fi': 'Palaute'}	feedback	0 / 5	0 / 0	Submitters: 0 0% View submissions

Figure 5.4: Screenshot of the second MVP: chapter in A+ LMS consisting of exam questions.

is for the whole system. In general the more important the hypothesis is, the sooner it should be tried to be validated.

There were four identified components of the electronic examination system: student interface, interface for constructing exams, grading interface and hardware, operating system and related infrastructure. Because student perspective was decided to be the most important research area for the second MVP, the student interface was emphasised when designing the MVP. The hardware and infrastructure was also closely related to the second MVP but it was viewed to be of lower priority, so it was implemented as easily as possible and feasibility and scalability of design was left for further tests. Exam construction and grading interfaces were concerned when the MVP was created and the hypothesis related to these components are derived from experience of creating the MVP combined with knowledge acquired in current state analysis phase.

The MVP was implemented as a course chapter in A+ LMS that utilizes similar exercises that students have done during the course and students can use same tools that they are familiar with. The exam was combination of

assignments similar to ones done earlier during the course: simulations and programming exercises and traditional exam questions, testing theoretical skills. The Figure 5.4 shows the first screen students saw when starting the exam.

The MVP was created by cloning course repository of the course in question containing course material including exercises, removing excess material reorganising it and drafting few new exercises. The process required quite a lot manual work because building the MVP was mostly copying old course material to new locations and modifying it. Building the MVP taught that the way used to construct the exam did not support reusing and versioning older exercises. An assumption was, that features to catalogue all available exercises and to easily generate new versions of exercises with different initial values would assist teachers in creating the exams and especially reduce the time required for the task.

The MVP was designed over several iterations with the responsible teacher for the course. After that content and usability and clarity of instruction was evaluated by two different course assistants. A+ LMS was chosen as a basis for implementing the MVP because it already provided means for delivering electronic assignments and also because most of the students are familiar with the system. During the design process it was concluded that the development work should be split into two completely independent research and development areas: the user interface of the A+ LMS for the examination use and the device, operating system and software tools used by students in the exam session. For this MVP it was chosen that presentation of the assignments and overall process flow of doing the exam was the most important priority. Other design aspects, for example security of used device and operating system, were not considered as important and so they were mocked for the MVP. Finally the MVP was tested with one student taking the exam in controlled environment. Because there was only one person taking the exam, the security and integrity of the exam session was easily fulfilled by author supervising the exam while also observing how the MVP was used.

One inconvenient usability issue found in the tests was that because of the size of the text boxes, there was substantial amount of scrolling back and forth required to revise related source code at the top of the page. While there are workarounds for this, e.g. opening same page to two tabs or windows and possibly utilising virtual desktops, these solutions might not be obvious to everyone and thus give unpleasant experience for not technologically oriented students.

The results of the user testing overruled some of the earlier hypothesis: it appeared that not at least every university student owns a laptop. The sample size of two was nowhere near representative, but this is an example what

kind of issues the current research setup might run into. Other observation during the user testing was that new software tools can have unexpectedly large impact on the usability. The tester needed a quick tutorial how to use the software tools provided for him. Even though this takes only couple of minutes, this scales very poorly and thus needs a solution to convey relevant information how to use provided system.

In the second MVP we managed to test the three original hypotheses but there were several more that will need further research, design and implementation. A1 represents the current idea how students are delivered with controlled and standardised exam environment. The fundamental idea is that invigilators supervise students booting the system so that USB drives are not given for students so that they will not be able to tamper with them. This solution would also reduce the required amount of USB drives to be maintained from several hundreds to maybe few dozens. The key assumption with this idea is that booting is fast enough so that setting exams does not take long time before the exam can be started. This idea could be validated with simulated scenario where one person acting as invigilator boots few dozen of laptops in a row and the elapsed time is measured. This gives a rough estimate how long it would take to get all laptops running with given amount of USB drives.

The A1 is however only one idea how the scalability and control can be ensured. While the study was already under way, our research team learned of parallel research project in Aalto University school of electrical engineering to solve a problem very close to ours: how to scale electronic examination for few hundred students simultaneously in lecture hall environment. Unfortunately we did not have chance to inspect their solution in the context of this thesis. It is recommended that in the future these projects coordinate together to effectively assign scarce development resources and to synergize on each other's work.

The hypothesis A2 and A7 differ quite much from the rest of the hypotheses. While the rest are new features or requirements for the system these two are prerequisites for the infrastructure. They are also uniquely difficult to assign priority: observing the current capacity is easy but on the other hand they may limit design space drastically for example if there are not enough electricity sockets for everyone taking the exam and fire regulations or capacity of electricity infrastructure prohibit the use of extension cords.

Another clearly different hypothesis is the A6 because its scope is so much smaller than for the others and because implementing it for the first testable implementation is going to require little to no effort.

5.3 Reflective Learning

This section presents key lessons learned from each applied tool in the action intervention phase.

5.3.1 Setting hypothesis

Hypotheses help to prioritize features. Forcing to analyse requested features from users' point of view, the process of setting hypotheses revealed insight of what needs to be designed and tested before the system would be ready to use.

Hypotheses are hard to simplify into one testable argument at least early in the project. At the start the scope of the project was huge and it felt difficult to isolate the most important hypothesis and ignore technological and organisational architecture required to eventually run exams in large scale.

Long iteration times do not corporate well with testing small, isolated changes. Partly due the project scheduling and partly due the restricted possibilities of testing, the iteration times were long during the project. This lead to only few iterations being tested so several hypothesis and design choices had to be tested at once.

Short description does not necessarily capture the complexity of the task. The sheer amount of listed hypothesis conveys very little information before entries have very different complexities and require varying amount of work to test. For example A6 is quite easy to test as a concept while A4 is likely going to take several iterations.

Formulating hypotheses instead of listing technical requirements forces deeper analysis and discussion. Stating that new features would make the examination system better is easy but it is more difficult to describe that benefit in measurable format. This forced us to consider benefits of features more thoroughly.

5.3.2 MVP

When discussing complex concepts, MVP helps to make implicit assumptions visible and so help discussing design options. The MVP utilised in the workshop proved to be especially valuable even if it was just a conceptual model of different solution ideas. On the other hand maybe exactly the conceptual nature of the MVP without any technical implementation helped to focus on fundamental design choices instead of for example being stuck with details of user interface. Also the MVP allowed inspecting

different aspects of electronic examination separately leading to a thorough analysis.

User testing in real use context emphasizes the parts of the solution that needs further attention. The test setup allowed use of workarounds that would not bode well in production environment. While the amount of these workarounds also clearly demonstrate all the aspects that need to be considered before solution would be ready for production, it also demonstrates that testing even critical systems is possible with right setup.

Mission critical nature of the system limits how agile the MVP can be. Because the test scenario was real exam for the students the MVP had to be somewhat polished and tested. The MVP required quite a lot of mocking features and infrastructure to run tests that would simulate real use scenarios.

5.3.3 Workshops

Group discussion reveals how differently people even in same position may view the process of arranging exams thus leading to misunderstandings. Discussing MVP together revealed small differences the teachers had about having exams and how they should be organised. This helped to understand how different meanings different people can have even if they use same words.

Workshops may be a good way to reach consensus over a important design choice but that may cause minority to be ignored. The goal of the workshop was to discuss available options and decide which of the solutions would best suit teachers. A consensus was achieved but it was not obvious if it was unanimous or if people kept to their contradicting views when they saw that they were in minority.

Workshops are a potential source of bias because of small sample size. While the goal workshop succeeded in meetings its goal of reaching consensus of solution concept to be tested, it has to be taken into account that four teachers involved are only small from all teachers at the department of computer science. Also all teachers had been involved in online education earlier so the sample might present too positive towards digitalisation of education.

ID	Hypothesis	How it is tested	Priority
O1	Students like the richer toolset offered by electronic examination	Exam taken on laptop with university's standard linux installation	Current MVP
O2	Interactive tools help students to demonstrate their learning	same as above	Current MVP
O3	Familiar environment of A+ LMS reduces exam anxiety	Implement old exam questions to A+ LMS	Current MVP
A1	Live OS can be booted to RAM in one minute from USB drive	Create image of the OS and simulate few dozens of boots	Waiting for collaboration
A2	University's wireless network can reliably support a couple of hundreds of connections simultaneously	To be decided	Low
A3	Teachers would like to construct exams same way as they construct course material	Implement new RST directives and exercise catalogue	Mid
A4	Teachers need tools to mass grade exams and assign exams for teaching assistants to be graded	Extension of current grading tools available in A+ LMS	High
A5	Students need designated exam mode in A+	Usability testing different layouts and limit navigation options	High
A6	When using standardised OS a cheat sheet of available softwares and functions help students in exam situation	A printed sheet how to use exam software	Mid
A7	Lecture halls have enough electricity sockets for everyone taking the exam	Discussion with university and capacity testing	Low

Table 5.1: The listing of current hypotheses for the examination system-

Chapter 6

Discussion

6.1 Requirements for the examination system

Only few iterations of the MVP was tested during the research process, so there is still much to study and test. However there are guiding findings that help with the future development. From the teachers' point of view the security and amount of required assignments for the exam are the most important characteristics. Hence based on current understanding, it's advisable to limit design space of the solution to options with limited time and space for the exam. In particular, the human supervision looks like the most convenient way to achieve required level of security for now because there is well established process between teachers and other faculty of the university and department to organise exams.

From the students point of view we found out that there are many delicate aspects affecting the usability of the system. While using A+ LMS was deemed as a neutral or positive design choice, the device, operating system and available software has to be picked carefully so that students are able to use them effectively when taking the exam. It was also found out that smaller than expected portion of the students might have their own laptops available to take into exam situation. This demands further discussion about the decision should the exam be taken on own or standardised device provided by the university. Addressing these aspects is important in the future to ensure that students feel confident and comfortable about the exam situation.

Electronic examination system is likely to require more detailed instructions for students. A requirement for this was discovered during the user testing.

6.1.1 Role of the exam in universities

The background study showed that exams serve three distinct roles in university education: first they are a form of assessment to measure knowledge and skills to assign grades equally. Second they enforce control to ensure that study credits and degrees are granted rightfully. Additionally exams provide the last chance to provide students feedback for their study efforts. Throughout the empirical study, the control aspect was emphasised and common view was that improvements on other two aspects should not be made if there is any trade-off with control.

The literature has pointed out that in computer science education electronic examinations and new assignment types enabled by them can possibly predict student success in further studies and as such be better measurement tool to assign grades. This claim is to be yet tested and our study did not have large enough sample to state anything about this. However our study showed that students might indeed benefit from familiarity between course assignments and exam questions possibly leading to better grades.

Electronic examination systems allow blending exam activity with other course activities, if systems manage to solve concerns with controlling student identities and prevent cheating and the benefits of giving better feedback and assessing students more accurately assess students are confirmed. In an attempt to further reduce how exams are detached from their application domain and other course activities exams might become indistinguishable from other learning activities and eventually exams as a concept may be abandoned altogether.

6.1.2 Electronic examination in regards with online learning

Figure 5.3 showed that examination systems are clustered towards controlled side of spectrum. To fully utilise the potential of online education the examination systems might need to be developed to allow more flexible environment for exams and thus better support online and blended education. This study relaxed only one of the three dimensions. Reason being that it allowed us to utilise existing processes better for faster testing. Also testing one perspective at time may better bring attention to special characteristics of each dimension. Below we discuss each of the three dimensions of online learning in more detail.

Our study found that human invigilators that come with controlled exam space are considered necessary by many teachers. Even EXAM system cannot solve the connection between control and space while it opens exam room

to be available any time to students with the use of technological control. Our literature didn't study solutions that control exam situation when exam is taken off-campus. The only referred study related to this kind of online learning was watermarking system by Simon [2005] that allowed the detection of one kind of cheating after the exam was submitted.

According to our study, when students take the exam and whether they have option to choose time when they take it or not does not affect the security of the exam as much as the other two components. However circulating exam questions between students or systematic cheating concerned teachers if students have extended schedule available when to take exam. While the setting was slightly different, Simon [2005] reported that students indeed try to use prohibited cooperation and share exam questions beforehand to gain advantage. One possible solution to battle this is having several alternative exam questions that are assigned to students randomly. This solution would in turn require both new features to A+ so that it can support this functionality and at least initially more effort from teachers, because they are likely going to need larger pool of exam assignments.

Use of own devices can be argued to be the most important aspect in online learning because if students are required to use devices provided by universities there would likely be some concerns that might be impossible to solve or at least would make arrangement very impractical. For example trust required to allow students to have university owned devices, tied capital and practicalities of transporting devices, just to name few. Even in situations where some of the dimensions of online education are controlled own devices can provide benefits. Based on our study we assume that familiarity of own devices would alleviate possible stress in exam situations because students trust that they know how to use their devices. On the other hand own devices pose maybe the biggest risks for control of the exam situation, because universities have very little ways to ensure that only allowed software is installed and that devices contain only allowed material. Additionally emphasis on own devices can drive inequality between students: those who may not have money to buy their own device may be forced to use random device offered by the university and thus be hindered in exam situation.

6.1.3 Subsystems of the electronic examination system

While building the second MVP, four distinct components for the electronic examination system were identified. The components does not map clearly with evaluation criteria found from the literature and presented in the Table 2.1. There are criteria like automatic marking and easiness of constructing exams that are clearly linked with a single component, grading interface and

teacher interface respectively in this case. However other criteria may not be that clear. For example security is mostly involved with operating system, software tools and network infrastructure, but also students interface plays a to fulfill that criteria.

6.2 Good practices of the lean startup

The lean startup and UCD practices chosen for this thesis were setting hypothesis, MVP and workshops. Setting hypothesis means that improvement ideas for the system or product are formulated as statements how the improvement would change the behaviour of customers and users instead of listing technical requirements. Hypotheses are characterised by being easily measured objectively so that expected change in behaviour can be verified. Ideally this measure data is collected directly from user behaviour when they interact with the system and not by secondary data collection methods like interviews. The MVP, Minimum Viable Product, is the smallest possible increment to the product which allows testing out the chosen hypothesis. MVP does not necessarily need to be functioning software. It can be anything that allows users to interact with the system eliciting the necessary knowledge of their needs and preferences. For example the first MVP used in this thesis, presented in Figure 5.2, allowed the author to investigate what kind of tacit limits teachers have for exam environments and what kind of exam settings they might be willing to test. Workshop is a facilitated collaborative work practice. Workshops were used in this research as interim user testing method because of schedule limits. Especially the hypothesis setting provided several insightful hypotheses even when the process was emergent and hypotheses were come up iteratively.

Promising lean startup practices left out from the study were at least vision and build-measure-learn loop. Vision could have provided valuable support for setting hypothesis with direction that would have persisted between iterations. Also build-measure-learn loop and pivoting are presented as key elements of the Lean Startup method and as such would have been interesting additions to studied practices. Build-measure-learn loop would have structured the development of MVPs better but it would have also required constant user testing that was unfortunately not feasible in this research. I believe that action similar to pivoting was done between MVPs but the process was not as explicit as specified by Ries [2011]. From the beginning there was an attitude that different options were kept open and focus was narrowed when there was more information available. The process was quite similar with pivoting but the same mindset has to continue when continuing

the project to avoid locking on a single solution.

6.3 The applicability of lean startup methods

In this section the applied Lean Startup methods are discussed and compared to existing literature and especially to other studies where these practices are tried. First the overall process chosen for the thesis is discussed and then each chosen practice in more detail.

The use of Lean Startup and User-centered design methods proved to be both challenging and valuable. A one of the fundamental issues was that there was not actually a team using those lean startup methods. This allowed a lot of implicit in methods and also cutting corners. The results of this study definitely need more validation from being used by larger teams. Overall lot of issues encountered when applying lean startup could be accounted for process deviating from original process presented by Ries [2011]. On the other hand good question is how easily large organisations can adapt their process completely and should there be some level of variation allowed.

The challenge of applying the planned design process related heavily to special characteristics of academic organisation and especially to application domain of examinations. One challenge was that courses run quite a long time and there are only few scheduled examinations when user testing could be organised. For this reason, iteration times between MVPs were quite long and there was also need to use other sources to simulate actual user testing. In this case we asked teaching assistants who were also students to review the exam. While they provided important observations, there were still lot of findings to reveal in the actual user testing because teaching assistants were more familiar with the domain of the exam and they likely had higher proficiency with the used software tools. The test setting also contained at least one flaw when the approximate user-testing was conducted: the exam environment wasn't fully simulated so the results were focused on the readability and clarity of the questions.

In our study the hypotheses were useful practice but slightly difficult to use. They helped in prioritizing tasks and keeping the user centered focus. At the same time hypotheses were little bit inadequate as a project management tools because of simplified structure that limits the amount of information conveyed. The benefits are completely in line with Ries. He states that hypotheses should indeed capture the most critical aspects of the system from the point of view of the users. However there were also difficulties in setting hypotheses. Sometimes ideas were hard to distill into short descriptions. Also short descriptions does not necessarily capture all

elements of the task especially if tasks are abstract or complex. Comparing to results by Khanna et al. [2018] there are some similarities. Because this thesis had long iteration times between MVPs the linkage between hypotheses and MVP blurred sometimes because several hypotheses were incorporated into one MVP. Additionally hypotheses would benefit from the use of other tools to manage the development work. List of hypotheses acts primarily as a task list but would require other tools, such as roadmap, to design and plan higher level aspects of the project.

The difficulties of setting hypotheses are also present in the Table 5.1. For example O1 is overly complex and difficult to understand because short description was tried to formulate so that it presents value to the user better. In reality O1 means that students prefer software tools like compilers and debuggers provided by electronic examination systems instead of producing source code with pen and paper.

The MVP helped in communicating the progress and abstract elements of the solution to users. This results is in line with the ones reported by Münch et al. [2013] who found that regular review sessions where MVP was discussed were important.

The official exam situation provided additional challenge for the user testing because the MVP had to be quite polished. The researchers were prepared for compensation in grade if there would have been any technical difficulties clearly limiting the students time or ability to answer given assignments.

6.4 Limitations of the study

The research problem of the thesis was to study the applicability of the lean startup practice in the context of education technology and examination system development. Action research methods requires applying studied methods in practice, the balancing between the development and observation of used methods. Because of this the analysis and reflection of used method were not done concurrently with the development and some insights may have been lost due to this. The applied practices were presented in literature as a holistic methodology without much room for compromises. However significant changes and limitations to the practices had to be made to fit the scope of this thesis. It may limit the validity of the study if studied lean startup practices have been used in the context that they were not intended to be used. One threat for the external validity of the thesis is that most of the research data was collected from a single source, a sample of four university lecturers. The usability testing with one student was not used mainly to validate the data collected earlier but to extend the perspective.

Chapter 7

Conclusions

This chapter first concludes the most important findings of this thesis and then suggests important and interesting research topics for future studies.

Electronic examination taken in lecture hall environment using students' own devices is a promising first step to extend the reach of electronic examinations to large courses. This approach was chosen because teachers wanted to hold on level of control provided by the traditional exams. This solution however is just a beginning on long journey of studying how growing use of online learning activities affect the role of examinations. Hypotheses helped also seeing the system from perspectives of different stakeholders that led directly to dividing the system into four subsystems.

Setting hypotheses and the Minimum Viable Product, MVP, supported the development work well by conceptualising and evaluating the complex domain area. The MVP was useful to elicit feedback and communicate complex ideas. Setting hypotheses helped in building the second MVP to recognise key assumptions from technical requirements. However just hypotheses were not sufficient project management tool when used alone. In the first iteration the MVP demonstrated well implications of each design option and linked them to a context familiar to teachers.

7.1 Future work

The most imminent continuation of this thesis is to keep on developing exam system and scaling tests with students gradually. Likely the most important component is students user interface, how the exam is presented in A+ learning management system. After that comes tools for grading exam because not all assignments are going to be automatically graded anytime soon. Third component is teachers' tools to construct exam quickly and allow easy use

of assignments from the course content and older exams. Concurrent to this development is recommended the development of examination infrastructure, OS and solution to control students' devices sufficiently. This development is recommended to be coordinated between other schools of the Aalto University and other polytechnic universities in Finland because our study found that there is already considerable research advancements in this particular area.

The tested lean startup practices need further studying as well. The results of this study can be validated further by extending the duration and the methods used in the study. It would be interesting to see how results evolve if more iterations are completed. Also introduction of additional lean startup practices is likely to enrich results of this study. For example vision, build-measure-learn loop and pivoting are all key topics in lean startup and so are likely to support the used design process well. One last interesting research topic for future studies is how studied practices suit the needs of a larger development team.

Bibliography

- A+ LMS. A+ lms public project page, 2019. URL <https://apluslms.github.io>.
- Aalto University. Aalto university examination guidelines, 2017. URL <https://into.aalto.fi/display/ensaannot/Aalto+University+Examination+Guidelines>.
- Aalto University. Aalto university online learning project page, 2019. URL <https://onlinelearning.aalto.fi/>.
- Abitti. Abitti examination system project page, 2019. URL <https://www.abitti.fi/fi/abitti/>.
- Kirsti M Ala-Mutka. A survey of automated assessment approaches for programming assignments. *Computer science education*, 15(2):83–102, 2005.
- David Arnow. Webtoteach project page, 1999. URL http://www.panix.com/~arnow/brooklyn_college/WebToTeach/.
- David Arnow and Oleg Barshay. On-line programming examinations using web to teach. *ACM SIGCSE Bulletin*, 31(3):21–24, 1999.
- David E Avison, Francis Lau, Michael D Myers, and Peter Axel Nielsen. Action research. *Communications of the ACM*, 42(1):94–97, 1999.
- Steve Blank. *The four steps to the epiphany: successful strategies for products that win*. K&S Ranch Press, 2005.
- Steve Blank. Why the lean start-up changes everything. *Harvard business review*, 91(5):63–72, 2013.
- Steve Blank and Bob Dorf. *The startup owner's manual: The step-by-step guide for building a great company*. BookBaby, 2012.

- Binglin Chen, Matthew West, and Craig Zilles. Do performance trends suggest wide-spread collaborative cheating on asynchronous exams? In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pages 111–120. ACM, 2017.
- Won Sun Chen and Adrian Yong Tat Yao. An empirical evaluation of critical factors influencing learner satisfaction in blended learning: A pilot study. *Universal Journal of Educational Research*, 4(7):1667–1671, 2016.
- Henry Corrigan-Gibbs, Nakull Gupta, Curtis Northcutt, Edward Cutrell, and William Thies. Deterring cheating in online environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(6):28, 2015.
- EXAM Consortium. Exam infrastructure description page, 2018. URL <https://wiki.eduuni.fi/pages/viewpage.action?pageId=64293094/>.
- EXAM Consortium. Exam project page, 2019. URL <https://wiki.eduuni.fi/display/csceexam>.
- Arto Hellas, Juho Leinonen, and Petri Ihantola. Plagiarism in take-home exams: help-seeking, collaboration, and systematic cheating. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, pages 238–243. ACM, 2017.
- Ville Karavirta, Petri Ihantola, and Teemu Koskinen. Service-oriented approach to improve interoperability of e-learning systems. In *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*, pages 341–345. IEEE, 2013.
- Dron Khanna, Anh Nguyen-Duc, and Xiaofeng Wang. From mvps to pivots: a hypothesis-driven journey of two software startups. In *International Conference of Software Business*, pages 172–186. Springer, 2018.
- Ari Korhonen and Jari Multisilta. Learning analytics. In Hannele Niemi and Jiyoun Jia, editors, *New Ways to Teach and Learn in China and Finland*, pages 301–311, Bern, Switzerland, 2017. Peter Lang. ISBN 978-3-631-69874-7. URL <https://www.peterlang.com/view/title/19509>.
- Sari Lindblom-Ylänne and Anne Nevgi. *Yliopisto-opettajan käsikirja*. WSOYpro Oy, 2009.
- Lauri Malmi, Ville Karavirta, Ari Korhonen, and Jussi Nikander. Experiences on automatically assessed algorithm simulation exercises with different resubmission policies. *Journal on Educational Resources in Computing (JERIC)*, 5(3):7, 2005.

- David V Mason and Denise M Voit. Integrating technology into computer science examinations. In *ACM SIGCSE Bulletin*, volume 30, pages 140–144. ACM, 1998.
- Beverly May. Applying lean startup: An experience report—lean & lean ux by a ux veteran: Lessons learned in creating & launching a complex consumer app. In *2012 Agile Conference*, pages 141–147. IEEE, 2012.
- M Dee Medley. On-line finals for cs1 and cs2. *ACM SIGCSE Bulletin*, 30(3):178–180, 1998.
- Joi L Moore, Camille Dickson-Deane, and Krista Galyen. e-learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, 14(2):129–135, 2011.
- Jürgen Münch, Fabian Fagerholm, Patrik Johnson, Janne Pirttilahti, Juha Torkkel, and Janne Jäärvinen. Creating minimum viable products in industry-academia collaborations. In *Lean Enterprise Software and Systems*, pages 137–151. Springer, 2013.
- Russell T Osguthorpe and Charles R Graham. Blended learning environments: Definitions and directions. *Quarterly review of distance education*, 4(3):227–33, 2003.
- Oxford English Dictionary. principle, n., 2018. URL <http://www.oed.com/view/Entry/151459?rskey=b3wT4Y&result=1>.
- Teemu Rajala, Erkki Kaila, Rolf Lindén, Einari Kurvinen, Erno Lokkila, Mikko-Jussi Laakso, and Tapio Salakoski. Automatically assessed electronic exams in programming courses. In *Proceedings of the Australasian computer science week multiconference*, page 11. ACM, 2016.
- Eric Ries. *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books, 2011.
- Neil C Rowe. Cheating in online student assessment: Beyond plagiarism. 2004.
- A. Rytönen and L. Myyry. Student experiences on taking electronic exams at the university of helsinki. In *EdMedia: World Conference on Educational Media and Technology*, pages 2114–2121. Association for the Advancement of Computing in Education (AACE), 2014.

- Simon. Electronic watermarks to help authenticate soft-copy exams. In *Proceedings of the 7th Australasian Conference on Computing Education - Volume 42*, ACE '05, pages 7–13, Darlinghurst, Australia, Australia, 2005. Australian Computer Society, Inc. ISBN 1-920682-24-4. URL <http://dl.acm.org/citation.cfm?id=1082424.1082426>.
- Jeffrey R Stowell and Dan Bennett. Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research*, 42(2):161–171, 2010.
- Pei-Chen Sun, Ray J Tsai, Glenn Finger, Yueh-Yang Chen, and Downing Yeh. What drives a successful e-learning? an empirical investigation of the critical factors influencing learner satisfaction. *Computers & education*, 50(4):1183–1202, 2008.
- Thierry Volery and Deborah Lord. Critical success factors in online education. *International journal of educational management*, 14(5):216–223, 2000.

Appendix A

Manuscript of the interview for current state analysis -phase (in Finnish)

Haastattelun tarkoituksena on kerätä aineistoa diplomityötä varten. Aineistoa tullaan käyttämään nykyisen tenttiprosessin analysointiin ja kehittämiseen. Kaikki vastaukset käsitellään anonyymeinä, eikä valmiista julkaisuista pysty erottamaan yksittäistä haastateltavaa.

Tarkoituksena olisi nauhoittaa tämä haastattelu. Nauhoituksia käytetään vain analyysin varmistamiseen ja kaikkia tallenteet tuhoaan analyysin jälkeen.

Esitiedot:

1. Mitä kurssia opetat?
2. Mistä kouluista opiskelijat ovat?
3. Mille vuosikurssille kurssi on suunnattu?
4. Kuinka monta opiskelijaa kurssilla on?
5. Entä kuinka moni suorittaa kurssin (lkm tai läpäisyprosessi)?
6. Kuinka pitkä kokemus sinulla on opettajana?
7. Kuinka kauan olet opettanut nykyistä kurssia?

APPENDIX A. MANUSCRIPT OF THE INTERVIEW FOR CURRENT STATE ANALYSIS -P.

Varsinaiset kysymykset:

8. Miten opiskelijoiden osaamisen arviointi järjestetään kurssilla?

- Kuinka oppilas saa arvosanan kurssilla?
- Mikä painoarvo tentillä on arvosanan muodostamisessa?

14. Miksi järjestät tentin?

10. Kuvaile miten kurssin tentit järjestetään?

(Mistä alkaa ja mihin päättyy?)

- opettajan näkökulmasta
- opiskelijan näkökulmasta

(hahmotellaan opettajan prosessi taululle vaikka postiteilla)

12. Kenen kanssa olet vuorovaikutuksessa tämän prosessin aikana

- Entä mitä järjestelmiä käytät?

11. Miten paljon yhden tentin järjestämiseen kuluu aikaa (kaikkien stakeholderien näkökulmasta)?

15. Mikä toimii mielestäsi hyvin nykyisessä prosessissa?

16. Entä mikä huonosti?

17. Mitä kohtia prosessista parantaisit, jos sinulla olisi riittävät resurssit?

13. Minkä tyyppisiä tehtäviä tentissä on?

- Mitä näistä sähköinen tentti tukisi?
- Onko joitain tehtäviä jotka "häviäisivät" sähköisen tentin myötä?
- Millaisia uusia tehtävätyyppejä sähköinen tenttiminen voisi mahdollistaa?

18. Millä tavalla sähköinen järjestelmä voisi parantaa nykyistä prosessia?

Kiitos haastattelusta. Tutkimus jatkuu siten, että analysoin tulokset ja kävisin mielelläni läpi tulokset kanssasi. Samalla, jos sinulle sopii, järjestäisin työpajan kaikkien haastateltavien ja muiden sidosryhmien kanssa tulosten läpikäymiseksi ja kehityskohteiden ideoimiseksi.