Doctoral theses at NTNU, 2021:301

Aparna Chirumamilla

Analysis of security threats, requirements, and technologies in e-exam systems

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Aparna Chirumamilla

Analysis of security threats, requirements, and technologies in e-exam systems

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-5579-3 (printed ver.)
ISBN 978-82-326-6341-5 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:301

Printed by NTNU Grafisk senter

To my mother

Abstract

Research context: The higher education (HE) sector is currently going through massive digital transformation by leveraging the use of technology to provide flexibility in teaching and learning. Increasing the usage of e-assessment is an integral part of digitisation in higher education. E-assessment presents several benefits over traditional paper-based assessment, including cost reduction, pedagogical improvements in assessments, and immediate feedback. Despite these benefits, the adoption of e-exams involves many challenges. Cheating has been an issue of concern for high stakes assessments. Another challenge has been lacking interoperability between e-exam systems and other supporting systems. Open digital ecosystems could achieve more flexible tool support for digital exams. However, development towards open digital ecosystems has been slow for many mainstream tools in the e-learning and e-assessment domain.

Research objective: This thesis aims to explore how e-exam systems can become key parts of an effective digital ecosystem for e-learning. We describe 1) functional features and 2) quality features for digital exams, mainly targeting key concerns of having sufficient security against cheating and satisfactory interoperability with related systems. The main research question and sub-questions for this thesis are:

RQ: How can e-exam systems contribute to achieving an effective digital ecosystem for e-learning?
SQ1: To what extent is the risk of cheating an obstacle to the adoption of e-exams, and how do e-exams compare to traditional pen and paper exams when it comes to cheating risks?
SQ2: What are the key requirements for e-exam systems, how are such requirements established, and how does the requirements process for acquisition and development of e-exam systems relate to approaches used for requirements in the field of software ecosystems?
SQ3: What are key obstacles towards achieving the interoperability needed for a digital ecosystem for e-exams and e-assessment?

Method: The research context of this thesis is framed between requirements engineering in software ecosystems, and e-exams in the higher education sector in Norway. This thesis consists of seven studies that present a systematic mapping review, threat modelling and risk analysis, penetration testing, case studies, and mixed-method research. Qualitative data is collected through interviews. Quantitative data is collected through surveys.

Results: Through the implementation of seven studies, we came up with five contributions through seven papers (P1 – P7):
C1: Improved understanding of cheating threats and countermeasures in paper exams versus e-exams and empirical findings on perceptions of teachers, students, vendors, and managers about such threats and countermeasures.
C2: A review of issues and potential research gaps in requirements engineering for software ecosystems through a systematic mapping review, producing essential

findings concerning requirements engineering activities and non-functional requirements for software ecosystems.
C3: Empirically grounded descriptions of the requirements process surrounding acquisition and development of e-exam systems in Norwegian higher education.
C4: Description based on empirical evidence of key features for e-exam software according to vendors, process managers, and higher education institutions in Norway.
C5: Identification of enablers and barriers for achieving open digital ecosystems for e-exams within a larger ecosystem of e-learning.

Conclusion: Our literature review and studies indicated that cheating had been a big concern towards the adoption of e-exams in the higher education sector, not only in Norway but in many universities around the world. Our findings indicated that e-exams have additional cheating threats. However, they also provide additional countermeasures against cheating, so they need not be less secure than traditional paper exams. Our empirical results suggest that using open digital ecosystems could reduce many challenges with security and interoperability between e-exam systems. Yet, the implementation of the digital ecosystem in the e-learning and e-assessment domain is still immature, and both vendors and customers prioritise new functional features and security against cheating higher than requirements for interoperability.

Preface

This thesis is submitted for partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D.) in the Department of Computer Science (IDI) at the Faculty of Information Technology and Electrical Engineering (IE) at the Norwegian University of Science and Technology (NTNU).

The PhD work was performed at the Department of Computer Science, NTNU, Trondheim. Professor. Guttom Sindre was the main supervisor, while Professor John Krogstie and Professor Monica Divitini were co-supervisors.

Acknowledgements

I must thank many people who stood by me during my PhD journey. First and foremost, I would like to thank my supervisor, Guttom Sindre, for giving me the opportunity to do a PhD under his guidance. This thesis would not have been possible without his best support. I am highly thankful to my co-supervisors, John Krogstie and Monica Divitini, for their guidance and support.

I also thank Anh Nguyen Duc, Thea Marie Søgaard, Shang Gao for co-authoring some of my works. I especially thank Anh Nguyen for his encouragement and support along the way. Special thanks to Randi Holvik and Camilla Thun Waaden for their support and quick assistance.

We received data from surveys and interviews for this PhD work. I thank all survey and interview respondents who spent their precious time participating in our studies. Especially I would like to thank domain experts from Unit, Inspera AS, UNIWise, NTNU, UiT, HVL, Kristiania University College for their continued availability in answering my questions during my studies.

I extend my gratitude to all my colleagues at the Department of Computer Science (IDI) and Information Systems and Software Engineering (ISSE) group.

A big thank you to my friends Rinku Singh, Shweta Tiwari, Sruti Subramanian for their encouragement and support.

Finally, my deep and sincere gratitude to my family for their love and support. My life partner, Praveen Kumar Chirumamilla, for his patience and continued support throughout the ups and downs of my PhD. My mother, brother, and in-laws for their support. Lastly, my little boy, Divit Parthiv, for his best support to make this PhD happen.

Aparna Chirumamilla,

Sep 2, 2021

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API	Application Programming Interface
ADTrees	Attack Defense Trees
AT	Attack Trees
BYOD	Bring Your Own Device
E-assessment	Electronic assessment
E-exam	Electronic exam
E-learning	Electronic learning
ESB	Enterprise service bus
FEIDE	Felles Elektronisk IDEntitet
FS	Felles Studentsystem
HARM	Hacker Attack Representation Method
ICT	Information and Communications Technology
LMS	Learning Management System
LTI	Learning Tools Interoperability
NSD	Norwegian Center for Research Data
NTNU	Norwegian University of Science and Technology
PCI	Portable Custom Interaction
QTI	Question and Test Interoperability
RE	Requirements engineering
SEB	Safe Exam Browser
SECO	Software Ecosystem
SLR	Systematic literature review
SMR	Systematic mapping review
Unit	the Norwegian Directorate for ICT and Joint Services in Higher Education and Research

Glossary

Assessment	Assessment (also called as test or examination or exam) refers to the process of evidencing and evaluating the extent to which a candidate has met or made progress towards the assessment criteria (JISC, 2006).
BYOD	Bring Your Own Device (BYOD) refers to the practice of students using their personal devices for educational purposes, in the educational institution or at home, after configuring it with the required settings in a couple of quick steps. In this thesis, the BYOD exam is concerned with the e-exam done on student-owned laptop in the university campus under the supervision of invigilator (also called proctor).
E-assessment	E-assessment (also called electronic assessment, digital assessment or online assessment or computer-based assessment) is an assessment activity that involves the use of computing devices.
Formative assessment	Formative assessment evaluates the actual level of students learning throughout the course and gives the student feedback to aid improvement (Dolin, Black, Harlen, & Tiberghien, 2018).
Summative assessment	Summative assessment (also called e-exam) provides information about what learning outcomes have been achieved by students at a certain time (Dolin et al., 2018).
E-exam	E-exam (also called electronic exam, digital exam, online exam or eExam) is the timed, supervised summative (final) assessments conducted via computing devices.
Item Bank	A storage facility for items which allows them to be maintained and used for automatic and manual test generation purposes (to create tests on-paper and/or on-screen). Today, almost all item banks are electronic although historically many were physical (JISC, 2006).
Cheating	Any intentional action or behaviour that violates the established rules governing the administration of a test or the completion of an assignment. cheating gives one student an unfair advantage over other students on a test or an assignment and decreases the accuracy of the intended inferences arising from a student's performance on a test or an assignment (Cizek, 2004).

Safe Exam Browser	Safe Exam Browser is a web browser environment to carry out e-assessments safely. The software turns any computer temporarily into a secure workstation. It controls access to resources like system functions, other websites and applications and prevents unauthorised resources from being used during an exam. In this thesis, SEB exams involved with e-exams that run based on configured settings on student-owned laptops under the supervision of invigilators.
FLOWlock	In WISEflow e-exam system, you make a flow based on what kind of exam you're holding, e.g., FLOWlock, FLOWassign. FLOWlock is the one type of flow for conducting secure exam, and it works like safe exam browser.
Digital ecosystems	A digital ecosystem is a business ecosystem based on an organisational network in the context of digital technology. Digital ecosystems are formed on the basis of digital objects (digital content, products, ideas, software, hardware, infrastructure) that are interchanged and shared between independent actors.
Software ecosystem	The interaction of a set of actors on top of a common technological platform that results in a number of software solutions or services (Manikas & Hansen, 2013). A broader concept than software ecosystem is digital ecosystem, which along with the software products found in a software ecosystem, also includes hardware and digital content interchanged and shared between several providers.
E-learning ecosystem	The learning community, together with the enterprise, united by a learning management system (LMS). It is formed by three categories of components: content providers, consultants, and infrastructure (Uden, Wangsa, & Damiani, 2007).
E-exam ecosystem	E-exam ecosystems (also called digital ecosystems for e-exams) is the intersection of the platform providing the basic functionalities (e.g., question delivery to candidates, collection of answers, marking) with plugins taking care of a plethora of more specialised needs (e.g. authoring or answering of specific question types in various speciality domains, i/o for students with special needs, specific grading schemes, advanced support for grading, etc.). One could consider e-exam ecosystem not as a separate digital ecosystem in its own right but rather as part of a bigger ecosystem for e-learning. One of the main goals for digital ecosystems for e-exams is to enable plug-ins for learning systems.

IMS Global	IMS Global Learning Consortium is a collaborative group of affiliates, including hardware and software vendors, educational institutions, publishers, government agencies, system integrators, and multimedia content providers.
Interoperability	The ability of two or more systems or components to exchange information and to use the information that has been exchanged (Geraci, Katki, McMonegal, Meyer, Lane, Wilson et al., 1991).

Part I: Synopsis

1. Introduction

1.1. Motivation and Problem Statement

The higher education (HE) sector is currently going through massive digital transformation by leveraging the use of technology to provide flexibility in teaching and learning (Selwyn, 2014). This digitisation affects every core activity of the universities, including education, research, and administration, and automates many previously labour-intensive business processes (Abad-Segura, González-Zamar, Infante-Moro, & Ruipérez García, 2020). Consequently, it helps teachers be more self-directed and allow students to get practical knowledge and skills before entering the job market (Bond, Marín, Dolch, Bedenlier, & Zawacki-Richter, 2018).

Increasing the usage of e-assessment is an integral part of digitisation in higher education, both for formative assessment throughout the semester, and for summative assessment which results in a grade, for instance, by means of an exam at the end of a course. Here, we use the term e-exam to cover high stakes graded tests done by digital devices, while e-assessment more generally covers any assessment activity using digital devices. E-exams present several benefits over traditional paper-based exams. One claimed benefit is cost reduction both by avoiding the printing and transportation of question papers (James, 2016) and simplification of administrative tasks surrounding the assessment, e.g. creating and delivering tests, collecting answers, and logistics related to planning and grading exams (Dermo, 2009). Of course, there may be other costs instead, especially if the university needs to equip huge PC labs for the exams, but less so if student-owned devices are used for e-assessments (Hillier & Fluck, 2013). E-assessments are often commended for pedagogical improvement in assessments, including immediate feedback, possibility to randomise questions and answers, and increased opportunity to use auto-scored questions, yielding more consistency and fairness through marking (Appiah & Van Tonder, 2018). E-assessments also offer enhanced question styles that include interactivity, multimedia, and greater flexibility for location and timing (Boyle & Hutchison, 2009).

Despite the benefits mentioned above, the adoption of high stakes e-exams involves many challenges. While many other processes in the HE sector have long since been digitised, including low stakes formative assessment activities where students typically deliver their answers through the university's learning management system (LMS), the usage of e-assessment for high stakes exams has been lagging behind (Hillier & Fluck, 2013). Several reasons have been reported for this reluctance towards adopting e-exams: scalability (universities have scarcity of resources, e.g., exam halls with desktop PCs), reliability, authentication (different countries have different regulations, some favouring students' privacy), training the staff and students, and appropriateness. For some disciplines, assessments are much based on projects and assignments, and in maths and engineering education, e-assessment tools have often had lacking or clumsy functionality for writing equations or making design diagrams, hence making pencil and paper preferable. However, in many universities, high stakes exams stuck with pen and paper, even in disciplines where typing on a keyboard would be clearly more efficient, such as plain essay writing or

programming. For example, the first-ever e-exam in the introductory Python programming course (course code TDT4110) at the NTNU took place in December 2017, prior to that, it had been pen and paper. A key concern causing reluctance towards adoption of e-exams – not just at the NTNU but in many universities around the world – is security, especially fears of cheating (Appiah & Van Tonder, 2018).

Cheating has long been an issue of concern with high stakes assessments (Cizek, 1999). Successful cheating is fundamentally unfair, creating advantages for cheaters over honest students. A grade achieved by cheating will not be a valid representation of the candidate's competence in the subject matter, and awarding degrees to incompetent candidates will ultimately damage the reputations of educational institutions (McCabe, Treviño, & Butterfield, 2001). Especially with e-exams done with student-owned devices (Bring Your Own Device, or BYOD), mitigating cheating would be more difficult (Dawson, 2016; Ifijeh, Michael-Onuoha, Ilogho, & Osinulu, 2015). Another reason behind slow adoption of e-exams was the need for training of staff and invigilators (Sim, Holifield, & Brown, 2004). Although e-exams may be monitored using technological surveillance (e.g., webcams, monitoring tools used together with e-exam systems, online proctoring), there is often also a need for effective human invigilators. However, most universities in Norway hire retired staff temporarily as invigilators during the exam season. Many of these have limited knowledge of IT and would thus likely be much less competent in mitigation of cheating for e-exams than for traditional paper exams. Hence, transitioning to e-exams might require a considerable amount of time for training the invigilators. Simultaneously, it requires training for the academic and administrative staff on how to design and run exams with the e-exam tool. The assumed cost savings from e-exams would not be immediate. In the transition period, there will likely be some initial years where the university spends more on exams than before due to investments in software and infrastructure, training of personnel, and changes in administrative routines. As new technologies arise, educators and institutions will probably be aware of cheating threats after students' uptake.

Another challenge has been lacking interoperability between e-exam systems and other supporting systems (Chituc & Rittberger, 2019). True, paper does not offer good interoperability with computerised information systems either, so poor interoperability of e-assessment tools may not be a key argument for sticking with paper. However, poor interoperability may cause a lot of double work, e.g., re-entering of data, meaning that the administrative simplifications and cost savings that one hoped to achieve from e-exams, may not materialise (or be smaller than expected). Together with limited functional features, poor interoperability may also cause pedagogical hindrances, forcing students and teachers to adapt learning and teaching activities to what the IT systems can support, rather than having the systems adapt to the preferred pedagogy. There are several challenges pertaining to achieving interoperability. A major issue is that universities will require the e-assessment systems to be integrated with supporting existing systems (Jakimoski, 2016). However, it is difficult for universities to ensure the level of interoperability of the e-assessment systems during acquisitions (Sclater, 2007). For interoperability to be ensured, vendors need to develop e-assessment systems using open standards and governance frameworks so that many different systems can collaborate smoothly

in a digital ecosystem (Uden et al., 2007). However, complexities and ambiguities in the standards are obstacles to achieving this vision in the e-learning domain (Chituc & Rittberger, 2019; Piotrowski, 2011).

1.2. Goal and Research Questions

Based on the problem outlined in the previous section, the purpose of this thesis is to investigate how e-exam systems can become key parts of an effective digital ecosystem for e-learning, where the actors would collaborate on a common technological platform for developing e-learning systems. Not just having the necessary functional features for digital exams, but also addressing the key concerns of having sufficient security against cheating, and satisfactory interoperability with related systems. The main research question and sub questions for this thesis are:

RQ: *How can e-exam systems contribute to achieving an effective digital ecosystem for e-learning?*

SQ1: *To what extent is the risk of cheating an obstacle to the adoption of e-exams, and how do e-exams compare to traditional pen and paper exams when it comes to cheating risks?*

SQ2: *What are the key requirements for e-exam systems, how are such requirements established, and how does the requirements process for acquisition and development of e-exam systems relate to approaches used for requirements in the field of software ecosystems?*

SQ3: *What are key obstacles towards achieving the interoperability needed for a digital ecosystem for e-exams and e-assessment?*

Our main research question addresses the effectiveness of an e-learning ecosystem. A review by Noesgaard and Ørngreen (2015) revealed that many different indicators of effectiveness had been used in e-learning, e.g., learning outcomes, performance, student or teacher satisfaction, usage of the product, etc., among them the most common indicator used to define effectiveness was 'learning outcome'. Many of the studies reviewed by Noesgaard and Ørngreen looked at only one or a few competing software products. The issue of effectiveness for an ecosystem, looking at a huge and dynamically changing set of partly competing, partly collaborating products and resources, will be even more complex. For the students and teachers, learning outcomes and satisfaction will still be key concerns, but there will also be issues about the interoperability between products, e.g., how easy it is for a new product to enter and get integrated into the ecosystem, how easy it is for stakeholders to switch from one product to another, and then also security, where it is important to find the right balance: too little security will hurt effectiveness because security breaches will be detrimental to the normal operation of the system. On the other hand, too much security may also hurt effectiveness, as systems with a lot of security barriers often become much more cumbersome to use. Given the complexity of e-learning ecosystems, effectiveness of such systems may be hard to define. Yet, as for any system, effectiveness comes down to how well it accomplishes its mission. For an e-learning ecosystem, the ultimate mission is to help students achieve their learning

outcomes – and Noesgaard and Ørngreen (2015) also found that although many different indicators had been used for effectiveness, learning outcomes was the top criterion concerning number of reviewed studies that used it. At the same time, from the perspective of teachers and other university staff, there will also be indicators of effectiveness related to how well the ecosystem supports their work. Similarly, there may be indicators for system managers and product vendors, e.g., related to the ease of integrating new products into the ecosystem. For the purpose of this thesis, we go with the following definition: Effectiveness of an e-learning ecosystem means how well it supports the students in achieving learning outcomes and how well it supports other stakeholders in constructively contributing to such achievement.

The first sub question aims at providing an analysis comparing cheating threats and countermeasures on controlled e-exams versus paper-based exams. Many universities are in the transition from traditional pen and paper exams to e-exams. As a result, the comparison between these exam types is gaining much attention. Researchers and educators have done considerable research comparing different factors, substantially addressing students' and teachers' perceptions on e-exams and paper exams. However, there is a need for a clear-cut view of particular advantages that e-exams would bring compared to paper exams concerning prevention and detection of cheating. Therefore, this thesis aimed to summarise comparisons between e-exams and paper exams, with a particular focus on proctored Bring Your Own Device (BYOD) e-exams where students use their own laptops or desktop computers for the examination in controlled settings.

The second sub question addresses key requirements for e-exam systems, both investigating what the requirements are, and by what process they are established and prioritised, in the interplay between customers (i.e. universities) and vendors of e-exam systems. While there are several publications stating requirements for e-exam systems on a conceptual level, particularly related to benefits that e-exam systems might bring compared to traditional paper exams, there has been much less published research about empirical investigations on how such requirements are captured and features are prioritised. Given that such empirical investigations are time-consuming, we cannot investigate this on a global scale, so this part of the investigation will look specifically at the HE sector in Norway. To relate SQ2 to the main RQ, and to the international level, we will however study international literature about requirements engineering for software ecosystems and consider how the empirically discovered requirements approaches for Norway's e-exam systems relate to requirements approaches proposed in the literature. Understanding the requirements engineering process (SQ2) also sheds further light on SQ1 – how are cheating concerns addressed in this process? and on SQ3 – to what extent are requirements developing in the direction of open digital ecosystems for e-learning. For connection with the other sub-questions, the investigations of SQ2 also pays special attention to requirements for interoperability and security, and whether requirements engineering methods and techniques used in the software ecosystems field could address challenges in e-exam system development.

The third sub question focuses on digital ecosystems for e-learning and e-assessment. Such ecosystems were optimistically proposed more than a decade ago, e.g., by Uden et al. (Uden et al., 2007), but although successful ecosystem

applications in e-learning exist (García-Holgado & García-Peñalvo, 2016), the development towards open digital ecosystems has been slow for many mainstream tools in the e-learning and e-assessment domain. This question is particularly aimed at identifying what might be hindering the development towards open ecosystems for e-exams – again with a special focus on the tooling for e-assessments in higher education in Norway. Understanding the obstacles towards achieving open digital ecosystems is an important step towards improving the situation and achieve more flexible tool support.

The Covid-19 pandemic caused an increased uptake of e-assessment as lockdowns prevented on-campus gatherings of students. Hence, many exams that would previously have taken place on campus (some already digital, some using paper), with proctors, were instead transitioned to online take-home exams, often with little or no proctoring, leading to increased concerns about cheating (Bilen & Matros, 2021; Lancaster & Cotarlan, 2021). Given that cheating is one of the main topics of this thesis, it may seem strange that there is no sub-question specifically about experiences with e-exams during the Covid-19 lockdown. Indeed, this thesis says very little about the concerns that emerged during the Covid-19 lockdowns. However, there are several reasons for this:

- The bulk of the research for this thesis was performed before Covid-19 struck, including all the empirical research for the included papers in part II. Although paper P6 was published in 2020, the data gathering through questionnaires and interviews were done in 2019, and the exam context investigated in that paper was proctored on-campus exams. Even for P7, most of the interviews were done in the autumn of 2019 and early 2020, before the Covid-19 lockdown (which started 12 March 2020 in Norway). The last couple of interviews for the study in P7 were done in March and April after the Covid-19 lockdown had taken effect. However, for consistency with previous interviews, it did not make sense to change the interview guide to have the last couple of respondents reflect on the Covid-19 lockdown. Moreover, even if the lockdown had then taken effect, no exams had not yet been held under the lockdown regulations since the spring term end-of-course exams in Norway typically take place from mid-May to mid-June, so the informants would not yet have had any experience with the altered exam-practices.
- As we have no empirical data and no substantial results about exam experiences during the pandemic lockdown, it did not seem appropriate to include much about the Covid-19 lockdown in the thesis introduction either, whose purpose is to summarise and synthesise the contributions from the collected papers.

Moreover, if there was increased cheating in exams resulting from the Covid-19 adaptations to exam practices, this need not be due to a shift from paper to e-exams. It could just as well be due to a shift from proctored on-campus exams to unproctored

take-home exams. Our purpose in SQ1 was to compare paper exams and e-exams under otherwise equal conditions (e.g., same type and level of proctoring).

1.3. Study Landscape

The investigation in this thesis has mainly conducted through seven studies (S1-S7) (see Figure 1). These studies have been conducted in collaboration with colleagues, and one of them with a master student at NTNU. We will cover more details on the studies in Chapter 3. A brief overview of seven studies is presented below.:

S1: A interpretive literature review (2015) focusing on the comparison between e-exams and paper exams.

S2: A case study (2016) conducted on the SEB lockdown browser

S3: A systematic mapping review (2014 – 2016) on software ecosystems related requirements engineering activities and quality attributes.

S4: A interpretive literature review (2018) focusing on the comparison between on-campus exams and remote exams.

S5: A case study (2019) at NTNU, based on our own practical experiences with the Inspera Assessment e-exam system and Blackboard learning management system. This study also included an analysis of documentation related to other e-assessment tools.

S6: A mixed-method research (2018 – 2019) consists of surveys and interviews with engineering students and teachers at NTNU.

S7: A case study (2019 – 2020) with the Higher education sector in Norway consists of interviews with e-exam systems tool vendors, system managers, process managers from different organisations.

1.4. Research Contributions and Papers

This section gives an overview of the papers included in this thesis. This thesis is based on seven peer-reviewed publications. Of the seven published papers, three are journal articles, two conference papers, one workshop paper, one book chapter. The collected papers are as follows[1]:

P1: Sindre, G., & Vegendla, A. (2015). *"E-exams versus paper exams: A comparative analysis of cheating-related security threats and countermeasures."* In Proceedings of Norwegian Information Security Conference (NISK), 8(1), 34-45. https://ojs.bibsys.no/index.php/NISK/article/view/298

P2: Vegendla A., Søgaard T.M., & Sindre G. (2016). *"Extending HARM to make Test Cases for Penetration Testing."* In Proceedings of Advanced Information

[1] It should be noted that in the midst of the project, the PhD candidate's last name changed due to marriage, hence being Vegendla for Papers 1-4 and Chirumamilla for Papers 5-7.

Systems Engineering Workshops: CAiSE 2016, vol 249, 254-265. Springer, Cham. https://doi.org/10.1007/978-3-319-39564-7_24

P3: Vegendla, A., Duc, A. N., Gao, S., & Sindre, G. (2018). *"A Systematic Mapping Study on Requirements Engineering in Software Ecosystems."* Journal of Information Technology Research (JITR), 11(1), 49-69. https://doi:10.4018/JITR.2018010104

P4: Vegendla, A., & Sindre, G. (2019). *"Mitigation of Cheating in Online Exams: Strengths and Limitations of Biometric Authentication."* In Kumar, A. (Ed.), Biometric Authentication in Online Learning Environments. 47-68. IGI Global. http://doi:10.4018/978-1-5225-7724-9.ch003

P5: Chirumamilla A., Sindre G. (2019). "*E-Assessment in Programming Courses: Towards a Digital Ecosystem Supporting Diverse Needs?"* In Proceedings of Conference on e-Business, e-Services and e-Society (I3E), vol 11701, 254-265. Springer, Cham. https://doi.org/10.1007/978-3-030-29374-1_47

P6: Chirumamilla, A., Sindre, G., & Nguyen-Duc, A. (2020). *"Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway."* Assessment & Evaluation in Higher Education, 45(7), 940-957. https://doi.org/10.1080/02602938.2020.1719975

P7: Chirumamilla A., & Sindre G. (2021). *"E-exams in Norwegian Higher Education: Vendors and managers views on requirements in a digital ecosystem perspective."* Computers & Education, 104263. https://doi.org/10.1016/j.compedu.2021.104263

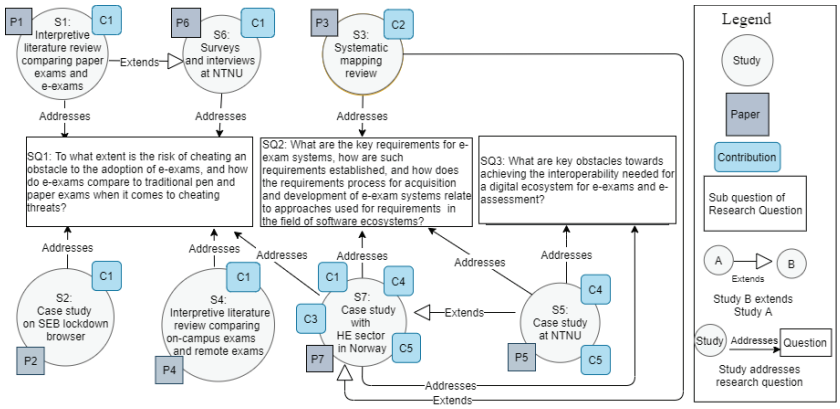

Figure 1. Research papers connection to studies, research questions and contributions.

The research contributions for the thesis are provided as below:

C1. *Improved understanding of cheating threats and countermeasures in paper exams vs. e-exams and empirical findings on perceptions of teachers, students,*

vendors, and managers about such threats and countermeasures. This includes analysis of various threats through threat modelling, results from surveys and interviews with students and teachers related to their perception of various cheating threats and countermeasures, further including insights from vendors and managers.

C2. *A review of issues and potential research gaps in requirements engineering for software ecosystems through a systematic mapping review, producing essential findings concerning requirements engineering activities and non-functional requirements for software ecosystems.* This review aims to show the major topics addressed in requirements engineering in software ecosystems and address potential research gaps in the literature.

C3. *Empirically grounded descriptions of the requirements process surrounding acquisition and development of e-exam systems in Norwegian higher education.* This includes results from a case study with vendors and managers about their experience with the requirements engineering process during procurement and development of e-exam systems.

C4: *Description based on empirical evidence of key features for e-exam software according to vendors, process managers, and higher education institutions in Norway.* This will present different key features considered by university staff, procurement managers, and vendors during the interviews.

C5. Identification of enablers and barriers for achieving open digital ecosystems for e-exams within a larger ecosystem of e-learning. This contribution will show how the digital ecosystems exist within the e-exams and what are considered enablers and barriers to open digital ecosystems for e-exams.

The relation between the papers, research questions, studies and contributions are represented in Figure1.

1.5. Thesis Structure

This thesis is organised as follows.

Part I: Following the current introduction, Part I includes the theoretical background and related work, research approach, results, discussion, conclusions and suggestions for future work. These sections are covered in the following chapters.

Chapter 2: Provides the theoretical background and related work.

Chapter 3: Describes the research approach used for this thesis.

Chapter 4: Presents the results by describing the papers attached to this thesis.

Chapter 5: Discusses the results of the thesis with respect to research contributions, implications, limitations, and evaluation of the PhD research work.

Chapter 6: Concludes the thesis and outlines some ideas and suggestions for future work.

Part II: Provides the collection of seven research papers included in this thesis.

2. Background and Related Work

This chapter provides a background and related work for the topics discussed in this thesis. The chapter is divided into four sections. Section 2.1 gives a brief background for e-assessment and e-exams, interoperability in e-learning, and an overview of e-assessment infrastructure in Norway's higher education sector. Sections 2.2 – 2.4 present related research on three main topics connected to the thesis research questions, 1) cheating and assessment security (SQ1), 2) key requirements for e-exam software (SQ2), and 3) digital ecosystems for e-exams (SQ3).

2.1. Background

2.1.1. Background on e-assessment and e-exams

The term assessment in this thesis refers to the process of evidencing and evaluating the extent to which a candidate has met or made progress towards the assessment criteria (JISC, 2006), establishing whether the student has achieved the intended learning outcomes of a module, course or degree program. The discussion of assessment often separates between two types of assessment - formative and summative. While the primary purpose of formative assessment is seen as assessing the actual level of students learning throughout the course – and give the student feedback to aid improvement - summative assessment provides information about what learning outcomes have been achieved by students at a certain time (Dolin et al., 2018), often for the purpose of grading. Hume and Coll (2009) consider '*assessment of learning'* as summative assessment and '*assessment for learning'* as formative assessment. Taras (2005) distinguishes between formative and summative assessment, but at the same time observes: *"it is possible for assessment to be uniquely summative where the assessment stops at the judgement. However, it is not possible for assessment to be uniquely formative without the summative judgement having preceded it."* (p.4). I.e., just like summative assessment, formative assessment also needs to know whether there is a gap between the intended learning outcomes at a certain point and the student's actual learning outcomes, since without knowledge about this gap, it is also hard to give advice for improvement.

Improving the quality of assessments is a key issue in the higher education sector. To achieve quality, assessments need to be valid and reliable. Assessments will be reliable if the performance gets the same grade independent of which censor is grading it, and the order of grading (Moskal, Leydens, & Pavelich, 2002). Assessment validity depends upon constructive alignment between intended learning outcomes, learning activities, and assessments (Biggs, 2003; Biggs, 2011). Threat to the validity occurs when learning outcomes, activities, and assessments are misaligned, leading to *construct-irrelevance* and *construct underrepresentation*. Here, construct underrepresentation means that some of the learning outcomes are not sufficiently addressed by the assessment (Downing, 2002), whereas construct-irrelevance means that there are factors beside achievement of the intended learning outcomes that will affect the grade (Haladyna & Downing, 2004).

Successful cheating might be one example of a construct-irrelevant factor. For some students, then, the grade is determined not based on their competence in the subject matter, but due to lack of scruples towards academic dishonesty and ability to cheat without being caught. Cizek (2004) defines cheating as *"Any intentional action or behaviour that: violates the established rules governing the administration of a test or the completion of an assignment; cheating gives one student an unfair advantage over other students on a test or an assignment and decreases the accuracy of the intended inferences arising from a student's performance on a test or an assignment."* Often cheating in academics is referred to together with the related term *academic dishonesty,* as opposed to *academic integrity.* Academic integrity tended to focus on the importance of integrity, especially about citing sources and on the awareness of honour codes (McCabe, Trevino, & Butterfield, 1999), whereas academic dishonesty includes cheating, fraud, and plagiarism, the theft of ideas, and other forms of intellectual property (Jones, 2011). There can be cheating by students, by university employees, or others, and it can take place before an assessment activity (e.g., leaking questions supposed to be confidential), during the assessment activity (e.g., using forbidden aids during an exam), or after the activity (e.g., illegitimately altering grades). We focus the investigations in this thesis to cheating during exams, not before or after the exam. The rationale for this choice is that cheating before or after the exam is less dependent on the choice of paper or PC as the medium for student's answer.

Electronic assessments are referred to by different terms depending upon how they are organised (Crisp, 2011; JISC, 2006). We define e-assessment (also called electronic assessment, digital assessment or online assessment or computer-based assessment) as *assessment activity that involves the use of computing devices.* Here, the computing device can be a laptop, desktop computers, or any other devices that use ICT to create, deliver and store assessments, report grades and feedback. Likewise, we define e-exam *(also called electronic exam, digital exam, online exam or eExam)* as *the timed, supervised summative (final) assessments conducted via computing devices.* Typically, an e-exam has the questions delivered to the student via a computing device, and the student also answers via such a device. Even with paper exams, some aspects or the process were digitised – typically the authoring of paper exams prepared in PC, using a word processor, although it was thereafter printed and handed out to the students on paper. Also, for the grading, digital tools may have been used for a long time, e.g., the teacher using a spreadsheet to record partial scores for various exam tasks and calculate grades for the candidates. Paper exams with Multiple Choice Questions could also have the students answer on paper forms that were then optically read to calculate scores automatically. Still, these would not be called e-exams. Some e-exams require one or more tasks that have students drawing design diagrams or solving math proofs and equations on paper and then scanning these documents to submit digitally. As long as such an exam also contains tasks that are answered digitally, it would likely be considered an e-exam, although also having some paper-based tasks.

There are different types of e-exams, e.g., depending on the location where the exam takes place (campus or home), equipment used (university equipment or BYOD), and degree of proctoring, e.g., from entirely unproctored to strictly proctored, and

the latter either by a human proctor, automated monitoring, or both. The use of tools (e.g., LMS or e-exam system) for e-assessment may vary from university to university (Martin, Lopez-Martin, Moreno-Pulido, Meier, & Castro, 2019). For instance, many American universities use Blackboard or Canvas as their learning management system (LMS) and even for e-exams. For high stakes exams, they may supplement the LMS with the lockdown software, e.g., Respondus LockDown Browser, to safeguard against cheating (Cluskey Jr, Ehlen, & Raiborn, 2011).

2.1.2. Interoperability in e-learning

The IT support needed for e-learning/e-assessment will likely require integrations and interoperability between e-learning/e-assessment systems with several supporting systems. Interoperability is defined by IEEE as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged." (Geraci et al., 1991). When two systems interoperate closely, we say that they are integrated (Lauesen, 2006). Interoperability is important to ensure effectiveness in exchanging and sharing information, aligning and orchestrating collaborative processes, and establishing decisions or policies (Daclin, Daclin, Chapurlat, & Vallespir, 2016). A typical symptom of poor interoperability is a waste of resources due to double work, as employees may have to re-enter manually information that was already captured digitally in another system, because of inability to transfer the data automatically. Non-integrated data sources may also cause decisions to be made based on poorer information than what should really have been available. To facilitate the integration of different systems, vendors provide application programmers' interfaces (APIs). A major obstacle to interoperability could be that different systems use different interfaces and data formats. Hence these systems should adhere to common specifications and standards. The e-learning/e-assessment tools that have been key to this PhD research use standards from IMS Global Learning Consortium (IMS, 2021). IMS Global Learning Consortium is a collaborative group of affiliates, including hardware and software vendors, educational institutions, publishers, government agencies, system integrators, and multimedia content providers. IMS has two fundamental goals: (1) to define specific guidelines which guarantee interoperability between applications and services in e-learning, (2) to support the application of guidelines in international products and services (Bianco, De Marsico, & Temperini, 2005). The three standards used by the e-learning/e-assessment tools central to this research are:

- Question and Test Interoperability (QTI)
- IMS Learning Tools Interoperability (LTI)
- IMS Portable Custom Interaction (PCI)

Question and Test Interoperability (QTI): This standard describes a basic structure for the representation of question (item) and test (assessment) data and their corresponding results reports. QTI is meant to enables the sharing of questions, tests and results data between different e-learning systems, for instance so that question and test data exported from one system can be imported to another system (Wills, Davis, Gilbert, Hare, Howard, Jeyes et al., 2009).

IMS Learning Tools Interoperability (LTI): This standard allows external tools to be launched within an application (Queirós, Leal, & Paiva, 2016; Severance, Hanss, & Hardin, 2010), e.g., from LMS to e-exam application. For an example of practical usage, assume a teacher for pedagogical reasons prefers to use in formative or summative assessment a certain question type that is not supported by that university's LMS or e-exam system X. However, the e-learning application Y has good support for this question genre. By using LTI to launch Y within X, the teaching staff could enable students to do an entire test or exercise as if working seamlessly in X, rather than having to start two different tools separately and juggle between them.

IMS Portable Custom Interaction (PCI): This standard enables the users to create new question types and plugins to their e-learning system. The IMS Global defines Portable Custom Interaction (PCI) as *"a standard way for technology-enhanced items (TEIs) or custom interaction types to be represented as part of the Question and Test Interoperability (QTI) and Accessible Portable Item Protocol (APIP) specifications*[2]*."*

Overall, this section provided a brief background on interoperability in e-learning/e-assessment systems. The overview of the e-assessment infrastructure in Norwegian higher education is further discussed in sec 2.1.3.

2.1.3. The e-assessment infrastructure in Norway's higher education sector

Universities in Norway have Blackboard or Canvas as learning management systems (LMS), and students deliver ungraded tests and coursework through such an LMS. However, in Norwegian universities, the LMS is not used for graded tests or coursework. Instead, a dedicated e-exam system is used, currently either Inspera Assessment[3] or WISEFlow[4], together with an external lockdown browser depending on assessment requirements. The Ministry of Education and Research of Norway has created a national Directorate for ICT and joint services in higher education and research, in brief named Unit[5]. This directorate is tasked with acquiring and developing joint ICT solutions for various tasks that higher education institutions typically have, and have for instance run joint acquisitions of e-exam systems for public universities and colleges. Unit currently manage Norwegian HE institutions' dialogue with e-exam system vendors. They also have responsibility for developing and maintaining a custom system of the HE sector, named FS (Felles Studentsystem), a Student Information System used by almost all higher education institutions in Norway.

---

[2] https://www.imsglobal.org/assessment/interactions.html#:~:text=Portable%20Custom%20Interaction%20(PCI)%20defines,%C2%AE%20(APIP%C2%AE)%20specifications (Accessed May 18, 2021).

[3] https://www.inspera.com/

[4] https://www.uniwise.co.uk/wiseflow

[5] https://www.unit.no/en

The architecture diagram in Figure 2 shows the systems and tools that are used in the Norwegian HE sector, which was the focus of the empirical part of the research in this thesis involving two e-exam systems and several other supporting tools used for storing student's information, assessments, and authentication, with links indicating information exchange. FS (second from left) contains authoritative information about students (e.g., personal information, enrolment, course registration, exams scheduled, grades received, etc.), courses, teachers, etc. StudentWeb (left) is a front-end to FS where students can register or withdraw from courses and exams, get information on the time and location of exams, view and appeal grades, etc. The learning management systems Blackboard (used by NTNU) and Canvas (used by most other Norwegian universities) handle communication within courses, such as announcing the course reading list, time and place of lectures, and providing learning materials like slides from lectures, instruction videos, and ungraded weekly exercises. Both Inspera Assessment (e.g., being used by NTNU, University of Oslo, University of Bergen) and WISEFlow (e.g., being used by University of Tromsø , University of South-Eastern Norway, Western Norway University of Applied Sciences, Kristiania University College) are proprietary software products, run as cloud services using lock-down browsers (top and bottom) to mitigate cheating. Further to the right are some other systems involved, the document archival system (ePhorte), the single-sign-on authentication, FEIDE or ID-porten (used with several systems, but we only show links to the e-exam systems to avoid messing up the diagram), and the plagiarism checking tool, where Norwegian HE currently uses Urkund.

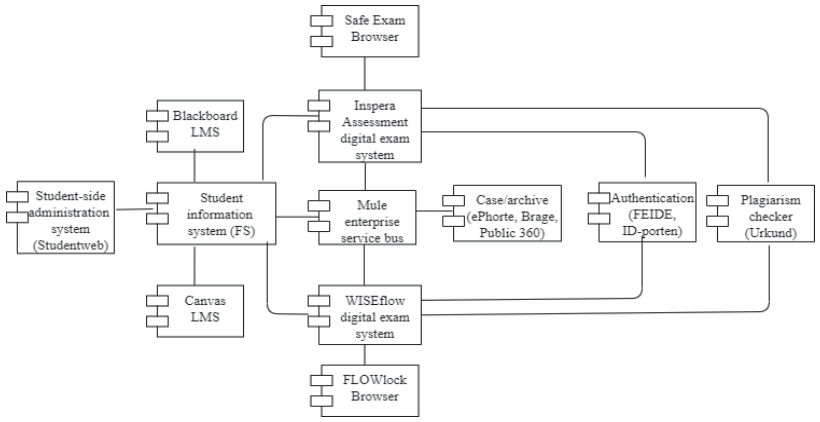

Figure 2. Exam solutions interfaces [Adapted from (Melve & Smilden, 2015)]

Overall, this section provided the background for this thesis. In the following sections 2.2 – 2.4, we provide related work for this thesis.

2.2. Cheating and assessment security

Cheating in connection with school and university exams has been a topic of research for hundred years already (Bird, 1929; Carter, 1928). It is beyond the scope of this thesis to present a full review of this literature. Cheating can happen at various stages, such as before, during, or after the test itself (Cizek, 1999). Before the test, a key problem is the leakage of confidential test questions. After the test, two typical modes of cheating are corrupt grading and illegitimate altering of answers post-delivery. The main focus in this thesis is, however, cheating during the test. Hence, we do not provide any detailed review of cheating before or after the test. The main motivation for focusing on cheating during the test is that this has the most likelihood of having some differences between e-exams and traditional paper exams. Cheating before or after the test, such as bribing teachers to leak questions or grade favourably, will more likely be the same regardless of the type of exam. The larger part of the research in cheating has focused on cheating by students (McCabe et al., 2001; Whitley, 1998), but there has also been research on cheating by university employees such as teachers or school administrators (Jacob & Levitt, 2003). Regarding employee cheating, most publications have focused on leakage of questions before the exam (Volante, 2004) or corrupt grading after the exam (Borcan, Lindahl, & Mitrut, 2014), though there have also been publications discussing employee cheating during the exam (Ettien, 2018). However, the main focus in this thesis is cheating by university students during exams.

There is a huge body of research on student cheating on exams, and it is beyond the scope of this thesis to give a full overview. Broadly, we group our discussion of related work in the following categories:

- Empirical research to discover and quantify actual cheating and evaluate the actual effect of mitigations
- Empirical research on stakeholder perceptions on cheating and mitigations, for instance, questionnaire or interview studies with students or teachers
- Analytical and design-oriented research on possible ways to cheat, and possible approaches to mitigation

For each of these categories, we will cover some general literature, then focus on the research specific to e-exams or comparison of e-exams and paper exams and discuss how the contributions in our thesis relate to this body of research.

2.2.1. Research to discover actual cheating

Empirical discovery of actual cheating can be performed either by directly observing and capturing students in the act of cheating, or by looking for evidence of cheating in the delivered answers after the exam. Since invigilators or teachers will often catch only a smaller fraction of students who actually cheated (Cerimagic & Hasan, 2019), researchers would have to establish a much more elaborate surveillance scheme to get useful data on the actual frequency of cheating by direct observation. This would make direct observation a very costly research approach and easily entail legal and ethical problems. Hence, the more common approach in this line of research is to

look for evidence of cheating in delivered answers or group performance statistics. Fendler, Yates, and Godbey (2018) compared the exam performance of two groups of students with different seating arrangements (free seating versus randomly assigned) on a multiple-choice test. The finding was that the free seating led to a higher degree of copying of answers, presumably because friends then took care to sit close to each other, or would-be cheaters took care to sit close to students known to be clever. Walker (2010) investigated the amount of plagiarism in delivered student work. Harmon and Lambrinos (2008) compared the performance in a proctored on-campus test with an identical online test without proctoring, finding indication of cheating based on significantly better performance on the online test.

Specific to e-exams: There are mainly two potential advantages of e-exams in the detection of cheating. First, the discovery of plagiarism is much easier with an electronic delivery than with handwritten answers. Second, e-exam systems can automatically register a lot more data than just the delivered answer, e.g., potentially also timing data for submitting each task – or in the extreme case, for every keystroke and mouse click during a test. Hellas, Leinonen, and Ihantola (2017) used not only plagiarism checking of the delivered code, but also pairwise comparisons of submission times of the various subtasks, to establish likelihood of cheating in a take-home Java programming exam.

Relation to our research: This thesis did not aim for observation or measurement of actual cheating, so the relation to our research is limited. Besides, our focus was mainly on proctored on-campus exams, not take-home exams – which limits the relationship with the above-mentioned works on take-home tests and plagiarism. However, the impact of seating arrangements, as explored empirically by Fendler et al. (2018), does corroborate our similar analytical claim in P1 – further validated via questionnaires to students and teachers in P6 - that specific seating patterns could be a countermeasure against the cheating threat of peeking at the answers of neighbour candidates. Our suggestion took the idea of seating one step further – to mixed seating – meaning that a student taking an exam in Programming would be surrounded by students taking other exams – say, Physics, Psychology, English Literature – whereas the approach of Fendler et al. was to suggest randomly assigned seating rather than free seating, but still within the same course. If a student has no close neighbour with the same exam, this might be assumed to reduce cheating by peeking or close-range collaboration in the exam hall even further. Of course, seating arrangements could be thought of as a generic countermeasure where it does not matter whether you have an e-exam or paper exam. However, an important point made in P1, and further explored in P6, is that e-exam systems may be an enabling technology even for some countermeasures that appear purely physical. A reason mixed seating is rather rare in large exam halls, is that it would complicate the logistics if invigilators could not walk down row by row with a large pile of copied exam questions (paper exam) to put on the desks, but rather had to take care to distribute question sets in an intricate pattern. Also, if the exams had different time limits, it would be disturbing to candidates in the longer duration exam X when the chief invigilator shouts out that writing must cease on the shorter exams Y, Z. With an e-exam there is no need to distribute paper, as each student might automatically get the appropriate question set on screen after authenticating, and there is no need

for invigilators to shout out information about timing, since the e-exam system could time out automatically and auto-save answers for any students who have not yet submitted when time is due. Hence, it might be easier to implement assigned mixed seating with e-exams than with paper exams.

2.2.2. Research on stakeholder perceptions of cheating

There is a huge body of research, especially on student perceptions of cheating, and also on perceptions of other stakeholders such as teachers and administrators. A common approach is questionnaire studies exploring the stakeholders' attitudes about cheating, own experiences with cheating (students), or the discovery of cheating (teachers), possibly combined with interview studies for increased understanding. Compared to studies observing actual cheating as mentioned in section 2.2.1, questionnaire and interview studies have the disadvantage of relying on self-reporting by students. As cheating is a potentially punishable behaviour, and some students are also uncertain exactly what constitutes cheating, there will be a tendency for under-reporting (Burrus, McGoldrick, & Schuhmann, 2007).

We present the research on stakeholders perceptions in the following order:

- How many students cheat
- Why students cheat relating to attitude towards cheating
- In what ways students cheat
- Approaches for mitigating cheating

How many students cheat – and whether cheating is on the rise, especially focussing on gender and discipline, e.g., Omonijo (2012) led survey with 199 students from three universities in Nigeria (who were actually caught while cheating with ICT tools in examination halls), focusing on differentiating e-cheating between science students and non-science-oriented students, male and female. The results from his study indicate that there is a significant difference. ICT students were more engaged in cheating than non-science students, and male students were more involved in cheating than female students. Teixeira and Rocha (2010) conducted a survey with 7213 economics and undergraduate business students from 42 universities located in 21 countries that mainly aimed to differentiate the magnitude of cheating through copying during onsite summative exams. Their results revealed that propensity to copy ranged from 5%, the lowest, in universities located in the Scandinavian countries (Denmark and Sweden) to 88% in the universities selected in the Eastern European countries (Poland, Romania, and Slovenia).

Why students cheat relating to attitude towards cheating, e.g., Carpenter, Harding, Finelli, Montgomery, and Passow (2006), conducted a survey with 643 engineering and pre-engineering undergraduates at 11 institutions, mainly focusing on why students cheat. Their results indicated that students were unaware of the distinction between cheating and permitted behaviour.

In what ways students cheat, e.g., Bernardi, Baca, Landers, and Witek (2008) performed a survey with 417 business students from Australia, China, Ireland, and the US. Their study shows that copying or exchanging in exams, tapping codes, and

bathroom notes apply to all exams regardless of the field. Especially cheating would be easier when multiple-choice and/or true-false questions are used. Trost (2009) carried out a questionnaire survey with Swedish university students, and her study results show that taking forbidden aids to examination (9%) was common while peeking (1%) less common, and no respondents reported impersonation and student–staff collusion. De Lambert, Ellen, and Taylor (2006), in their survey with New Zealand institutions, report that students and staff believed impersonation is rare while forbidden aids and peer collaboration appeared to be more common. Colnerud and Rosander (2009) carried out a questionnaire survey with students considering both assignments and examinations. They reported that students considered collaboration with another student leaving crib notes in restrooms as one of the most obvious cheating practices during examinations. Some studies have looked at the situation in academia in general, e.g., questionnaires with a huge number of respondents, while others have looked specifically at the situation within one university or one type of study program. An example of the latter by Sheard, Dick, Markham, Macdonald, and Walsh (2002) looks specifically at IT students, which is particularly interesting to us since many of our informants in P6 also came from similar study programs. Their results show that copying through swapping assignments with a friend are common.

Approaches for mitigating cheating. Related to the Sheard et al. (2002) study mentioned above, a follow-up-study 10 years later (Sheard & Dick, 2011) indicated that their university had been able to reduce several types of cheating using various mitigation approaches, including revised cheating and plagiarism policy, raising awareness on academic integrity by introducing courses related to cheating and plagiarism into the curriculum, paying greater attention to the structure of assessment, e.g., conducting oral interviews to help reduce cheating. The study by Bernardi et al. (2008) with business students also indicates that using different question sets and scrambling the questions on these sets, and using more essay examinations would deter cheating. Mellar, Peytcheva-Forsyth, Kocdar, Karadeniz, and Yovkova (2018) conducted an exploratory case study with teachers about their perceptions on using TeSLA (an Adaptive Trust-based e-assessment System for Learning) e-authentication system for both on-campus and remote exams. Their results indicate that though teachers felt that e-authentication technology was not the primary mitigation approach in addressing cheating and plagiarism, but rather they saw it as major element to be used together with other mitigation approaches. Specifically, teachers thought that the TeSLA system could mitigate impersonation threat, but still they thought that it might not prevent accessing forbidden aids and assistance/collaboration during e-assessments in both proctored on-campus exams and un-proctored remote exams. Further, they provided a number of approaches to mitigate cheating: using of technology for lockdown and automatic logging, effective assessment design, increased surveillance, implementing security policies.

On the other hand, some researchers pointed out potential downsides to technological surveillance using assessment security tools. Dawson (2020) claimed that technologies used for assessment security establish large databases of student work that could be used for different purposes than improving academic integrity. However, this threat was assumed to be lesser earlier as the invigilators need to report

back to the institutions. He also indicated that routine checks on entire student populations rather than specific students suspected of cheating (suspicionless surveillance) would increase the workload for staff by more cases of suspected cheating. Ross and Macleod (2018) argue that technology-mediated practice of plagiarism detection force students to comply with it as well as create distrust among students, teachers and institutions. Introna (2016) argues that surveillance technologies turn the issue of cheating and assessment security into an algorithmic problem with technical rules. A survey by Woldeab and Brothen (2021) show that surveillance might be unfair, being hostile not only to students who want to cheat but also hurting the performance of students with high exam anxiety. Gilliom and Monahan (2012) provide more insights on concerns with surveillance in general (e.g., through performance monitoring technologies, social media, at workplaces), including some examples from academics.

Specific to e-exams: There has not been much research into the amount of cheating specifically in e-exams or whether there is more cheating in e-exams than in traditional pen and paper exams. There has been research related to perceptions of stakeholders on e-exams. Jamil et al. (2012) carried out a questionnaire survey about teacher perceptions of e-exams and paper exams. However, their study did not explicitly compare perceptions towards cheating in paper exams and e-exams. They have one question in their survey whether easier shuffling of items in e-exams could reduce cheating. Their results indicate that more than 62% of the teachers perceived shuffling as effective countermeasure against cheating. There has been research explicitly looking at cheating in online exams compared to on-campus exams, such as the study by Harmon and Lambrinos (2008). However, the key comparison here was between a proctored test and an unproctored test, while our preferred comparison would be e-exam versus paper exam, but everything else being the same (e.g., both having the same level of proctoring).

Relation to our research: Our research did not at all focus on why students cheat or their moral attitudes about various ways of cheating. Nor did we try to establish any figure for the actual frequency of cheating – although the questionnaire study in P6 asked students whether they had own experiences cheating. Our results indicated a much lower percentage of cheating than what has been found in many other studies, but as also admitted in the paper, our results were based on self-reporting from a rather small population of students and may have suffered from under-reporting, and P6 did not claim any findings related to the overall frequency of cheating.

Our research mainly focused on various ways of cheating, countermeasures against cheating, and the comparison of such cheating threats and countermeasures between e-exams and paper exams. The part of the research that looked at stakeholder perceptions was reported in P6, and the key contribution was on the perceived ease of various types of cheating in e-exams versus paper exams, and the perceived viability of various countermeasures against cheating. Our findings are in line with (De Lambert et al., 2006; Trost, 2009), related to ease of cheating in exams in general, and corroborate with suggested mitigations in Jamil et al. (2012). However, these papers only provided stakeholders perceptions of various cheating methods and mitigations. They did not provide comparisons specific to e-exams and paper exams

similar to what we did in P6. Thus, our research could add to their contribution by explicitly providing comparisons between e-exams and paper exams.

2.2.3. Analytical and design-oriented research

Unlike the research in 2.2.1 to discover actual cheating, or in 2.2.2 to have students, teachers, or other stakeholders report on their perceptions and opinions about cheating, the research in this subsection is more about analysing how students *might* cheat – for instance, using threat analysis or risk analysis, or via penetration testing of e-exam systems to discover vulnerabilities that might be exploited for cheating, and possibly also to analyse or even design solutions that can mitigate such cheating.

We present some research that has analysed cheating threats and/or proposed mitigations based on the following criteria:

- Categorisations of cheating threats and mitigations.
- Penetration testing of e-exam tools.
- Proposal of best practices for mitigating cheating in exams.
- Design of systems to mitigate cheating.
- Research on stakeholder perceptions of e-authentication system.

Categorisations of cheating threats and mitigations. Küppers et al. (2020) analysed cheating threats using attack defense trees (cf. Section 3.3.4), specifically comparing e-exams and paper exams. Their threat analysis was focused on specific cheating threats - impersonation, assistance/collaboration, and the use of unauthorised aids for exams, pointing out that e-exams are more secure than paper exams. Rosmansyah, Hendarto, and Pratama (2020) presented various countermeasures for impersonation attack using an attack defense tree model. Meland, Bernsmed, Frøystad, Li, and Sindre (2019) conducted experiments with security graduate students and security experts comparing two different threat modelling methods - misuse case diagrams and Bow-tie diagrams. In their study, digital exams is used as an exemplar case. Their results show that threat modelling methods were helpful in finding cheating threats for e-exams.

Penetration testing of e-exam tools. Dawson (2016) presents five potential hacks against BYOD e-exams: copying contents of the USB stick to hard disk, running the exam on a virtual machine, USB key injection, modifying e-exam software, and cold boot attacks. The five attacks were tried out in practice through testing on various BYOD e-exam tools. From his analysis, he reports that bringing unauthorised aids to the exam, removing the examination paper from the venue, and receiving live assistance from outside experts as significant threats to BYOD exams. He stated that BYOD exams have all the vulnerabilities of paper exams and exams conducted in university PCs, as well as some additional vulnerabilities of its own. Rosmansyah, Ritonga, and Hardi (2019) analysed various cheating threats and mitigations for e-exams using an attack defense tree and further evaluated ADTree through penetration testing against a server running the e-exam application.

Proposal of best practices for mitigating cheating in exams. Buzzetto-More and Alade (2006) provided best practices for e-learning and e-assessments in general.

Grünigen, Souza, Pradarelli, Magid, and Cieliebak (2018) presented best practice guidelines for lecturers to design secure e-assessments, suggesting a variety of question types for different formats (whether exam is conducted in BYOD devices or university PCs) of e-assessments and different surveillance methods. Williamson (2018) addressed best practices for overcoming the challenges of cheating for his teaching course exam.

Design of systems to mitigate cheating. The TeSLA[6] project (An adaptive Trust-based e-assessment System for Learning) is a Horizon 2020 project and is one of the worldwide initiatives for design of secure authentication system for e-assessment in online and blended educational settings, involving a consortium of 18 expert organisations (both technological and educational institutions) from 12 countries (Kocdar & Dirkx, 2017; Noguera, Guerrero-Roldán, & Rodríguez, 2016). The TeSLA system involves several instruments such as face recognition, voice recognition, keystroke dynamics, forensic analysis, and plagiarism tools for authentication and authorship checking in e-assessment. Especially, the TeSLA system was designed to connect or integrate with already existing learning environments through both LMS plug-ins and LTI connectors. Baró-Solé, Guerrero-Roldan, Prieto-Blázquez, Rozeva, Marinov, Kiennert et al. (2018) reported design and implementation experiences by members of the TeSLA project during the preparation of the pilots of the project.

Research on stakeholder perceptions of e-authentication system: Okada, Noguera, Alexieva, Rozeva, Kocdar, Brouns et al. (2019) conducted a mixed-method study examining the concerns and practices of teachers who used the TeSLA system in six countries (UK, Spain, the Netherlands, Bulgaria, Finland and Turkey). Their findings revealed some technological, organisational, and pedagogical issues related to accessibility, security, privacy and e-assessment design and feedback. They provided some recommendations: providing technical FAQs for teaching staff to deal with problems, and an audit report with results of each instrument with guidelines for university staff to interpret results and ensure quality of e-assessment, to raise awareness about data security and privacy, to develop policies and guidelines about fraud detection and prevention, e-assessment best practices and course team support. (Okada, Whitelock, Holmes, & Edwards, 2019) conducted a mixed-method study investigating the attitudes and experiences of 328 students who used TeSLA. Their findings suggest a broadly positive acceptance from distance education students. As for some notable findings, including men were less concerned about providing personal data than women, middle-aged participants were more aware of the nuances of cheating and plagiarism, while younger students were more likely to reject e-authentication.

Relation to our research: In this thesis, we did analytical research for P1, P2, and P4. When we conducted our threat analysis, there was relatively little work available on threat modelling, specifically for e-exams. Our threat analysis in P1 and P4 presented various cheating threats for e-exams through impersonation, assistance/collaboration, and the use of unauthorised aids. We explored similar

[6] https://tesla-project-eu.azurewebsites.net/

threats and countermeasures as discussed by Küppers et al. (2020) and Rosmansyah et al. (2020). While our research in P1 focused specifically on cheating during the exams, Küppers et al. (2020) also focused on cheating after the exams. Indeed, Küppers et al. and Rosmansyah et al. improved upon our attack tree models. For instance, for verifying candidate ID, we suggested biometrics, whereas Küppers et al. suggested public key infrastructure and Rosmansyah et al. suggested monitoring as mitigation. In P2, we conducted penetration testing of an open-source lockdown browser using HARM modelling as a method to analyse threats. This is different from the penetration testing approach taken by Rosmansyah et al. (2019), who used the Acunetix web scanner to find the vulnerabilities in an e-exam system whereas our testing approach was analytical based on experiments.

Our findings from P4 are in line with (Mellar et al., 2018), indicating that biometric authentication was one of the elements for mitigating cheating, however, it would be most effective if used together with other cheating mitigation approaches such as effective assessment design and increased surveillance. Admittedly, our study was not directly related to perceptions of teachers, rather more from the analytical perspective.

2.3. Key requirements for e-exam software

Many researchers have come up with a number of requirements to e-exam systems through case studies with universities and e-exam systems vendors, studies with students and teachers concerning usage of e-exam systems, or via design research proposing their own prototypes for such systems. We group our discussion of related work for key requirements for e-exam software in the following categories:

- Case studies about the piloting and usage of e-exam systems,
- Design research projects proposing their own e-exam applications or proposing more generic requirements frameworks for such systems.
- Works that explicitly discuss concerns and features for security and cheating mitigation of e-exam systems, and similarly publications that specifically consider integration and interoperability for such systems.

For each of these categories, we will cover research specific to e-exams and discuss how the contributions in our thesis relate to this body of research.

2.3.1. Case studies about piloting and usage of e-exam systems

Several papers discussed usage of e-exam systems through case studies. Kuikka, Kitola, and Laakso (2014) compared three LMSs (Moodle, Optima, and ViLLE) and two e-exam systems (Soft Tutor, Tenttis) through a survey with teachers in TUAS, Finland. From their analysis, they suggested a few essential features required for teachers to ease the introduction of e-exams. Fluck (2019) provided several key features for e-exam tools, based on in-depth interviews with 17 academic staff from five countries (Australia, UK, Finland, Iceland, Austria).

Table 1. Similarities and differences of findings from P7 with some of the existing case studies

Author / Feature or requirement	Kuikka et al. (2014)	Fluck (2019)	Fluck, Adebayo, and Abdulhamid (2017)	Wibowo, Grandhi, Chugh, and Sawir (2016)	Fitzharris and Kent (2020)	P7
Logistics management	*					*
accessibility		*			*	*
authoring						*
Marking	*	*		*		*
feedback	*	*		*		
Question analytics	*					*
Learning analytics					*	
Scalability			*			*
Usability	*	*	*	*	*	*
Integration, interoperability	*					*
Security		*	*	*		*
Reliability	*	*	*	*		*

Fluck, Adebayo, et al. (2017) compared two e-examination systems from Australia and Nigeria, and came up with some requirements for e-exam systems. Wibowo et al. (2016), conducted a pilot study of the ExamPro e-exam system with students and academic staff in Australia through focus groups, to improve the future adoption of e-exams. Fitzharris and Kent (2020) conducted a case study with students from Brunel University, London. Their study mainly focused on learning analytics and presented a few requirements for e-exam systems in that respect. Overall, these studies were typically eliciting viewpoints of students and teachers who are end-users of e-exam systems to inform further development and operation of such systems. Further, these papers reported various advantages for e-exam systems - auto-marking, enhanced scalability, usability, reliability, and security.

Relation to our research: Our case study in P7 provided findings about participant views about key features of e-exam systems, and how these were supported by the two e-exam systems (Inspera Assessment, WISEflow) that are being used in Norway. Table 1 compares the above-listed case studies and our empirical findings

in P7 (rightmost column in Table 1). Several of the above-listed papers did not address some of the basic features (e.g., authoring, integration, and interoperability) considered important by participants in our study. Except for Kuikka et al. (2014), none of the other papers focused on the logistic feature that enables additional planning for e-exams, e.g., for students to reserve time to take the exam without teacher involvement. Similar to Kuikka et al., our findings in P7 also show that logistics support is an essential feature for e-exams. Kuikka et al. (2014) addressed interoperability. However, their study did not discuss other aspects of interoperability, such as integration with other applications that universities typically have. Fitzharris and Kent (2020) performed analysis on the same system (WISEflow) as we did, but their research centered on learning analytics, which has been not discussed in P7 and other papers. However, among all features, P7 has not explicitly discussed the 'feedback' feature, as this feature has been considered as a straightforward feature for e-exam systems by participants in our study. Overall, Table 1 shows that most of the papers considered usability, marking, security and reliability as the most required qualities and features for e-exam systems.

2.3.2. Design research

Several researchers have reported the design and implementation experiences with e-exam systems, with combinations of in-house development and already available software. Fluck, Pálsson, Coleman, Hillier, Schneider, Frankl et al. (2017) discussed some of the benefits and barriers for e-exams conducted in university-PC exams and/or BYOD exams, proposing design decisions towards usability, reliability, and security of e-exams. Adebayo and Abdulhamid (2014) conducted mixed-method research with students and staff from Nigerian universities, to detect security issues in the existing e-exam system, further to propose a secure e-exam system. Brink and Lautenbach (2011) reported experiences with an in-house developed e-assessment system in a South African University. Isaias, Miranda, and Pífano (2019) made a framework of eight evaluation criteria for e-assessment systems: variety of design options, scalability, security, access and usability, feedback features, personalization, cost, and interoperability. The framework was validated by a questionnaire gaining responses from academic staff across 37 countries. Among the criteria, the highest level of agreement was for variety of question types and feedback features and for interoperability. When it comes to more general frameworks, Striewe (2019) based on a literature review, proposes components (front-end components, educational components, knowledge representation and storing components, and connector components) and design alternatives for e-assessment systems for each component, pointing out the features that the component would be supposed to cover. Khlifi (2020) proposes a prototype to ensure continuous authentication of students during exams, which collects information of students and their behaviour during course activities and uses them to generate authentication questions for students to respond to during the exam. (Okada, Noguera, et al., 2019).

Relation to our research: Our research in P7 discussed several features and requirements for e-exam systems. However, it does not present any comprehensive design effort, so that the relationship may be weaker than for studies in the previous

paragraph. Still, the empirical findings in P7 could suggest improvements to the features of Inspera Assessment and WISEflow e-exam systems that are being used around the world (Though, issues with some of the features pointed out in P7 might have changed already by now, with new releases of those products). Our results from P7 are consistent with the framework suggested by Striewe (2019). The features highlighted by our participants have parallels in all four major components proposed in the framework by Striewe, including front-end components (user interface), educational components (underlying logic of tests and question types, pedagogical modules), knowledge representation, and storing components (e.g., question pools, tests, exam answers, and results), connector components (related to integration and interoperability). Admittedly, the development of features for knowledge representation in the existing e-exam systems that we analyzed is currently much less advanced than such features suggested by Striewe.

2.3.3. Security features for e-exams

Mitigation of cheating has been especially challenging for BYOD e-exams since students can control the hardware in their devices to cheat during exams (Frankl, Schartner, & Zebedin, 2012). To make BYOD e-exams more secure, one must prevent students from accessing files on their computer and prevent them from accessing other programs than the e-exam system (e.g., email, chat, google). There are primarily two ways to achieve this, as presented in literature (Dawson, 2016): Either requiring booting from a memory stick or accessing the exam through a lock-down browser, both have their pros and cons. In these two ways, exams can be configured to run on lockdown browsers. When exams run on the host operating system, a lockdown browser can run as a plug-in or desktop application. On the other hand, exam systems that use live booting operating systems, run entirely off from USB stick where examiners can customize the operating system to run on lockdown browser, disallowing access to the computer's hard disk, and limiting internet access. In the existing literature, some authors have discussed security features for e-exams in general, while some explicitly addressed the security of BYOD e-exams.

Security features for e-exams in general. Huszti & Petho (2010) discussed security requirements for e-exam systems in general, focussing on authentication, anonymization of candidates, and confidentiality of questions. Dreier, Giustolisi, Kassem, Lafourcade, and Lenzini (2015) developed a formal framework for analysing the verifiability of exams and evaluated it with both traditional and e-exams. This framework verifies all exam phases, including registration verifiability, questions validity, exam-test integrity, marking correctness, marking integrity notification integrity. Kassem, Falcone, and Lafourcade (2015) implemented a solution to monitor the answers of e-exams. On the one hand, all these papers centered on providing security solutions for e-exams. On the other hand, Saini, Grispos, Liu, and Choo (2017) examined the ways to compromise e-exam applications, mainly extracting examination questions from question pools of e-exam applications. Further, their study involved proposing recommendations for enhancing credibility and integrity of exam questions.

Security features for BYOD e-exams: Several papers discussed the security features of Moodle LMS in the BYOD e-exams setting. Frankl et al. (2012) provided the so-called "Secure Exam Environment" implemented for Moodle LMS at AAUK, Austria, to run BYOD e-exams that mainly use wired LAN and boot from USB or DVD. Kaiiali et al. (2016) provided a "Secure Exam Management System" for the security of BYOD exams run on Moodle LMS through WIFI. Hillier & Fluck (2017) reported a technical solution to develop a BYOD laptop-based e-exam system, which runs exams in offline BYOD and bootable USBs, encouraging more authentic assessment practices.

Relation to our research: Unlike some of the above-mentioned research, this thesis did not implement any own security solutions for e-exams. However, we evaluated the security features of some of the international-level e-exam tools used for e-exams. In P2, we performed penetration testing on an open-source lockdown browser – Safe Exam Browser (SEB), which is widely used for e-exams internationally. The testing revealed some of the vulnerabilities that can be exploited to bypass SEB security properties. Further, interview responses in P7 added some empirical evidence of issues with SEB to P2.

Related work on interoperability requirements will not have a subsection here, as this will instead be discussed in Section 2.4. All in all, the publications reviewed above do present key requirements and features for e-exam systems in various ways. Still, none of them investigate the experiences and perceptions of vendors and managers. This thesis has provided empirical evidence on key requirements for e-exam systems, especially concerning security and interoperability, through a case study with vendors and managers in the Norwegian higher education sector in P7.

2.4. Digital ecosystems for e-exams

This section describes how different requirements mentioned in the previous section can be addressed through open digital ecosystems for e-exams. There are only a few publications available directly on digital ecosystems for e-exams. In this section, we will also discuss ecosystems in e-learning and e-assessment in general. We group our discussion of related work on open digital ecosystems for e-exams in the following categories.

- Research proposing overall architectures or principles for software ecosystems or digital ecosystems within e-learning and e-assessment, to be discussed in section 2.4.1
- Research presenting requirements engineering processes and methods for e-learning or e-exam ecosystems, section 2.4.2
- Research on obstacles towards achieving open digital ecosystems in e-learning, section 2.4.3.

For each of these categories, we will cover some general literature in e-learning, then focus on the research specific to e-exams and discuss how the contributions in our thesis relate to this body of research.

2.4.1. Architecture or principles for e-learning ecosystems

The ecosystem metaphor comes from biology and highlights the evolving nature of software, with many products partly competing, partly complementing each other (García-Holgado & García-Peñalvo, 2018). In software ecosystems, the overall needed functionality can be achieved by composing software services from several providers. Manikas and Hansen (2013) define software ecosystem as "the interaction of a set of actors on top of a common technological platform that results in a number of software solutions or services." (p.1297). A broader concept than software ecosystem is *digital ecosystem*, which along with the software products found in a software ecosystem, also includes hardware and digital content interchanged and shared between several providers (Kallinikos, Aaltonen, & Marton, 2013; Selander, Henfridsson, & Svahn, 2013). These ecosystems also include human as a component, and they are affected when the other components in the ecosystems evolve (García-Holgado & García-Peñalvo, 2016). Many software companies let others extend their core products by opening their platform and also generate profit from that. Such an approach has also been argued within e-learning (Uden et al., 2007) and e-assessment (Llorens, Molina, Compañ, & Satorre, 2014; Luo & Lin, 2013).

Principles for software/digital ecosystems within e-learning. CLO[7] media define the digital ecosystem for e-learning (also called e-learning ecosystems) as "*the learning community, together with the enterprise, united by a learning management system (LMS)*." Uden et al. (2007) outline the e-learning ecosystem as formed by three categories of components: content providers, consultants, and infrastructure. The meaning of these categories is that content providers offer content for learning solutions, consultants help organizations develop strategies, and infrastructure helps deliver and track e-learning. Uden et al. further discussed limitations of e-learning technology, and benefits of an integrated approach with e-learning ecosystems. Similarly, Pettersson (2009) presents principles for designing software ecosystems within e-learning, looking specifically into the standards and practices in e-learning systems. He pointed out concerns towards teachers' adoption of e-learning ecosystems and reusability of digital content in schools. All these definitions focused on the technological aspects of learning ecosystems, whereas Chang and Guetl (2007) defined learning ecosystem (LES) as consisting of the stakeholders incorporating the whole chain of the learning process and the learning utilities, the learning environment, within specific boundaries such as learning environmental borders, focusing more on the human aspect.

Architecture for software ecosystems and digital ecosystems within e-learning. Dong, Zheng, Yang, Li, and Qiao (2009) proposed an architectural model for an e-learning ecosystem based on cloud computing architecture that allows stability, efficient resource use, and sustainability for e-learning ecosystems. Chang and Uden (2008) discussed a framework to support the implementation of e-learning ecosystem governance, which consists of four characteristics, including structures, processes, communications and relational mechanisms, pedagogies and instructional

[7] https://www.chieflearningofficer.com/2004/08/30/e-learning-ecosystems-the-future-of-learning-technology/

designs. Guetl and Chang (2008) provided a review of ecosystem-based models for learning that are applicable for e-learning 2.0. Aparicio, Bacao, and Oliveira (2016) provided a literature review on e-learning concepts and proposed an e-learning systems theoretical framework consists of three main components: people, technology and services. García-Holgado and García-Peñalvo (2016) provided a formalized architectural pattern to improve the definition and implementation of e-learning ecosystem. They implemented it in the real context in Spanish Public Administration. This pattern would be helpful for learning, training, and knowledge management of public employees and enhancing the sharing of courses and resources between public schools (García-Holgado & García-Peñalvo, 2014). Chang and Guetl (2007) proposed an architectural model for learning ecosystems and applied it to small-and-medium-sized organizations (SMEs).

Specific to e-exams: As far as we know, there has not been any published research specifically on e-exam ecosystems. Thus, there has been no standard definition provided for e-exam ecosystems in the literature. Here we define a potential digital ecosystem for e-exams or e-assessment as *intersection of the platform providing the basic functionalities (e.g., question delivery to candidates, collection of answers, marking) with plugins taking care of a plethora of more specialized needs (e.g. authoring or answering of specific question types in various speciality domains, i/o for students with special needs, specific grading schemes, advanced support for grading, etc.).* Thus, one could consider e-exams not as a separate digital ecosystem in its own right but rather a key part of a bigger ecosystem for e-learning, so e-learning ecosystems will naturally contain components for formative and summative assessment. Ullrich, Forell, Houy, Pfeiffer, Schüler, Stottrop et al. (2021) presents platform architecture for specifically supporting the e-assessment of various diagram modelling tasks that are relevant in IT education.

Relation to our research: This thesis did not aim for proposing new principles and architectural models for e-learning/e-assessment ecosystems. However, our empirical findings from P7 did align to the definition of e-exam ecosystem through the e-exam systems platforms that support plug-ins for providing basic functionalities for, e.g., authoring and designing specific question type. Also, P7 (Cf. Figure 2 in this thesis) provided the architectural model of the e-assessment ecosystem particular to the Norwegian context. Our architectural model resembles the model suggested by Chang and Guetl (2007): learning utilities (Blackboard and Canvas LMSs, Inspera assessment and WISEflow e-assessment systems), ecosystem conditions related to learning environmental boundaries (using dedicated e-assessment systems for exams and LMSs for course management), however, our model did not provide components specific to learning stakeholders (biotic units) as suggested by (Chang & Guetl, 2007). Moreover, our architectural model was still in the early design stage; thus, it did not offer layered architecture, as provided by García-Holgado and García-Peñalvo (2016).

2.4.2. Requirements engineering process in e-learning ecosystems

Requirements engineering (RE) is a crucial process of software development (Dick, Hull, & Jackson, 2017; Roman, 1985). However, there is not much research on RE specific to e-learning and e-assessment ecosystems. First of all, there has not been much published research on how the RE process will be carried out in software ecosystems or digital ecosystems (Immonen, Ovaska, Kalaoja, & Pakkala, 2016). Most of the research in software ecosystems was focused on technical and architectural aspects rather than the requirements engineering perspective. Some papers discussed individual RE activities - requirements elicitation (Fricker, 2010; Jansen, Brinkkemper, & Finkelstein, 2009), requirements negotiation (Knauss, Damian, Knauss, & Borici, 2014; Valença, Carina, Virgínia, Slinger, & Sjaak, 2014), requirements specification (Boucharas, Jansen, & Brinkkemper, 2009; Van den Berk, Jansen, & Luinenburg, 2010a), however not explicitly discussed the whole RE process/method. On the other hand, Immonen et al. (2016) proposed an RE method specifically for digital service ecosystems. They pointed out that continuous communication between ecosystem members is one of the key issues in achieving the goals of ecosystems.

Specific to e-exams: We have been only able to find one paper particularly relevant to our own work about RE for e-exam systems. Foss-Pedersen and Begnum (2017) performed a case study with 27 institutions in Norway especially looking at universal access requirements for the Inspera Assessment and WISEflow e-exam systems. Their findings indicated that the RE process was somewhat ad-hoc. There was unclear division of responsibility between vendors and buyers, and nobody had stepped up to take a clear responsibility for accessibility requirements.

Relation to our research: P7 provided findings on the RE process being used for e-exams systems acquisitions in Norway. For this work, we interviewed vendors and managers of two mass-market products (i.e., Inspera Assessment and WISEflow), some of the same stakeholder groups that were interviewed by Foss-Pedersen and Begnum (2017). Our results indicate that the RE process was still somewhat ad hoc, in line with the findings of Foss-Pedersen and Begnum (2017), though we found that the RE process had been improved during the 3,5 years between their study and ours, specifically with a clearer division of responsibilities between stakeholders. Thus, the findings of P7 indicates an RE process becoming more similar to what was recommended by Immonen et al. (2016).

Our paper P3 provided a systematic mapping review of requirements engineering in software ecosystems. Although our findings from P3 did not show a novel approach for RE in software ecosystems as presented by (Immonen et al., 2016), P3 still presents some of the issues and potential research gaps in RE for software ecosystems.

2.4.3. Obstacles towards achieving digital ecosystems in e-learning

Nowadays, universities use multiple systems for supporting various learning and assessment tasks. In general, these systems require interoperability with other systems, e.g., e-exam systems, student information systems, and single sign-on systems for authentication, to support appropriate learning and assessment tasks. Major obstacles to interoperability could be that different systems use different interfaces and data formats. Thus, these systems use several standards and frameworks to support communication. Dagger, Connor, Lawless, Walsh, and Wade (2007) pointed out that lacking support of semantic interoperability in standards and framework is the main issue for information exchange between different service-oriented e-learning platforms. (Ouf, Abd Ellatif, Salama, & Helmy, 2017) which appears to agree that e-learning ecosystems is a good idea, but at the same time criticizes previous research on e-learning ecosystems to focus too much on one-size-fits-all mass education, not paying enough attention to the need to personalize the tool set to the needs of individual learners, which plays important role to build effective e-learning ecosystem. Jakimoski (2016) presents the analysis of several interoperability frameworks and further identified core interoperability challenges for the efficient personalization of learning environments. Alharthi, Spichkova, and Hamilton (2019) discusses requirements for e-learning systems in a sustainability perspective. Although, they did not make much direct mention of ecosystems, still address issues that would be relevant for achieving an effective ecosystem, such as openness, interoperability and reusability etc.

Specific to e-exams: Based on research literature on challenges for interoperability in e-learning, we have identified three main challenges:

- Lacking openness
- Lacking interoperability between components.
- Lacking priority for interoperability from vendors and customers

Lacking openness: Lacking openness is often addressed in the literature about concerns related to availability of APIs from vendors, flexibility of standards. Our system managers felt that APIs from e-exam systems vendors were strictly controlled and even the access was limited to only administration workflows. Similarly, Striewe (2019) argued that APIs of software would usually be strictly controlled by vendors to ensure the security of platforms. Flexibility of standards is often discussed as the main issue for interoperability (Kelly, Wilson, & Metcalfe, 2007). For instance, if a standard is too flexible, different tool developers may implement the standard in different ways, so that gradually different dialects of the standard emerge, and if these dialects are not entirely compatible, that will hurt interoperability. On the other hand, if a standard is very rigid and lacks flexibility, maybe some tool developers will feel that it does not fit their purpose, and not use it at all - which also hurts interoperability. The QTI standard allows tests and questions exported from one university's e-exam system to be imported to another university's e-exam system, however, QTI standard has been considered problematic (Chituc &

Rittberger, 2019; Wills et al., 2009). Piotrowski (2011) pointed out that the QTI specification is too complex, ambiguous, and challenging to implement.

Lacking interoperability between components. Our findings indicate three challenges in relevance to lacking interoperability between components. First, interoperability of the e-exam system with the other software products that the university already has, is considered as main challenge, as also similarly addressed by Jakimoski (2016). The second challenge was concerning interoperability of e-exam systems with other e-exam systems, for instance, universities X and Y exchanging or jointly developing question banks, whether they use e-exam systems from the same provider or different providers (Chituc, Herrmann, Schiffner, & Rittberger, 2019; Chituc, 2020). Research shows that there has still not been much development from vendors about sharing question banks between universities (Chituc et al., 2019). However, the universities could only be able to save effort and increase the quality of tests if they could be shared between universities and across countries. The third challenge was related to interoperability with third-party tools for e-assessments. For instance, a programming exam might need compiler support for students to run their code, and the compiler is not integrated with the e-assessment system. Kurniawan, Lee, and Poskitt (2020), provided their experience on implementing programming exams with access to an Integrated Development Environment (IDE) during exams. Their findings show that IDE slowed down performance sometimes.

Lacking priority for interoperability from vendors and customers. Several e-learning tools support the IMS Learning Tools Interoperability (LTI) standard that allows external tools to be launched within an application (Queirós et al., 2016; Severance et al., 2010). This feature enables various learning tools to interoperate, similar to single-sign-on, e.g., from LMS to e-exam application. García Peñalvo, Conde García, Alier Forment, and Casany Guerrero (2011) addressed this feature through a service-based framework connecting Moodle LMS and Basic LTI. Their framework could make it easier both for commercial vendors and free software developers to make plugins supporting the authoring, solving, and grading of various question types in connection with LMS or e-exam platforms. Sclater (2007) argued that despite large investments by vendors and educational bodies, the specifications had not reached a critical mass of adoption due to the insufficient demand by users, particularly in higher education institutions, where the assessment process is strictly controlled by single institutions. On the other hand, he recommends universities for interoperability testing of systems during the procurements, so it must be included at acquisition to ensure that the vendors really deliver on this.

Overall, these papers have given good input for research related to the larger theme of e-learning ecosystems, indicating that ecosystems would be the right way to go. But due to various obstacles (e.g., higher priority for functional extensions, organizational barriers, etc.), it would take time to get there.

Relation to our research: This thesis addressed interoperability issues within e-exam systems in P5 and P7. One of the main goals for digital ecosystems for e-exams (as mentioned in the definition of e-exam ecosystems in 2.4.1) is to enable plug-ins for learning systems. We addressed interoperability issues for programming exams

focusing on Parson problems and code writing, mainly focusing on interoperability challenges with the Inspera Assessment e-exam system. P5 pointed out several challenges with interoperability between LMS and e-exam systems with our practical experience with tools, which is in line with findings of (Chituc & Rittberger, 2019). But, our research was limited to findings related to only one e-exam system - Inspera Assessment, and also we focused specifically on the (partly lacking) support for questions in programming, such as Parsons problems and code writing, rather than providing a broader view of interoperability challenges as done by Chituc and Rittberger (2019).

P7 addressed interoperability issues between components within a broader scope than P5 with the expert interviews. Our results indicate lacking priority for interoperability from customers (i.e. Universities) similar to the findings by (Sclater, 2007). Our findings in P7 also show that system managers in our case study considered interoperability between different e-exam systems is a challenge for exchanging questions between different systems. Also, similar to (Chituc et al., 2019), our findings show that participants considered integration of question banks as useful, however, not a highest priority as improved functional features, as well as better usability and security required more attention. Our findings also pointed out that e-assessment systems lack the support for some of the third-party integrations (e.g., Matlab) required for BYOD e-exams. The main issue with third-party integrations, we believe, in general, would be that BYOD devices would not deal with licensing of some third-party tools, e.g., Matlab, used for e-exams in client-side, thus requiring the installation of the needed tools in BYOD devices before exams. However, empirical findings from P7 point towards the same conclusions as presented in P5 – that the support for making plug-ins is currently limited.

3. Research Approach

Information systems (IS) is a multi-disciplinary research field, including both technical and non-technical issues. Several research methods are available for information systems research. Palvia, Leary, Mao, Midha, Pinjani, and Salam (2004) present fourteen research methods that are commonly used in information systems research. This thesis has used different research methods based on the needs of the various research questions. The rest of the chapter is structured as follows: Section 3.1 discusses the context of the research. Section 3.2 provides methodological research conducted in this thesis. Section 3.3 presents the overall research design, arguing why various research methods were selected. Section 3.4 then goes into more detail about how each research method was used in this particular research project. Section 3.5 discusses the data collection and analysis methods used. Finally, section 3.6 provides research ethics considered in this thesis.

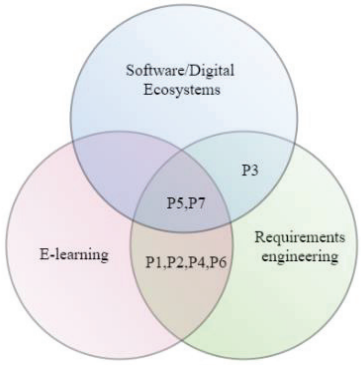

Figure 3. Venn diagram with fields of research

3.1. Research Context

The research described in this PhD thesis is framed between three fields (Figure 3): Requirements Engineering, Software/digital ecosystems and E-learning. When it comes to topical context, Papers P1, P2, P4, and P6 were concerning requirements for e-exams, especially regarding cheating and security. Specifically, P1 and P6 were about the comparison of paper and e-exams, P2 was about penetration testing performed on an open-source lockdown browser (SEB) software, P4 was about the comparison of on-campus exams and remote exams. P3 was about requirements engineering in software ecosystems. Papers P5 and P7 were about requirements for e-exams in digital ecosystems. Hence, P1, P2, P4, and P6 are located at the intersection of E-learning and Requirements engineering, P3 is located at the intersection of software/digital ecosystems and requirements engineering, and P5 and P7 are lies in the area where three circles intersect. Admittedly, that the thesis would have been more coherent if every study had stayed in the centre where all themes intersect. Here, P2 could be placed at the intersection of all themes since SEB

is an open-source tool that is used around the world, and it could be integrated with other tools and be seen as a potential component in digital ecosystems. For P3, it turned out very little work had been published just for requirements engineering of ecosystems in e-learning and e-exams, as most work here had been more on the idea level. Hence, we looked at RE for software ecosystems in general in a broader scope for the mapping study. Also, it turned out that current e-exam technology, at least as used in Norway, had not progressed very far in the direction of an open ecosystem, and functional features and security requirements especially to mitigate cheating were currently of far more interest to stakeholders.

When it comes to geographical setting, here, e-exams are organized in Norway using Feide SSO, FS student information system and dedicated tools for summative assessment (Inspera Assessment, WISEflow), together with lockdown browsers (SEB, FLOWlock) for some exams. The way LMSs like Canvas, Blackboard are used for coursework, i.e. ungraded/formative assessments. Some of the research has an even more local context (NTNU). Even if we are looking specifically at the Norwegian or local context in some of our research, these tools are also used in a lot of universities around the world. Blackboard and Canvas are well-known learning management systems globally. Safe exam browser is also a well-known lockdown browser internationally. Even the dedicated e-exam systems - Inspera Assessment[8] and WISEflow [9]are used by universities in many countries. However, Inspera Assessment and WISEflow e-exam systems lack the necessary features to fill other tasks of an LMS, outside assessment, so the universities abroad may have split usage not necessarily identical to what NTNU has, but at least somewhat similar, i.e., Inspera Assessment and WISEflow taking care of some (at least exams) or all assessment tasks in the course, while the LMS does the rest.

Table 2. Geographic setting of papers

Aspect / Papers	Human informants	Tools concerned by the study	Foundation literature
P1	N/A	N/A	international
P2	N/A	international	international
P3	N/A	N/A	international
P4	N/A	N/A	international
P5	N/A	national, international	international

[8] Inspera's webpage, https://www.inspera.com/about claiming to have end users in 160 countries (which is more than half countries around world), and showing logos of well-known universities in many countries as customers.
[9] WISEflow is currently serving more than 20 educational institutions in Denmark, Norway, Sweden, Iceland and Greenland, Netherlands, UK and Korea.
Cf. https://www.uninett.no/sites/default/files/webfm/Seminar_Uniwise_wiseflow.pdf

P6	local	national, international	international
P7	national	national, international	international

Table 2 presents the papers' basis concerning national, and local settings. The papers P1, P2, P3, and P4, were described in the international setting (P1 – since the comparison of paper and e-exams here did not presuppose any specific tool or university, P2 – since the penetration testing performed in paper 2 is about SEB, which is an open-source software product used in many places around the world, and the paper also focused on the modelling approach rather than just the penetration testing of SEB system, P3 – since the research performed in the mapping study is not depending especially on the situation in Norway, and P4 – since the risk analysis here is also independent of geographical setting). P7 has a national context since various Norwegian universities were involved. Although WISEflow is a Danish company, they were interviewed in the capacity of being the vendor of one of two e-exam systems (Inspera Assessment, WISEflow) used in Norway. Also specific to Norway (Cf. Figure 2), is the FS and Feide tools, and the role of UNIT, and the stakeholders who were informants in the case study. P5 also has a national context, but Inspera Assessment does also sell to several other countries than Norway, so to some extent, it is wider than just national. P6 has a local NTNU setting since all the participants were from the NTNU.

3.2. Research Methodology

This section provides the basis of methodological research conducted in this thesis. There have been a variety of research methodologies used in IS field (Oates, 2005). Figure 4 gives an overview of the research process used to address research questions (indicated in grey boxes) based on Oates research process recommended for Information Systems and Computing research.

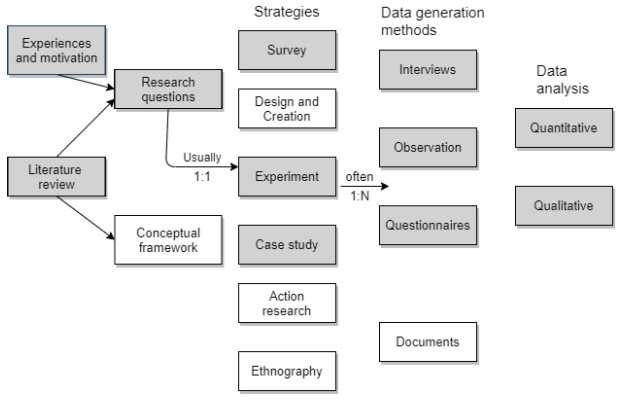

Figure 4. Model of research process adapted from (Oates, 2005)

Oates (2005) in her conceptual framework defines research methodology as “the combination of research strategies and data generation methods that you adopt” (p 31). Many researchers do not clearly distinguish between terminology used in the research process. Twining (2010) provided comparisons of different terminology used in the literature research. For consistency, in this thesis, research method is used for denoting research methodology. The type of research methodologies has also been indicated by researchers in the literature differently. Palvia et al. (2004) considered literature review and interviews as research methodologies in their research. Whereas Oates (2005) denoted literature review as a preliminary step and continuous process to keep up with the published literature, and interviews as data generation method.

The research process (Cf. figure 4) of this thesis started with defining research questions emerging from experience and motivation and literature review (P1, P3, P4), as suggested by (Oates, 2005). This thesis used three research methodologies covered in the existing literature by Oates and Palvia et al.: literature review, case study, experiment. In addition, due to the nature of the research questions, threat modelling and risk analysis and mixed-method research methodologies were used. Penetration testing was used in experiments to find vulnerabilities in e-exam software. The reason for using threat modelling and risk analysis, penetration testing, and mixed-method research is provided in detail in the section about research design (Section 3.3). Questionnaires, observations, and interviews were used for data generation.

Both quantitative and qualitative data analysis have been used for research in this thesis. Quantitative data is numerical (e.g., measurements and statistics) and qualitative data is non-numeric (e.g., interview transcripts, documents). Relative preference for quantitative or qualitative methodology depends on philosophical issues related to the question of ontology (assumptions about nature of reality) and epistemology (assumptions about the nature of knowledge and how it can be obtained) (Oates, 2005). Oates (2005) also defines research paradigms (positivism, interpretivism, critical research) while explaining how different research mythologies are linked to different paradigms. Paradigms specifies the view of researcher. Oates (2005) suggests that the positivism paradigm underlies the scientific method that mainly uses experiments; however, in IS research, positivists also use surveys since experiments are often not feasible. She provided various characteristics of positivism, of which quantitative data analysis, universal laws (i.e., generalizations) are relevant for this thesis. As per interpretivism, she suggests that it is concerned with the interpretation of interpretivist through exploring and explaining particular social setting through observation (e.g., interviews). In this thesis positivism (for experiments and surveys) and interpretivism (for case studies/interviews) paradigms have been used. However, as per differences between paradigms, regardless of quantitative and qualitative, they can be used by positivist, interpretivist, but the paradigms are mainly distinguished by their ontology and epistemology (Oates, 2005). In this thesis, we use quantitative methodology with realist/objectivist ontology and empiricist epistemology in surveys (P6) and experiments/penetration testing (P2), whereas qualitative methodology with interpretivist epistemology and constructionist ontology in personal experiences

with e-assessment and supporting tools (P5), and in interviews (P6, P7) as suggested by Tuli (2010).

3.3. Research Design

This section mainly aims to describe the overall research design and choice of the research methods, i.e., why a particular method or combination of methods were chosen. We will not provide more details on our exact usage of each method here, this will be done in sec. 3.4. The method used for each research question is discussed in subsections 3.3.1 – 3.3.3. Table 3 provides the overall research design of this thesis.

Table 3. Mapping between studies, papers research methods, research questions, and contributions.

Study	Aim	Paper	Selected research method	Data collection	Data analysis	Rationale for using the method	RQ	C
S1	Compare paper exams and e-exams	P1	Interpretive literature review	Used ADTrees	Categorization of threats and countermeasures	- To conduct review in a small set of high-quality and relevant published literature - less time-consuming than systematic literature review	SQ1	C1
S2	Extend previously proposed HARM method to achieve a systematic approach to penetration test development for cheating scenarios in exams	P2	Experiment	Penetration testing on SEB lockdown browser	Categorization of testing scenarios into success or fail	To test the viability of threats identified in P1 in practice	SQ1	C1
S3	Systematic mapping study on software ecosystems related RE activities and quality attributes.	P3	Systematic mapping review	44 relevant papers	Categorization of findings related to RE activities and quality attributes	Compared to SLR, SMR will not perform more meta-analysis of primary studies and less time consuming than SLR	SQ2	C2

S4	Identify mitigation of cheating in online exams	P4	Interpretive literature review	Used ADTrees	Categorization of threats and countermeasures	To conduct a review in a small set of high-quality and relevant published literature	SQ1	C1
S5	Identify how digital ecosystems within the e-exam system improve e-assessments	P5	Case study on e-assessment tools used at NTNU	- tests on Blackboard, Inspera Assessment - Review documentation about WISEflow, Canvas, Moodle	Manual analysis	Case study was the best choice as research questions for this paper are exploratory in nature	SQ3	C5
S6	Investigate perceptions of engineering student and teachers in Norway on comparison of cheating in e-exams and paper exams concerning six-different cheating practices and seven different countermeasures	P6	Mixed-method research	- 212 students and 162 teachers survey questionnaire responses - Semi-structured interviews with 6 students and 5 teachers	- Statistical analysis using One-sample t-tests and Independent t-tests - Qualitative data analysis in Atlas.ti 8 using the constant comparative method from grounded theory	Mixed method research collects and analyses both qualitative and quantitative data rigorously in response to research questions	SQ1	C1
S7	- Investigate key features for e-exam software according to vendors, process managers, system managers in the Norwegian HE sector. - Investigate how key features are identified and agreed upon. - Investigate vendors and managers views on requirements from a digital ecosystem perspective	P7	Case study on e-exam tool support for higher education in Norway.	- 12 Semi-structured interviews from 7 different organizations - 12 participants (1 process manager, 3 from vendors organization, 8 system managers from four universities)	Qualitative data analysis in NVivo 12 using the constant comparative method from grounded theory	Case study was the best choice as research questions for this paper are exploratory in nature	SQ1 SQ2 SQ3	C1 C3 C4 C5

3.3.1. Research methods for SQ1

SQ1 considers cheating-related risks in e-exams and how these compare to paper exams. This question is addressed using four approaches, literature review, threat modelling and risk analysis, experiment, and mixed-method research. The reasons for choosing these methods are provided below.

Literature review. Literature review of publications about cheating in paper exams and/or e-exams can be used to determine threats that have already been observed in practice and have been discussed in scientific publications (Cf. sec.2.2). However, a weakness especially related to the fact that e-exam technology is developing quite fast, and technology for cheating is also developing fast, is that there may be threats that have become more important recently, or that will be important in the future, which have not been discussed in scientific papers. Moreover, there may also be some threats that have been discovered but not published about, e.g., tool developers may not openly publish findings of security vulnerabilities in their products and may also be somewhat vague about the exact countermeasures they are using against cheating, as these may be business secrets. Hence, literature review would not give a complete picture of threats and countermeasures, instead a picture that might be somewhat dated and skewed by what has been treated in publications. On the positive side, literature review tends to give a good overview of a problem domain since different publications will have treated the problem from different angles (Jesson, Matheson, & Lacey, 2011).

Threat modelling and risk analysis (or threat analysis). Threat analysis is a method to imagine various ways an attacker could achieve illegitimate goals (such as successful cheating) in a system (Shostack, 2014). It depends on the analyst's creativity and experience and is a somewhat subjective process but doing the process in a structured manner. The aim is to end up with a relatively complete and structured view of the most important threats and which countermeasures may be applied against these threats. Compared to literature review, threat analysis has the advantage of being able to identify novel threats and threats that will be of increasing interest in the future, even if these have not been published about or materialized in a lot of actual attacks yet. Hence, literature review and threat analysis complement each other – literature review reveals what threats have already been observed in practice and written about in publications, while threat analysis can give a more complete picture, including other threats that are possible. What threat analysis by its own does not tell, however, is to what extent the imagined attacks will be successful in various actual system implementations.

Experiment. Penetration testing was used for experiments. Penetration testing is a method to test the computer system to find the security vulnerabilities in the system before an attacker could exploit them to achieve illegitimate goals (Bishop, 2007). Actual penetration testing of e-exam software may have some advantages where literature review has weaknesses – namely in exploring vulnerabilities of actual software products even if they have not been addressed by anyone in scientific papers. On the other hand, contrary to literature review, testing of actual products does not give the same overview, as one typically has to try out products one by one,

and for each product, various test cases one by one, which is time-consuming. Moreover, software products tend to be released in new versions every now and then, where some previous vulnerabilities will have been fixed (while others may have accidentally been introduced). Hence, penetration testing gives a deeper technical insight and possibility to look at vulnerabilities that exist in the software but have not been published about yet. At the same time it provides a snapshot of the situation just at the time of testing, without the same overview as could be obtained by literature review. Compared to threat analysis, penetration testing shows which attacks are practically possible in which e-exam systems, whereas threat analysis, instead deals with imagined scenarios and would not have to be restricted to one specific tool setup.

Mixed-method research. Mixed-method research could enable balance with joint analysis and triangulation of quantitative and qualitative data (Creswell & Clark, 2017). The combination of quantitative and qualitative data often provides better understanding of the phenomenon. This research investigates stakeholders' perceptions about the importance of various threats and the viability of various countermeasures through quantitative surveys and qualitative interviews. This method has advantages and disadvantages, e.g. relating the threats and countermeasures that have come up through threat analysis. The investigations with students and teachers tell more about what cheating may be taking place in practice, in the local context, rather than what is technically possible in a tool or what is theoretically possible in a generic e-exam system, or what has been taking place in other universities before. Again, a disadvantage with this method would be that participant with little or no personal experience with cheating would have trouble answering questions, especially in surveys where there is a lesser chance to get clarification on questions unless the survey is responded to at direct administration.

3.3.2. Research methods for SQ2

The research question SQ2 about ecosystems mandates the combination of literature review (P3) to get an overview of published research internationally, plus a case study to explore how relevant ecosystem thinking is in the context of e-exam systems development in Norwegian Higher Education (P7). While the systematic mapping review of publications about requirements engineering can be used to find out what has been published in literature about requirements engineering in software ecosystems in general, the case study with stakeholders (tool vendors, system managers at universities) could provide in-depth understanding of the requirements or features and the process surrounding the acquisition of e-exam systems. The disadvantage here is that vendors may be reluctant to reveal everything about their products, in order to protect business secrets. Hence, the choice for the SQ2 was a research method that is flexible in design. In the flexible design process (e.g., case studies), key parameters that effect the findings, may be changed during the course of study, while in a fixed design process (e.g., surveys), all parameters are defined at the launch of the study (Robson, 2002). For instance, if the interviews turned out not to reveal reasonable findings to provide clear evidence, additional data could be

added from other forms of data collection, e.g., surveys, archival analysis and observations (Runeson & Höst, 2009).

3.3.3. Research methods for SQ3

Robson (2002) classifies research methodologies based on the purpose for research – Exploratory (finding out what is happening, seeking new insights and generating ideas and hypotheses for new research, Descriptive (portraying a situation or phenomenon), Explanatory (seeking an explanation of a situation or a problem, mostly but not necessary in the form of a causal relationship) and Improving (trying to improve a certain aspect of the studied phenomenon). The research question SQ3 is of an exploratory nature that made case study a natural choice. A quantitative survey might have been used for this study if the purpose was to compare stakeholder views on features of e-exam tools in terms of ranked preferences. However, the research questions are not about the relative importance of the features in numerical terms, but rather what the key features are and why they are important. As argued by Yin (2002) such research questions are best suitable for case studies. Thus, the exploratory nature of the research questions points towards inductive reasoning (Braun & Clarke, 2006; Twining, Heller, Nussbaum, & Tsai, 2017), using the exploratory case study approach (Yin, 2017) to develop theory from the case.

We have conducted two case studies for SQ3. However, the data collection method was different in these two studies. One was targeted on the case 'e-exam tool support at NTNU' in a local setting, whereas the other one was about 'e-exam tool support in Norwegian higher education' more in a national setting. In the former, the case was analysed more based on our own practical experiences with the tools, while the latter was conducted through interviewing experts in Norwegian higher education. The advantage of having experts as informants is that they will have a broader overview of the e-exam system functionalities than researchers, as they receive more feedback from users in addition to the practical experience they have.

When it comes to the type of case study, the first case study at NTNU was quite obvious to be a single case. The second case study, although it spans several universities and two different system providers, was also considered to be most appropriately defined as a single case study, rather than multiple case. A key rationale for this is that all these organizations are stakeholders in a larger procurement and development process coordinated by Unit. The two system providers also have customers in other countries, but since only Norwegian higher education institutions were included, the case is "e-exam tool support for higher education in Norway".

3.4. Research Methods

The previous section explained the overall research design and why a certain combination of research methods was chosen. The current section will go into more detail about why a specific method was used and how the particular research methods were conducted.

3.4.1. Literature review

Most projects in information systems and software engineering involve an element of literature review. There are mainly two different types of reviews: interpretive literature review, systematic literature review.

Many researchers distinguish between two main types of literature review: systematic and interpretive (Dyba, Dingsoyr, & Hanssen, 2007; Jesson et al., 2011). However, as argued by Schultze (2015), while the term "systematic" says something about the process, "traditional" says nothing particularly about the approach, so a better distinction would be "systematic" vs "interpretive". Systematic then focuses on rigorous, documented procedures and automated searches that strive for completeness in coverage of sources, such that the literature review in principle should be possible to repeat by other researchers. Interpretive literature review focuses more on the researcher's personal interpretation of the literature, and thus does not make the same claims for repeatability or complete coverage. Also argued by Schultze, it is deceptive to look at systematic vs interpretive as a clean dichotomy. It is rather a continuous spectrum where papers could also be somewhere along the middle. interpretive (or traditional) literature reviews may typically also have used keyword searches in publication databases, and criteria for inclusion or exclusion of sources, and even the most rigorous systematic literature review would not be void of personal interpretation by the involved researchers, for instance in conducting inclusion/exclusion decisions.

Below, we provide description of these types, following how we have used these approaches in our research.

Interpretive literature review. The interpretive literature review does not strive for completeness in coverage of sources, rather cover the most relevant and important sources on a topic. Hence, it can address a smaller set of literature than a systematic review, and thus be less time-consuming – or if spending the same time as a systematic review, allowing to go deeper in the analysis and interpretation of just the most important sources. The disadvantage of the approach is that it is more ad hoc than a systematic review, relying more on the researcher's interpretation of what is relevant and important, and the results may thus less reliable or repeatable - with a higher risk of missing some relevant publications (Dyba et al., 2007; Jesson et al., 2011).

Systematic literature review. Two types of systematic review approaches are presented in literature: systematic literature review (SLR) and systematic mapping review (SMR). SLR is more thorough and time-consuming than SMR because the research question in SLR is used to identify the primary studies and, consequently the data extraction process is applied to each primary study following the aggregation of the extracted data (Kitchenham, Pearl Brereton, Budgen, Turner, Bailey, & Linkman, 2009). Whereas, SMR classifies the relevant literature in that specific domain and aggregates studies concerning categories, e.g., author's names, authors affiliations, publication source, publication type, publication date, venues, topics, etc. (Kitchenham, Budgen, & Pearl Brereton, 2011).

Specific method choice in our research. We chose interpretive literature review for P1 and P4 because we needed to also do threat analysis and penetration testing, hence, a full systematic literature review would be too demanding. Also, due to the quick development of technology, a systematic literature review might entail spending a lot of time analysing papers of marginal importance, e.g. discussing cheating threats that are now dated, while our goal was rather to get a reasonably good overview of the most important groups of threats. Hence an interpretive review was found sufficient, and just discussing background and related work about cheating in the papers reporting on threat analysis, penetration testing and empirical investigations with stakeholders about cheating.

For P3, we performed a systematic mapping review. The search process was an important part of SMR. It was conducted following five steps:

1. Seeding key research involving categorizing requirements modelling and security issues in software ecosystems from already studied research papers.
2. Defining search protocol to identify published research on RE activities and non-functional requirements in software ecosystems.
3. Conducting a systematic search on different databases.
4. Additional manual search
5. Quality assessment of the papers involving removal of low-quality papers using inclusion and exclusion criteria.

These steps are rather briefly presented here, for more details on search process see the full text of P3 in Part II of this thesis. One common threat to systematic literature reviews is not to discover all relevant studies. To reduce this threat, we defined the search strings to retrieve as many documents as possible related to the research topic. To reduce the bias in selecting papers, we defined a review protocol with clear inclusion and exclusion criteria for each selection phase. We also adopted a well-prepared quality assessment checklist to judge the paper's quality (Nguyen-Duc, Cruzes, & Conradi, 2015). The selection and extraction were made by at least two authors together, which helped to reduce researcher bias.

3.4.2. Threat modelling and risk analysis

Threat modelling is recognized as one of the most important activities in software security. It is often used to show the structured representation of threats that affects the security of the application. As mentioned in (OWASP, 2021), threat modelling is a process for capturing, organizing, and analysing all the information, enabling decision-making about application security risk, further, produce a list of improvements to the requirements, design, or implementation. Early threat elicitation as soon as the initial architectural model of a system is available would help to elicit the security requirements of a system, and, consequently, help to reconsider and refine the architectural model (Myagmar, Lee, & Yurcik, 2005; Scandariato, Wuyts, & Joosen, 2015). It is recognized by the most well-known secure software processes, Touchpoints (McGraw, 2004), OWASP's CLASP

(Chandra, Wohleber, Feragamo, & Williams, 2007), and Microsoft's SDL (Howard & Lipner, 2006).

There is no standard modelling approach used for threat modelling. Shostack (2014) recommends three strategies for threat modelling: focusing on assets, focusing on attackers, and focusing on software. Some of the most often used threat modelling techniques in research include Microsoft STRIDE (Scandariato et al., 2015), attack trees (AT) (Schneier, 1999), attack defense trees (ADTree) (Kordy, Mauw, Radomirović, & Schweitzer, 2014), misuse cases (Sindre & Opdahl, 2005). ADTree is the extension of traditional attack trees (Saini, Duan, & Paruchuri, 2008; Schneier, 1999), where there will be a representation of defense nodes in addition to attack nodes. Karpati et al. (2013) proposed Hacker Attack Representation Method (HARM) to represent hacker attacks in various ways, combining attack sequence diagrams, misuse sequence diagrams (MUSD), misuse case diagrams (MUD) (Katta, Karpati, Opdahl, Raspotnig, & Sindre, 2010), misuse case maps (MUCM) (Karpati, Opdahl, & Sindre, 2015) and attack trees. The MUSD and MUD methods use similar notations from Unified Modelling Language (UML). The methods such as ADTrees, AT, MUSD, MUD, and MUSD focus on the attacker perspective with variations in the sequence of attack representation. Whereas, MUCM focuses on attackers and software, showing the relationship between the attack sequence and the architecture, and each step in its architectural context. Compared to MUSD and MUCM, which show details of one particular type of attack, MUD shows a broader overview.

Specific method choice in our research: For the threat analysis in P1 and P4 we chose attack defense trees (ADTrees) (Kordy, Mauw, et al., 2014) for threat modelling and analysing security threats and countermeasures for both paper exams and e-exams. Because we wanted to look into both cheating threats and countermeasures, ADTrees had advantages compared to the other modelling approaches mentioned above, in being able to capture both threats and countermeasures in the same diagrams in a nice visual way. This graphical security modelling method was a new and rapidly growing modelling method at the time of our study (Kordy, Piètre-Cambacédès, & Schweitzer, 2014).

3.4.3. Experiment

Penetration testing was used for experiments. Penetration testing is the primary method used to ensure that vulnerabilities or weaknesses in networked environments, web applications, and physical premises are identified and can be tackled before they are abused in an actual attack (Tang, 2014). Furthermore, such testing is useful to identify the risks associated with the attacks. It requires detailed analysis of the threats and potential attackers in order to be most valuable (Bishop, 2007). Arkin et al.(2005) suggest that penetration testing can be effective if combined with the security-related findings from the earlier software life cycle stages, e.g., findings from requirements analysis and risk analysis, but less effective performed completely ad-hoc. The penetration testing stage in software development life cycle from their research is shown in Figure 5.

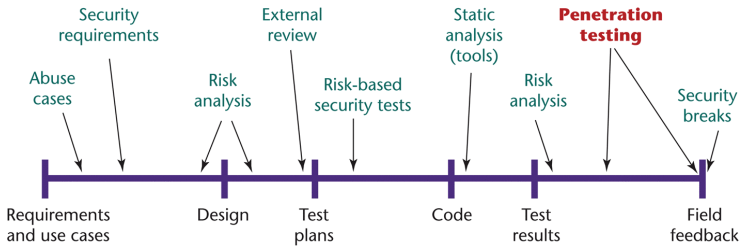

Figure 5. Penetration testing stage in software development life cycle (Arkin et al., 2005)

Specific method choice in our research: In P2, we chose to conduct penetration testing on the SEB (SEB 2.0) open-source lockdown browser software. We have performed the tests in a lab setting, not during a real exam. The testing was done by two persons – the phd student and a master student who was supervised by the phd student during her master thesis research. We used a laptop with SEB installed (used by the master student role-playing as a cheating test-taker) and then another laptop used by an illegitimate assistant (role-played by the phd student), and then tried out various cheating scenarios. It might have been even more representative to do such penetration testing in a real exam setting, not just to see if it was technically possible but also check if it could go unnoticed by the invigilators. But this was not chosen as real testing would be more expensive, potentially disruptive for real candidates taking the exam, potentially distracting for invigilators – who, if they were to discover the test-cheating, might end up not discovering some other real cheating that happened at the same time, potential ethical problems with doing such a test during a real exam. As an alternative, one might construct a realistic experiment as a mock-up exam with hired "invigilators" and "test-takers" – some of whom were role-playing honest students, and some role-playing cheaters with various approaches. But this would also be more expensive than a lab test with two researchers, and if invigilators knew it was a mock-up to test cheating approaches, they might be more alert than the average invigilators in real exams, thus not giving representative data. Hence, the purpose of finding out whether attacks are likely to succeed in practice may hold bigger challenges than in the relaxed lab setting. Since real-world tests are more time-consuming and expensive than lab tests, it is good to perform lab tests first. If it turns out that some type of attack was not even possible in the lab, it may be a waste of time to develop a real-life test for it, so resources should rather be spent on other attacks that were more likely to succeed. Again, here for penetration testing, the HARM threat analysis (Karpati et al., 2013) method was used to establish a clear connection between the threat analysis and penetration test cases. The main reason for choosing HARM was that it was proposed specifically as a method for going from attack trees to penetration test cases. Another reason was that there was also a wish in the research group to test out the HARM method in practice.

3.4.4. Case study

Case study research is concerned with the researcher gaining an in-depth understanding of particular phenomena in real-world settings. This thesis relies on case studies as described by Yin (Yin, 2002; Yin, 2017). According to Yin, case study is, *"An empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident."* (Yin, 2002) (p. 13), Evidence is typically gathered from a large number of sources like documentation, interviews, direct and participant observation, and physical artifacts.

Yin proposes (Yin, 2002) a process consisting of the following phases:

1) Case study design: objectives are defined, and the case study is planned,
2) Preparation for data collection: procedures and protocols for data collection are defined,
3) Collecting evidence: execution of data collection on the studied case,
4) Analysis of collected data,
5) Reporting

Specific method choice in our research: We have conducted single case studies investigating requirements/features for e-exam systems in P5 and P7. The first two phases in the process proposed by Yin for planning case studies are provided here. Before data collection, we defined research questions for both studies. The case study for P5 was intended to be conducted in the local context at NTNU, while P7 was targeted in the national context with the Higher education sector in Norway. However, the vendors of the studied e-exam systems have customers in several countries, in Europe and beyond. Hence, the case study would likely be of much broader interest. Since our case studies were restricted to Norway, some findings may be specific to the national setting or even to our local university. The data collection in P5 was based on our practical usage experience with the features of e-assessment tools. To gain more practical knowledge on the requirements/features of e-assessment tools, we aimed for expert interviews in P7. We planned for semi-structured interviews in P7, so questionnaires were prepared in advance for participants (vendors, process managers, and system managers at universities) depending on their roles. The questionnaires were designed with some closed questions to improve consistency between interview responses and research questions and some open questions to get other important information from participants that we might not have thought about; to give participants freedom to express their own viewpoints. Data collection for P7 was planned to be done with vendors, system managers and process managers, which notably omits students and teachers. However, the requirements of teachers for e-assessment systems would likely differ a lot from person to person, e.g., depending on the discipline taught and the assessment practice followed. There will also be much variation among students, e.g., based on personal preferences and special needs. The advantage of interviewing system managers at a university is that they receive feedback from many students

and employees, e.g., concerning dissatisfaction with the system or requests for new functionality, thus they should be able to present a more aggregate idea of requirements. Similarly, process managers who acquire e-exam systems on behalf of universities and vendors who sell e-assessment systems to several universities should also possess such aggregate views on needs. Hence, these three groups of people should be knowledgeable about general trends concerning requirements for e-exam software.

Further detail about data collection and analysis (3-5 phases) in the case studies will be provided in section 3.5.

3.4.5. Mixed-method research

Mixed-method research could enable balance with joint analysis and triangulation of quantitative and qualitative data (Creswell & Clark, 2017). Creswell defines mixed method research as *"Combination or integration of quantitative and qualitative research and data in a research study (...) resided in the idea that all methods had bias and weaknesses, and the collection of both qualitative and quantitative data neutralized the weakness"* (Creswell, 2014) (p.563). Creswell and Clark (2017) propose a mixed-method process which entails that the researcher

- collects and analyses both qualitative and quantitative data rigorously in response to research questions and hypotheses,
- integrates (or mixes or combines) the two forms of data and their results,
- organizes these procedures into specific research designs that provide the logic and procedures for conducting the study, and
- frames these procedures within theory and philosophy.

Quantitative surveys could be used alone to examine the participants' attitudes, beliefs, opinions, or practices at one point in time (Creswell, 2013). These can reach many more respondents but have weaknesses, especially if respondents may have trouble answering the questions. An example of this in our research would be questionnaires about cheating, where participants with little or no personal experience with cheating, might have trouble answering questions about the relative ease of various ways of cheating. Whereas in interviews, the interviewer can directly clarify any questions that are obscure and can also ask respondents to expand on answers that are unclear (Fraenkel, Wallen, & Hyun, 2011). Although interviews allow for more open questions and more in-depth answers, they are time-consuming to conduct and analyze, thus reducing the number of respondents that can realistically be involved.

Specific method choice in our research: Mixed-method research was conducted for P6, combining questionnaires to a larger number of persons and interviews with a much more limited number of persons. While P7 only employed interviews, P6 was more suitable for a quantitative part than P7 since it focused on comparing ease of cheating and ease of defending against cheating for paper exams vs e-exams. P7, on the other hand, did not seek a similar comparison, rather exploring what are the key

requirements/features for e-exam systems, and how the requirements process is being conducted.

We chose to conduct research with both students and teachers for P6, as the perceptions of these two groups would complement each other. Students may not have complete knowledge about the ease of cheating, nor of the university's countermeasures, though they would likely have some insights either from their own experience or from peers. Teachers may not directly catch students while cheating, as exams are mostly proctored by other employees in Norwegian universities. Hence, responses from teachers would mainly be on their experience with catching or suspecting cheating when grading answers and possibly considerations they have made about the effectiveness of countermeasures when designing tests so as to be hard to cheat on.

Further detail about data collection and analysis in the mixed methods research will be provided in section 3.5

3.5. Data collection and analysis

This subsection gives more details of data collection and analysis, particularly for the studies reported in P6 and P7. The quantitative and qualitative data were collected using inductive and deductive approaches (Braun & Clarke, 2006; Twining et al., 2017). Before collection of the data from survey respondents and interview participants, authors notified Norwegian Centre for Research Data (NSD) for their approval of the studies, and data were only collected after such approval was received, see Appendix for approval, information letter and consent form. Table 3 provides the summary of data collection and data analysis method applied corresponding to research questions and contributions. This section provides more detailed explanation of the data collection process.

3.5.1. Data collection with students and teachers

Data collection for SQ1 (P6) was conducted with students and teachers at the researcher's own university (NTNU) through questionnaire surveys and semi-structured interviews. This study aimed to compare ease of cheating in three types of written examinations: paper exams, e-exams using university-owned equipment (which would typically be thin clients) and e-exams using student-owned laptops (so-called Bring Your Own Device – BYOD). The study also sought to investigate perceptions about the effectiveness of some typical countermeasures towards cheating across these examination types. The questionnaire part of the survey was carried out from Nov 2018 – Jan 2019 using the SelectSurvey online tool (SelectSurvey, 2008) with web-based questionnaires. Teachers and students were from various departments at the NTNU: computer science, electronic systems, electric power engineering, cybernetics, information security and communication technology, mathematical science, natural sciences. Students were invited through mass emails to various student groups at NTNU. A total of 259 students responded to the email invitation, but 47 of these skipped the survey. Of 212 participants, 149 completed the whole survey while 63 only partially completed it. Out of the 149 fully

completed, 84 were computer science students who responded to the survey by direct administration in a classroom, while others responded remotely on the web. Teachers were invited to participate by emails containing a link to the web-survey. A total of 736 teachers were invited, whereof 197 responded, though 35 of these skipped the survey. 162 then participated, whereof 95 (13% of those invited) completed it fully and 67 partially. Participation in the survey was voluntary, and the participants were assured that their data would be confidential and treated anonymously. Additional demographics such as age, gender, nationality were not collected from the respondents. Partly, the motivation for omitting such data was to keep the questionnaire short, and especially for students, it was also to avoid fear that respondents admitting to cheating might be identified via the survey. Given the general demographics of the study programs in question, the majority of the students surveyed would likely have been in their early twenties, Norwegians, and about 2/3 male. Thus, respondents deviating from the average demographics – e.g., being female in a male-dominated field of study, and/or being of somewhat older age than the typical student, might fear indirect identification from responses about demographics.

The interviews were conducted with six 3rd year bachelor students in February and March 2019 and five teachers in September 2019. All the interviewees were from the Computer Science department. Participants received invitation emails explaining the purpose of the interview, and selection was made based on their experience with e-exams in different courses (e.g., how many e-exams they had experience with). All participants had experience both with paper exams and e-exams, since e-exams had been broadly adopted at the NTNU just recently, and not in all courses. All the participants were informed of, and consented to, the audio recording of interviews. Interviews were designed as semi-structured so that participants could be prepared up front receiving the pre-planned questions by email before the interview. There were also other questions beyond the pre-planned ones, e.g., follow-up questions. On average, these interviews lasted approx. 40 minutes.

3.5.2. Data collection with e-exam system vendors and managers

Data collection for SQ2 and SQ3 (P7) was conducted through a case study on e-exam tool support in Norway with vendors, system managers, and process managers through semi-structured interviews. This study aimed to identify key requirements for e-exam systems and enablers and barriers towards achieving an open digital ecosystem around e-assessment in larger ecosystems of e-learning. Vendors (i.e., development managers, head of product development, product managers) were from two mass-market software product companies, Inspera Assessment and WISEflow, both delivering systems used for e-exams in Norway (and elsewhere). The Ministry of Education and Research employed the process manager, Unit (Directorate for ICT and Joint Services in Higher Education and Research) to run joint procurement processes on behalf of all public universities. The group labelled “System managers” consisted of license administrators, project managers, team leaders, advisors, and engineers doing e-exam technical support from several Norwegian universities.

Figure 6 illustrates the process of data collection and analysis used in the case study with providers and managers. Based on the research questions of the study, suitable participants were identified. We used a combination of key informant sampling (contacting persons known to be in central roles, with expertise on the topic) and snowball sampling (getting suggestions from initial participants about other potential participants), to cover both providers Inspera Assessment and WISEflow, and a selection universities using each system. All in all, we interviewed n=12 participants from 7 different organizations. Participants were providers, process managers, system managers at universities who were primarily involved in the acquisition and implementation of e-exam systems, and/or in the planning, support, and execution of digital exams. All consented to have their interview data researched and published in anonymized form.

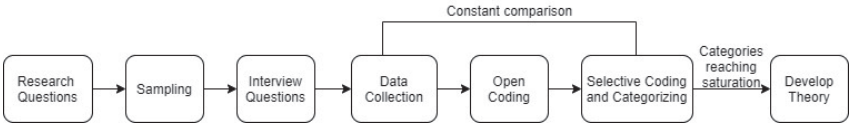

Figure 6. Figure illustrating case study design

Semi-structured interviews (Kallio, Pietila, Johnson, & Kangasniemi, 2016) were chosen as the most appropriate for our purpose, keeping some structure for comparability between participants while at the same time allowing for participants to bring forth issues we might not have thought about. Interview guides were prepared and distributed to participants before the interviews. All interviews were done by the first author, during April 2019 – Aug 2020, some face-to-face (4) and some (8) via Skype and Zoom video calls. The reason for using video calls was partly as this was recommended March 2020 onwards due to the Covid-19 pandemic, and partly because several respondents were in other cities than Trondheim, so that face-to-face interviews would have led to extensive travelling. Each interview lasted approx. 40 mins.

3.5.3. Data analysis

Statistical Package for Social Sciences (SPSS) version 25 (SPSS, 2017) was used to analyze the questionnaire data. For comparing various types of examinations, one-sample and independent t-tests were used, with the neutral alternative 3 as test value, to check if respondent preference went significantly to one of the sides (e.g., whether one type of examination was perceived as allowing for easier cheating than another). Similarly, for comparing student and teacher perceptions, independent t-tests were used to check if differences were significant, for more details on validity of tests, see sec. 5.3.5.

Two different qualitative analysis tools were used for extracting the data from interviews with students, teachers for P6, and for interview data from providers, and managers for P7. For the analysis of data from students and teachers, Atlas.ti 8 was used. For analysing data from providers and managers, NVivo 12 was used. First, the interviews were transcribed line by line by the first author. The constant

comparative method was used to extract data from interviews (Corbin & Strauss, 1990). This has the advantage of making the analysis more explicitly theoretical (Urquhart, Lehmann, & Myers, 2010) and encourages the researcher to be both rigorous and theoretical. It was first proposed by Glaser, Strauss and Strutzel (1968) in their grounded theory methodology and further practically explained by others (Charmaz, 2006). We did not use grounded theory in full but used constant comparative analysis outside of grounded theory. We used it for data analysis in the same way as (Fluck, 2019). The purpose of our qualitative studies was to focus on answering the paper's research questions rather than identifying emerging theory (Boeije, 2002).

We had a set of predefined categories from our research questions of P6 and P7. Hence, analysis was focused on finding relations between the concepts that emerged from the analysis and grouping those concepts under predefined categories. First, data analysis involved open coding of the responses. This open coding abstracts concepts from the data. Second, axial coding was performed to group the codes from open coding and further categorized the codes. Lastly, selective coding was conducted to interpret the relation between codes and categories from axial coding. The data collection, coding, and analysis were done together to enrich the existing category. During the coding process, the constant comparison of data and codes was made to compare responses and decide what data will be gathered next until the data got saturated.

3.6. **Research ethics**

In recent times, there has been increasing regulation of educational research when data is collected (Hammersley & Traianou, 2012; James & Busher, 2015). This section discusses research ethics considered in this thesis.

Informed consent is one of the pillars of research ethics. As said earlier in this chapter, research conducted for this thesis has been notified to NSD and received prior approval from NSD, to collect and process data from participants. All participants were informed of the ethical considerations of the studies. In online survey questionnaires, a check box was provided to make sure that participant would be aware of providing their consent for processing and publishing their data and signed consent letters were collected from directly administered survey participants. Whereas, during the interviews, the participants were reminded of their consent. Informed consent does not decrease risk to research participants, but the participants should be given the opportunity to decide whether they can take those risks (Macnish & van der Ham, 2020). Further, we group our discussion of ethical issues which should be considered for the research conducted in this thesis in the following categories.

- Potential harm to research participants (e.g., interview subjects revealing sensitive information about themselves, which could later be used against them)

- Potential harmful effects of the publication of results of the research (e.g., students or criminals obtain ideas for ways to cheat based on this thesis work which they would not have gotten otherwise).

We cover details of how we mitigated these two ethical issues below.

Potential harm to research participants. Although we tried to avoid too sensitive questions about cheating and software vulnerabilities in the surveys and interviews, there are always certain issues that may come up concerned to the research ethics. We have collected most of the empirical data for surveys and interviews online. First, privacy and confidentiality of the data should be protected (Stockley & Balkwill, 2013). For confidentiality, survey data were kept in the Selectsurvey data processor until the data has been published. Concerning confidentiality of interview data, data has been deleted from the researcher's university laptop after the research was published. Data collected for this thesis is made private. Still, anonymity and privacy might get affected when technology is used as the data processor. However, Selectsurvey is university-owned software, not free software, and it performs authentication of the user and logs user actions when the user accesses the tool. The literature also shows that anonymization is another important process that needs to be applied before analyzing the online data (Dawson, 2014). We have anonymized data before analysis. We have not published any data that reveal the identity of the participants. From the interviews, direct quotes were published still anonymity was ensured by not revealing participant identity, e.g., names, email addresses, and neither does the published data contain information that could be combined into inferring the identity of participants. In paper 7, Interview participants position details were provided in the appendix to show the expertise of the participants. Still, the position was not revealing the direct identity of participants. On the other hand, a reader who has been heavily involved in digital exam processes in Norway might be able to infer who some of the interview participants have been, since they are recruited from a much more limited set of experts / key stakeholders in various universities, vendor organizations and process manager - Unit, rather than from a larger body of students. However, the interview subjects in P7 participated as professional experts and did not respond about sensitive issues (e.g., own experience with cheating), rather about their experiences with software development, acquisition and operation processes. Investigator triangulation (Twining et al., 2017) was performed on the collected data, which involved another researcher in the analysis together with the investigator. However, information is kept confidential during triangulation. We could have published identifying data, but this might have highly impacted the sample and their response since the questions were mainly related to cheating, software issues and vulnerabilities, participants might not be interested in providing such information.

Potential harmful effects of the publication of research results. It is impossible to rule out that somebody got new ideas for cheating by reading this research. For instance, the discovery that SEB could be circumvented by a Skype conversation if one had it running before starting SEB (Cf. Paper 2). However, researchers contacted SEB about the weakness as soon as they found it, long before the paper was published. Moreover, not publishing it might also have been ethically wrong. There are e-exam systems built on top of SEB, for instance, Inspera Assessment, and for

vendors of such systems, it might be important to know about any such weakness of SEB so that they know that they must themselves add extra safeguards in their software to handle threats that SEB does not handle. Not publishing results could have left software vendors unaware of issues that they should have known. Macnish and van der Ham (2020) provided an overview of some of the ethical issues where they note that on small-scale, limited participation experiments researching without prior informed consent when the harms are minimal is typically taken to be acceptable. Robinson and Halderman note "if researchers know about problems and there is sufficient time to mitigate them, they may have an obligation to publicly disclose them." (Robinson & Halderman, 2011) (p 127). They further note that when researcher studied system without authorization, advance notice about researcher findings to vendors prior to public disclosure, may run the risk of lawsuits attempting to suppress publication of results, hence researchers often choose not to disclose problems to vendors in advance of publication.

4. Results

The research of this thesis project has resulted in seven papers. Of the seven published papers, three are in journals, two in conferences, one in a workshop, and one is a book chapter in an anthology. The papers are presented in Part II of this thesis, following guidelines from the editors of the venues. This chapter summarizes the main findings of research conducted for the papers. For each paper the following information is given:

- Title
- Authors' names
- Publication Venue (Where the paper is published)
- CRediT authorship contribution statement (cf. https://casrai.org/credit/)
- Abstract of the paper
- Main findings of the paper
- The paper's relation to the research questions of the thesis

It should be noted that in the midst of the project, the PhD candidate's last name changed due to marriage, hence being Vegendla for Papers 1-4 and Chirumamilla for Papers 5-7.

4.1. Paper 1

Title: E-exams versus paper exams: A comparative analysis of cheating-related security threats and countermeasures

Authors: Guttorm Sindre, Aparna Vegendla

Publication Venue: Norwegian Information Security Conference (NISK), Volume 8, Pages 34-45, 2015

CRediT authorship contribution statement:

- Aparna Vegendla: Methodology, Validation. Writing – Review & Editing
- Guttorm Sindre: Conceptualization, Methodology, Validation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision.

Abstract: E-exams can have a lot of advantages over traditional paper-based exams, and if using a BYOD approach (Bring Your Own Device) they can also scale to large classes and peak exam days. At the same time, BYOD adds extra security challenges by using studentcontrolled equipment. To be viable, BYOD e-exams need not have perfect security, only about the same level of security as paper-based exams have. This article uses attack-defense trees to provide an analysis comparing the threats and countermeasures against cheating at controlled exams with paper-based exams versus BYOD e-exams. The conclusion is that neither has a clear advantage from a security perspective.

Main findings of the paper: This paper investigates cheating threats for on-campus exams by comparing threats and countermeasures in e-exams and paper exams to provide security requirements for BYOD e-exams. The assumed setting for the comparison is the typical large exam hall with many students taking exams at the same time, and invigilators to prevent cheating – each invigilator responsible for several students. For the paper variant, students will be provided with a printed question set at the start of the exam, as well as with dedicated sheets on which to write their answers – which will be handed into invigilators at the end of the exam. For the e-exam, students receive and answer questions via a PC, and BYOD (Bring Your Own Device) entails that a student-owned laptop is used for this purpose, rather than university-owned equipment. Communication with the exam server is typically through wireless internet, and lock-down browsers are typically used to prevent students from using the internet in other ways than communication with the exam server. Apart from the difference paper vs. student laptop, the comparison assumes equal conditions, e.g., same ratio of students per invigilator.

The paper used Attack-defense trees (ADTrees) in its analysis, where the countermeasures can be represented through defense nodes in addition to threats as attack nodes. The analysis was focused on three threats: impersonation, assistance/collaboration, and using unallowed aids for the exam. On the one hand, the analysis shows that e-exams do have more opportunities for cheating. In addition to still being able to perform old-fashioned cheating (e.g. cheat notes, whispering), students may also cheat via the exam PC. For instance, forbidden aids might be stored in the PC, or accessed via the web, and the PC might also be used for illegitimate collaboration with peers or getting assistance from an outsider. Such cheating would be easier with a BYOD exam, where the student owns the device and may have rigged it for cheating before the exam, than for e-exams using university owned equipment that the candidates do not have the ability to manipulate before the exam. On the other hand, e-exams also offer more countermeasures against cheating. Partly, there may be entirely new countermeasures, such as biometry for candidate authentication (rather than relying on a manual inspection of the old-fashioned student ID card), and automated monitoring of student actions during the exam. Moreover, some countermeasures are possible but logistically demanding and thus expensive for paper exams, but easier to implement for e-exams that include:

- Mixed seating: Rather than having all the students doing a CS1 exam sitting in one room or one area of the exam hall, then all those doing Physics sitting in another room or area, etc., mixed seating would imply seating students in such a way that nobody has a close neighbour doing the same exam. This would make a number of old-fashioned close-range cheating approaches more difficult, e.g. whispering, passing of notes, peeking at neighbour answers, using combinations of hand and feet positions to convey answers to Multiple Choice Questions s, etc. Mixed seating is clearly possible for paper exams, but may be logistically complicated for the distribution of question sets, enforcing rules for aids, timing etc. (if inter-mingled exams have differences in this respect). For e-exams it is easier because no paper needs be distributed.

- Variation in question sets: Peeking can be mitigated by shuffling the order of questions and answer options (esp. for Multiple Choice Questions, where otherwise the answer pattern migth be easily peeked), and even more if students are given different questions altogether, for instance as variations over some template. This is also possible with paper exams, but often with practical limitations related to printing, copying, and distribution – for instance having three different variants of the question set. With e-exams, with no paper involved, it is easier to support much larger variation, like each student having an entirely unique question set, by pulling questions randomly from a bank as the exam takes place.
- Moving calculators and books into the exam system: Calculators are often indicated as a common vehicle for cheating, for instance, by storing forbidden information on the calculator or illegitimately using an advanced calculator with communication features. Instead having a calculator app inside the e-exam system, extra calculator devices could be prohibited. Simiarly, exams that allow some books (e.g., formula collections in Physics) can have challenges with students hiding additional information in the same book. Instead having the formula collection as an electronic attachment provided by the e-exam system would mitigate this challenge.
- Strict question/answers sequence: This is also called forced navigation by some, meaning that the candidate only sees one question at a time and cannot navigate back to previous questions. The opposite would be free navigation, where a cheating candidate can identify at the outset of the exam the questions for which help will be needed, and then somehow outsource these to peers or outsiders, to collect answers in due time. With forced navigation, help for each question must be obtained as that question emerges in the sequence, thus the cheater needs to establish a more elaborate communication with the helper, with higher risk of being caught, and the helper has less time to develop a good answer, since the cheating candidate cannot progress with the exam until the answer for that question arrives.
- Automated plagiarism checking: with digital answers, tools for plagiarism checking can be effectively used, not only for direct copy paste, but also for various rephrasing tricks. Hence, plagiarism carries a higher risk of getting caught in e-exams than for traditional paper exams, where discovery often hinged on the grader recognizing the plagiarized source. On the other hand, it must also be acknowledged that plagiarism is much more quickly performed with e-exams, e.g., copy-paste rather than rewriting the plagiarised text by hand.
- Biometric authentication: having somebody else sit the exam for you - is a rare cheating threat, but potentially the most effective of all if uncaught, since even an F candidate can get a perfect A if the impostor has strong subject knowledge. Authentication with ID card is often insecure as

candidates could fake ID cards. Biometric authentication could give far better in detecting impersonation and assistance/collaboration between candidates. Of course, this could be employed both for paper and e-exams, but e-exams have an advantage that infrastructure for the authentication is then anyway in place, e.g., using recognition of face, voice, and keystroke dynamics via the same PC that is used for answering the questions. For a paper exam, equipment for such authentication would instead have to be provided by the university and e.g., carried around by the proctors, giving extra cost. Although, our results show that enforcing biometrics were not seen as the only important approach to prevent cheating, they were seen as effective together with improved assessment design with variety of questions and increased surveillance.

The paper's relation to the research questions and contributions of the thesis: This paper addresses SQ1 and C1 through a comparative analysis of cheating threats and countermeasures in e-exams and paper exams using ADTrees. The paper provided a starting point for later technical and empirical investigations, namely the penetration testing done in P2, the questionnaire investigations about cheating with students and teachers in P6, the case study about cheating with vendors, managers in P7. It was also followed up with similar analytical investigations about on-campus versus take-home exams in P4.

4.2. **Paper 2**

Title: Extending HARM to make Test Cases for Penetration Testing

Authors: Aparna Vegendla, Thea Marie Søgaard, Guttorm Sindre

Publication Venue: Proceedings of Advanced Information Systems Engineering Workshops: CAiSE, volume 249, pages 254-265, 2016.

CRediT authorship contribution statement: This paper refined the results of the project of a master student (Søgaard). Vegendla did the detailed supervision of Søgaard, and Sindre had overall supervision of the work. Credit authorship contribution for this paper is:

- Aparna Vegendla: Conceptualization, Methodology, Software, Validation, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration.
- Thea Marie Søgard: Conceptualization, Methodology, Software, Investigation, Writing – Review & Editing.
- Guttorm Sindre: Resources, Writing – Review & Editing, Supervision.

Abstract: [Context and motivation] Penetration testing is one key technique for discovering vulnerabilities, so that software can be made more secure. [Question/problem] Alignment between

> modeling techniques used earlier in a project and the development of penetration tests could enable a more systematic approach to such testing, and in some cases also enable creativity. [Principal ideas/results] This paper proposes an extension of HARM (Hacker Attack Representation Method) to achieve a systematic approach to penetration test development. [Contributions] The paper gives an outline of the approach, illustrated by an e-exam case study.

Main findings of the paper: The study took attack-defense trees for cheating through e-exam systems as a starting point, and used the HARM method (Karpati et al., 2013) to develop test cases for the attacks. The purpose of the study was both to find out whether the HARM method was suitable for this task, and to find out more about the actual vulnerabilities of a key e-exam technology, namely the Safe Exam Browser (SEB) (Schneider, 2014) – a lock-down browser which had been used as a cheating countermeasure in many e-exam systems. The results shows that SEB effectively prevented some of the attempted cheating approaches:

- Starting the exam from a virtual machine was not possible with SEB. Otherwise, an easy cheating approach would have been for the candidate to thereafter switch from the locked-down virtual machine to the unlocked real machine, where websites, chat, email etc. could then be accessed.
- Pasting text from the clipboard was impossible. Otherwise, in a closed book exam, the candidate might have copied important information to the clipboard just before the exam, then pasting it into e.g., a free text essay answering field during the exam, from where it could be further used to answer various exam questions.
- Using remote desktop control was neither possible. Otherwise, a student might share the exam questions with an accomplice outside the room and receive help with answers via remote keyboard control.

On the other hand, several of the attempted test cases succeeded in bypassing SEB. These include:

- Injecting text into the exam answer (e.g., using a free text essay field) by means of a USB keystroke injector. This vulnerability might be utilized by cheating exam candidates in several different ways. One example would be in closed book exams to have stored important material on the injector before the exam. Another usage might be to obtain a keystroke injector from a peer or outsider during the exam, containing answers to questions. As USB injectors are rather small objects, they might be passed around when the invigilator is looking another way, or hidden in a pre-agreed spot e.g., in the restrooms.
- Running SEB on a remote computer.
- Accessing SEB on a remote computer.

- Communicating with audio/video conference. Opening the video conference application Skype after having gone into lock-down, was impossible, however if having the Skype call already running and then starting SEB, the Skype conversation turned out to be still live afterwards, enabling speaking as well as written chat communication between candidate and helper (e.g., outside accomplice or other candidate)

We informed SEB developers about the vulnerabilities as soon as we had found them. Since several new versions of SEB have emerged after these tests were done in 2016, it is likely that many of the vulnerabilities have been addressed. It should also be noted that changing SEB is not the only possible way to address the found vulnerabilities, some might also be addressed in the e-exam systems working on top of SEB

The paper's relation to the research questions and contributions of the thesis: This paper addresses SQ1 and C1 by using penetration testing to investigate the viability of cheating-related threats and attacks during e-exams. Thereby, the results enrich the understanding of security vulnerabilities of current e-exam technology, more specifically lock-down browsers, which are a central component in many BYOD e-exam setups, hence enabling more precise specification of security requirements relating to cheating.

4.3. **Paper 3**

Title: A Systematic Mapping Study on Requirements Engineering in Software Ecosystems

Authors: Aparna Vegendla, Anh Nguyen Duc, Shang Gao, Guttorm Sindre

Publication Venue: Journal of Information Technology Research (JITR), Volume 11, Pages 49-69, 2018.

CRediT authorship contribution statement:

- Aparna Vegendla: Conceptualization, Methodology, Validation. Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.
- Anh Nguyen Duc: Formal Analysis, Validation, Writing – Review & Editing.
- Shang Gao: Writing – Review & Editing.
- Guttorm Sindre: Validation, Writing – Review & Editing, Supervision.

Abstract: Software ecosystems and open innovation processes have been claimed as a way forward for the software industry. A proper understanding of requirements is as important for SECOs as for more traditional ones. This article presents a mapping study on the issues of RE and quality aspects in SECOs. Our findings indicate that among the various phases or subtasks of RE, most of the SECO specific research has been accomplished on elicitation,

analysis, and modelling. On the other hand, requirements selection, prioritization, verification, and traceability has attracted few published studies. Among the various quality attributes, most of the SECOs research has been performed on security, performance and testability. On the other hand, reliability, safety, maintainability, transparency, usability attracted few published studies. The article provides a review of the academic literature about SECO-related RE activities, modeling approaches, and quality attributes, positions the source publications in a taxonomy of issues and identifies gaps where there has been little research.

Main findings of the paper: The paper mapped 44 articles published in 6 journals, 16 conferences, 7 workshops, and in one book chapter. The mapping study revealed that at least 3 papers focused on RE in software ecosystems (SECOs) each year from 2009-2016 (cf. Table 4 from P3). The peak publication period is from 2013 to 2015 (22 papers, around 50%). The publication channels contributing more than two studies were: Requirements engineering (RE) conference, European conference on software architecture (ECSA), Journal of systems and software (JSS), and International workshop on software ecosystems (IWSECO).

Given that the field of SECOs was still growing at the time of this study, findings show the relative importance of RE in SECOs. Most of the collected papers mainly discussed software development in SECOs in the context of open-source SECOs, where the main focus was on technical and architectural views rather than business and strategic views essential for RE. Many papers addressed RE activities of SECOs were elicitation, analysis, and specification, whereas requirements verification, validation, and management have attracted few published studies. Identifying stakeholders' roles and relationships was the most addressed topic for requirements elicitation in SECOs, and the goal modelling approach was favoured for elicitation. A few studies proposed modelling approaches for requirements negotiation in SECOs, using negotiation and network theories. Moreover, most of the studies proposed meta-models, conceptual architectures, and formal models for both requirements negotiation and specification. However, there were no publications on tools specific for conducting RE process for SECO.

Our findings indicated that usability, reliability, maintainability, security, safety, performance, transparency, and testability were the key topics addressed on quality attributes for SECOs. Usability, security, safety, and reliability were most addressed quality concerns during the platform operation phase, while maintainability, performance, and testability concerns were addressed during the platform development phase. Security was the most addressed quality attribute in SECOs. The results correspond to the mapping between quality attributes with the research topics studied in SECOs was:

- Usability - software platform usability and extendability
- Safety - embedded open-source SECOs.
- Security - certification and policy management, security patterns, software redundancy, authentication, accountability, transparency.

- Reliability and maintainability - dependencies between unreliable components.
- Performance - developer performance, performance management of SECOs, ecosystem health - robustness, productivity, interoperability, stakeholder satisfaction, and creativity
- Testability - software testing requirements for mobile applications, static analysis of vulnerabilities in web applications and their plug-ins, review and approval methods for platform extensions.
- Transparency - RE confidentiality.

The paper's relation to the research questions and contributions of the thesis: This paper mainly addresses SQ2 and C2. It presents a review of issues and potential research gaps in requirements engineering for software ecosystems through a systematic mapping review. Further, the findings from this study contributed to guiding the investigation of SQ3. A good overview of issues and concepts of requirements processes was essential both for shaping the interview guide for P7, and for interpreting the data from interviews.

4.4. Paper 4

Title: Mitigation of Cheating in Online Exams: Strengths and Limitations of Biometric Authentication

Authors: Aparna Vegendla, Guttorm Sindre

Publication Venue: In the book titled Kumar, A. (Ed.), "Biometric Authentication in Online Learning Environments". IGI Global, Pages 47-68. 2019

CRediT authorship contribution statement:

- Aparna Vegendla: Conceptualization, Methodology, Validation. Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.
- Guttorm Sindre: Validation, Writing – Review & Editing, Supervision.

Abstract: E-exams have different cheating opportunities and mitigations than paper exams, and remote exams also have different cheating risks than onsite exams. It is important to understand these differences in risk and possible mitigations against them. Authenticating the candidate may be a bigger challenge for remote exams, and biometric authentication has emerged as a key solution. This chapter delivers a categorization of different types of high stakes assessments, different ways of cheating, and what types of cheating are most relevant for what types of assessments. It further presents an analysis of which threats biometric authentication can be effective against and what types of threats biometric authentication is less effective against. Insecure aspects of various biometric authentication

approaches also indicate that biometric authentication and surveillance should be combined with other types of approaches (e.g., how questions are asked, timing of the exam) to mitigate cheating.

Main findings of the paper: The risk analysis in this paper was mainly focused on cheating with three different methods: impersonation, assistance/collaboration, and using unallowed aids for the exam during three different types of exams: written exams, oral exams, and in unproctored course work. Our findings indicate that remote online exams do introduce several new cheating threats not because they are e-exams, rather because they are unproctored, or with a remote proctor rather than having a proctor in the same room. The analysis resulted in four different attack-defense trees for various types of exam: online oral, on-site written, online written, and unproctored coursework – as well as an additional attack-defense tree specifically addressing attacks to circumvent biometric authentication.

Biometrics can be easier to implement in online exams with built-in authentication features that come with e-exam tools. Biometric authentication seemed to be mainly intended to address the impersonation risk. It will be ineffective if the accomplice stays outside the view of surveillance cameras during assistance/collaboration in exams. Hence, in the case of assistance/collaboration, biometrics mainly helps to increase the hassle of cheating rather than fully prevent it. Biometrics have no use for detecting forbidden aids (e.g., cheat notes, searching for answers on the internet, etc.), however, it could still be used to detect the movements of candidates. Biometrics are just one of several potent mitigations against cheating for online exams, so additionally, they do require an increase in surveillance (both human online proctors and technological surveillance) and different variations of question types (e.g., oral follow-up questions, to mitigate cheating). Moreover, while onsite exams have traditionally been believed to be quite secure compared to remote exams, the book chapter observes that this is not necessarily true. The vulnerability of on-site exams towards cheating may be increasing due to the emergence of cheap communication equipment (e.g., tiny, wireless earpieces) that can enable candidates to get substantial help from accomplices outside the proctored exam venue yet go unnoticed by invigilators. This increased availability of cheating technology through devices means that the cheating threats of onsite exams vs. remote exams which were previously assumed to be quite different, are gradually converging. In addition to the attack-defense trees, the book chapter also provided three lists of recommendations for how to conduct remote oral exams, remote written exams, and unproctored graded coursework – beyond technical countermeasures, focusing on question genres and the exam process.

The paper's relation to the research questions and contributions of the thesis: This paper addresses SQ1 and C1. It shows a literature review, especially on mitigate cheating in online exams using biometric authentication. It provides recommendations to avoid cheating during written exams, oral exams, and in formative assessments conducted as part of course work (i.e., term papers, projects, etc.) which may both inform further research on mitigation of cheating and inform examination practice.

4.5. Paper 5

Title: E-Assessment in Programming Courses: Towards a Digital Ecosystem Supporting Diverse Needs?

Authors: Aparna Chirumamilla, Guttorm Sindre

Publication Venue: In Proceedings of Conference on e-Business, e-Services and e-Society (I3E), volume 11701, Pages 254-265, 2019.

CRediT authorship contribution statement:

- Aparna Chirumamilla: Conceptualization, Methodology, Validation. Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.
- Guttorm Sindre: Validation, Writing – Review & Editing, Supervision.

Abstract: While a number of advantages have been discussed on digital ecosystems, little research has been reported on the e-learning/e-assessment domain so far. Today, the different types of questions have been used in exams based on course type, e.g., text-based questions, mathematical questions, and programming questions. All these question types require supporting plug-ins for e-assessments. This study provided our practical experience on programming exams in Inspera Assessment and Blackboard Learn, focusing on Parsons problems (drag-and-drop questions) and code writing questions. Our findings indicate that the tools have basic support for programming exams, and there is a low-level integration between the tools. However, the adaptability of any exam system could depend on the interoperability between platforms and external plug-ins. Hence, more improvements can be made with implementing e-assessments in digital ecosystems while applying a lot of changes internally and externally to the institutions. In the paper, we explained how a digital ecosystem within e-assessment could improve assessments and how it supports the diverse needs of programming exams.

Main findings of the paper: Our observations revealed that at the time the paper was written, the e-exam systems Inspera Assessment and WISEflow and the learning management systems Moodle and Canvas all included some support for drag-and-drop questions, but not in an ideal way for programming tests using Parson problems. From the tools used at the NTNU, Inspera Assessment had better question type support for Parsons problems than Blackboard had, but still with substantial limitations. The user interaction for drag and drop questions was somewhat tedious for students, especially if reordering, and teacher authoring of questions was even more tedious. For code writing questions, both had the shortcoming that the code would not run and must be manually graded, and Blackboard did not even have syntactic support. The findings from web documentation show that WISEflow, Canvas, and Moodle also supported the basic functionality of drag-and-drop questions but not ideal for Parsons problems. Code writing was supported in both

Inspera Assessment and Moodle, both had syntactic support (i.e., code indentation and highlighting), and it seemed Moodle had better support than Inspera Assessment[10]. Moodle had a plugin (CodeRunner) that allowed comparing the candidate's code to a test suite and scoring according to the number of tests passed, whereas, for Inspera Assessment[11], the Programming question genre did not allow compiling and test running. The authoring tool supported in Inspera Assessment and Blackboard could also be improved by supporting third-party extensions and plug-ins with the adoption of open API in digital ecosystems. Our findings revealed that Blackboard would need an integrated plug-in to support drag-and-drop question type. Inspera Assessment would also need an integrated plug-in for better support of Parsons problems and code writing. Especially, Inspera offers open APIs and supports open standards for customers to build specific question types. However, integration and interoperability between the systems still remained a challenge as stakeholders (tool vendors, users) refrained from considering essential technological and organizational aspects to prioritize the basic functionalities.

The paper's relation to the research questions and contributions of the thesis: This paper contributes to SQ3, C4, and C5. It identifies enablers and barriers for achieving open digital ecosystems for e-exams within a larger ecosystem of e-learning.

4.6. Paper 6

Title: Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway.

Authors: Aparna Chirumamilla, Guttorm Sindre, Anh Nguyen Duc

Publication Venue: Assessment & Evaluation in Higher Education, Volume 45, Pages 940-957, 2020.

CRediT authorship contribution statement:

- Aparna Chirumamilla: Conceptualization, Methodology, Validation. Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.
- Anh Nguyen Duc: Validation, Writing – Review & Editing.
- Guttorm Sindre: Resources, Validation, Writing – Review & Editing, Supervision.

Abstract: A concern that has been raised with the transition from paper examinations to electronic examinations is whether this will make cheating easier. This article investigates how teachers and students perceive the differences in ease of cheating during three types of written

[10] Inspera was working on another question genre, Code Compile, which would allow such test runs of code both during the exam and during grading, but at the time this paper was published, that feature was not yet released. It emerged as Beta functionality in 2020, expected to go into normal usage some time in 2021.

examinations: paper exams, bring your own device (BYOD) e-exams and e-exams using university-owned devices. It also investigates perceptions about the effectiveness of some typical countermeasures towards cheating across these examination types. A mixed-method approach was used, combining questionnaires and interviews with students and teachers from STEM disciplines in the authors' own university. A total of 212 students and 162 teachers participated in the questionnaire survey, and then, more limited members were interviewed to get deeper understanding of the results. Scope of the survey was limited to six-different cheating practices: impersonation, forbidden aids, peeking, peer-collaboration, outside assistance and student-staff collusion, and seven different counter measures: proctors, biometry, mingling, shuffling, random drawing, sequencing and broadcasting. Some of these countermeasures do not only target cheating through exam PC but also traditional types of cheating whispering and use of concealed phone and other equipment. The results show that both students and teachers provided cheating as easier with e-exams, and especially with BYOD device. They also thought that specific countermeasures against cheating would be more effective and easier to implement with e-exams.

Main findings of the paper: In the survey, both students and teachers reported little first-hand experience with cheating in proctored on-campus exams. As for frequency of various types of cheating, both students and teachers considered impersonation, outside assistance and student-staff collusion to be rare while peeking and forbidden aids were considered more common during on-campus exams. A clear majority of respondents assumed less than 5% of the delivered student examination answers utilized cheating with peeking and forbidden aids being the most common cheating approaches. Almost half of the respondents assumed that the likelihood of getting caught if cheating to be 0-20%. Both students and teachers perceived BYOD exams to enable easier cheating than paper exams and university-PC exams. When paper exams and BYOD exams were compared, students assumed BYOD to enable easier cheating with forbidden aids, peeking, peer collaboration, and outside assistance while teachers believe forbidden aids, peeking, peer collaboration, student-staff collusion would be easier. When BYOD exams compared to University PC, teachers perceived BYOD exams as enabling easier cheating for all six cheating threats while students believed BYOD enable easier cheating with only impersonation, forbidden aids, peer collaboration and outside assistance. Comparing paper exams with university PC exams, students perceived paper as enabling easier cheating with impersonation and forbidden aids, while university PC would allow easier peeking. Teachers did not perceive any significant differences either way. Both groups believed that cheating is easier with e-exams especially with BYOD exams, when comparison was performed between students' and teachers' data, teachers perceived BYOD exams as enabling easier cheating than students had. As for countermeasures against cheating, respondents considered proctors to be more effective with paper exams than with e-exams. On the other hand, several other countermeasures were considered easier to implement with e-exams. Both students and teachers thought that forced sequencing of tasks and broadcasting of clarifications would be more

effective for e-exams than for paper exams. Students also considered biometry, shuffling of question order, and random drawing of questions from larger item banks as more effective countermeasures against cheating for e-exams than paper exams.

The paper's relation to the research questions and contributions of the thesis: This paper addresses SQ1 and C1 by investigating student and teacher perceptions on cheating threats and countermeasures for proctored on-site exams. In particular, the paper compares perceptions on ease of various ways of cheating, and effectiveness of select countermeasures, for three types of exams: traditional pen and paper exams, e-exams using university equipment, and BYOD e-exams (i.e., students using their own laptops). The empirical findings from this study supports the analytical results from P1, namely that e-exams have more cheating threats, but also a wider arsenal of effective countermeasures.

4.7. **Paper 7**

Title: E-exams in Norwegian Higher Education: Vendors and managers views on requirements in a digital ecosystem perspective

Authors: Aparna Chirumamilla, Guttorm Sindre

Publication Venue: Computers & Education, 104263, 2021.

CRediT authorship contribution statement:

- Aparna Chirumamilla: Conceptualization, Methodology, Validation. Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.
- Guttorm Sindre: Resources, Validation, Writing – Review & Editing, Supervision.

Abstract: E-assessment has been supported in Learning Management Systems for decades. More recently, dedicated e-exam systems have emerged on the market, more specifically supporting the workflow and security needs surrounding high stakes exams. For instance, in Norway, LMS's Canvas and Blackboard are only used for ungraded assessment tasks, while e-exam systems like WISEflow and Inspera Assessment are used for graded ones. Since the systems are mass-market software, vendors must satisfy the needs of several customers, and needs that are specific to only one or a few customers will receive low priority, perhaps forcing teachers to adapt their assessments to what the tool supports, rather than having the tool adapt to the preferred pedagogy. So far, there has been considerable research on views of students and teachers on e-exam systems, much less on the views of vendors and system managers. In this paper, we investigated what these stakeholder groups consider to be the key features of e-exam systems and by what process they are determined. An exploratory case study

was conducted based on interviews with 12 participants belonging to three groups: vendors, process managers, and system managers in Norwegian universities. Our findings indicate much agreement among these groups about key features of e-exam systems, though observing that not all functionality requested by end-users will be prioritized. Also, there was much agreement that a movement towards standardization, open interfaces, and digital ecosystems would allow smoother integration with other information systems in the higher education sector and easier addition of plug-ins for specific functionality – but that there still is a way to go to reach the ambitions of a flexible ecosystem. Currently, vendors give more priority to adding functional features in e-exam systems rather than better interoperability, and integration with third-party tools remains a challenge.

Main findings of the paper: This paper has reported on an interview study with 12 persons having central roles in the development, procurement, and operation of e-exam systems in Norwegian higher education, some being vendor employees, some system managers (i.e., license administrators, project managers, team leaders, advisors, engineers) at various universities, and one being a process manager at the national organization Unit, coordinating the acquisitions and IT infrastructure development for all Norwegian public universities. Participants generally seemed satisfied with the requirements process, feeling that the coordinating role of Unit had improved the process compared to the previous situation where each university was running separate acquisitions and integration projects – though it was observed that parts of the requirements process were still somewhat ad hoc. There was much agreement between participants about both functional and non-functional features of e-exam systems. System managers were much focused on security and interoperability. As for security, while there might be some vulnerabilities for cheating, especially with the BYOD type of e-exam, stakeholders generally felt that systems were satisfactory and believed that more cheating was taking place outside the digital exam infrastructure – e.g., old-fashioned cheat notes, concealed mobile phones. As for interoperability, stakeholders generally agreed that long-term ambition should be a move towards a digital ecosystem where open standards and APIs would allow for smooth integrations between various tools. However, many integrations were challenging at the current stage – both considering the usage of third-party tools during the exam and exchanging information between e-exam systems and other tools, such as Learning Management Systems. Table 4 shows participants responses on some of the enablers and barriers for achieving open digital ecosystems for e-exams. Since our case study was restricted to Norway, some findings may be specific to the national setting. However, both e-exam system vendors have customers in several countries, in Europe, and beyond. Hence, the challenges surrounding requirements specification, security, and interoperability of e-exam systems are likely of much broader interest. Interesting avenues for further work could be to perform similar studies in other countries or across several countries and compare the views of the stakeholder groups interviewed here with those of students and teachers. It will also be exciting to see how the e-exam systems

of these two vendors – and other competitors – develop over the next couple of years. While the move towards open digital ecosystems is a fine ideal that many agree about, there are also obvious business interests that would go in favour of maintaining a situation dominated by proprietary software.

Table 4. Enablers and barriers for achieving open digital ecosystems for e-exams

Enablers	*Barriers*
• Availability of Open APIs • Use of open standards for development • Increased reusability • Culture	• Lacking openness • Lacking interoperability between components: Lacking content sharing between components, Lack of support for third-party tools access from tool vendors • Lacking priority for interoperability from vendors and customers • Non-uniformity in choosing solutions for integrations

The paper's relation to the research questions and contributions of the thesis: This paper contributes to research questions SQ2 and SQ3, and contributions C1, C3, C4 and C5. It gives an empirically grounded description of the requirements process surrounding the acquisition and development of e-exam systems in Norwegian higher education and challenges regarding functionality, security, and interoperability.

5. Discussion

This chapter will discuss the research contributions, implications of findings, limitations, and evaluation of this PhD research work.

5.1. Contributions

5.1.1. Cheating threats and countermeasures in paper exams vs. e-exams

C1. *Improved understanding of cheating threats and countermeasures in paper exams vs. e-exams and empirical findings on perceptions of teachers, students, vendors, and managers about such threats and countermeasures.* This first contribution is based upon literature review, case study, and data from quantitative and qualitative studies. The comparison of cheating threats and countermeasures in e-exams and paper exams was investigated through threat modelling and risk analysis with ADTrees as presented in P1 (Sindre & Vegendla, 2015), penetration testing on an open-source lockdown browser software using the HARM method in P2 (Vegendla, Søgaard, & Sindre, 2016), and empirical findings on the perceptions of stakeholders (teachers, students, system managers, tool vendors) about cheating risks in paper versus e-exams in P6 (Chirumamilla, Sindre, & Nguyen-Duc, 2020) and P7. Similar to the analysis done in P1, a threat analysis comparing on-site and remote exams was done in P4, especially focussing on strengths and limitations of biometry (Vegendla & Sindre, 2019). All these papers enrich the understanding of cheating threats and countermeasures in paper exams versus e-exams and help to enable more precise specification of security requirements relating to cheating in exams.

The threat analysis in P1 indicated a group of cheating threats that could be done via PC. The penetration testing on ways to beat the lockdown browser in P2 showed that several such cheating threats were viable in practice. These tests add to the knowledge provided by Dawson (2016). Although some cheating approaches are similar to what Dawson proposed, different target systems were tested. Furthermore, the idea that e-exams may be more vulnerable to cheating than paper exams was also supported empirically, as both students and teachers believed this to be the case for several different cheating threats. The combination of threat analysis, testing, and empirical study of stakeholder perceptions strengthens the results compared to having just one of them.

However, the key finding here is not that e-exams have additional cheating threats, which is rather apparent. A perhaps more important contribution of this thesis is pointing out that at the same time, e-exams also enable practical countermeasures against cheating. In some cases, these are countermeasures that would be possible also for paper exams. For instance, mixed seating of candidates from various courses would likely be effective mitigation towards close range collaboration such as peeking and whispering in the exam hall (mixed seating here meaning that a candidate up for a CS1 exam would not be surrounded by other CS1 candidates, but

instead by candidates from, e.g., Physics, Statistics, Psychology, English Literature). Mixed seating is possible both with paper exams and e-exams. However, it complicates the logistics a lot for paper exams, as invigilators cannot go down the aisle with huge piles of one exam set. They instead need to interleave sets from different piles in an intricate pattern. In cases where exams have unequal duration, mixed seating also complicates enforcement of the timing (i.e., prevent students from writing after time is out). For e-exams, there is much less of a logistics problem since question sets need not be distributed on paper – every candidate simply gets the set for the right course based on their authenticated login. Similarly, there is no need for invigilators to enforce time since the e-exam system could automatically auto-save and deliver when the time is out for those students who did not deliver already. Hence, there is a higher chance that mixed seating could be enforced within reasonable cost with e-exams than with paper exams. This illustrates that e-exams enable countermeasures, even some countermeasures that are of a purely physical nature, such as the seating arrangement.

In addition, there are, of course, countermeasures of a more digital nature, which might be entirely impossible for a paper exam. One example would be the detailed logging of the growth of the student's answer, potentially down to every keystroke. Suppose a student near the end of the exam suddenly receives an electronic message with a nice essay from an outside helper (e.g., by somehow being able to circumvent the lock-down browser) and pastes this into his answer field. In that case, the e-exam system might detect such a suspiciously fast growth of the student's answer and flag it as suspicious. On the contrary, consider the paper-based variant of this cheating approach. The cheater has been able to drop off empty answer sheets in an agreed place in the restroom, which the accomplice have then obtained and filled in with an A-grade essay, put back in the same spot to be obtained by the candidate on a second restroom visit, and then smuggled back to the exam hall under his clothes. Here, it is unlikely that any quick growth of the candidate's answer will be observed since each invigilator is typically watching many candidates, not really aware of how many sheets they have written at any point. The claim that e-exams can support additional countermeasures is made analytically in P1 and P4 and supported empirically from the perceptions of students and teachers in P6. It can also be noted that expert stakeholders interviewed in P7, although acknowledging that e-exams were not 100% secure, felt that there was much more cheating taking place outside of the electronic exam infrastructure (e.g., using old-fashioned cheat notes or concealed mobile phones) than what was done inside this infrastructure (e.g., breaking lockdown and using forbidden aids via the exam PC itself). It is also observed by Küppers et al. (2020) that cheating outside the e-exam system, e.g., using unauthorized material via the toilet, is a more common threat even for proctored on-campus e-exams. Similarly, Dawson (2020) noted that even perfect authentication by e-exam systems would not protect closed-book examination from cheating via unauthorized material (e.g., using crib sheet). Thus, control of assessment circumstances is required instead, indicating that the e-exam system may not be the weakest link to cheating in proctored on-campus exams.

For take-home exams, as discussed in P4, the cheating threats would increase compared to on-campus e-exams, but not because they are e-exams, instead because

they are unproctored, or with remote proctoring rather than having a proctor in the same room. Increased ease of cheating for take-home exams was observed by Harmon and Lambrinos (2008). The most common threat for take-home e-exams would be plagiarism – either from web sources or in the form of illegitimate collaboration where candidates share answers. However, the e-exam also provides good chances to discover plagiarism, using an automatic plagiarism checker, which would be much harder for a take-home exam using pen and paper. On the other hand, plagiarism checking does not prevent cheating where the candidate gets help from a third party providing original answers beyond the candidate's competence. The attempted mitigation for this is through biometric authentication and remote proctoring. However, help from a third party would only be detected by such an approach if the helper is visible or audible to the remote proctoring system. If the helper is outside the angle of the web cameras, and avoids speech that is caught by the microphone, biometry will not discover anybody but the correct candidate. Hence, additional measures must be taken together with biometrics. Nevertheless, a take-home e-exam would have fewer mitigations against cheating than an on-campus e-exam, but again, a take-home paper exam (which would be rare nowadays) would have even fewer mitigations against cheating.

Apart from the research context for this contribution of thesis, Covd-19 caused emergency changes in exam practices, driving many courses to remote / take-home (and often unproctored) e-exams which would otherwise have had on-campus e-exams (or even on-campus paper exams in case of e.g., some math courses). This thesis mainly looked at e-exams and cheating for on-campus proctored situations. As much of the research in this thesis was already done and about to wrap up when the lockdown came about. If this thesis project had been done 2 years later, it would likely have ended up focussing more on remote exams. Many of the new features added by Inspera and UNIwise just recently are also about remote exams. At the same time, P4 provided categorization of threats and countermeasures specific to remote exams, and as well as discussed some of the countermeasures (e.g., biometry, variation in question style, and enforcing sequential answering of questions with no back option in the exam) generally to mitigate collaboration between candidates are relevant regardless of whether it is on-campus or take-home.

5.1.2. Issues and research gaps in requirements engineering for software ecosystems

C2. *A review of issues and potential research gaps in requirements engineering for software ecosystems through a systematic mapping review, producing essential findings concerning requirements engineering activities and non-functional requirements for software ecosystems.* This contribution is presented through a systematic mapping review (P3) (Vegendla, Nguyen-Duc, Gao, & Sindre, 2018) of requirements engineering (RE) in software ecosystems (SECOs) between 2009-2017.

The objectives for systematic review in P3 were:

- to present what RE activities have been studied in the literature
- to explain how the non-functional requirements were considered in SECOs.
- to identify whether the RE process used for traditional systems can cope with the context of SECOs.

Table 5. Research topics across RE activities in SECOs

Activity	*Topics*	*Relevant Paper from literature*
Elicitation	• Goal modelling	• (Yu & Deng, 2011)
	• Reference model	• (Pettersson, Svensson, Gil, Andersson, & Milrad, 2010; Van den Berk, Jansen, & Luinenburg, 2010b)
	• Non-functional requirements • Identifying stakeholders' roles	• (Axelsson & Skoglund, 2016; Scacchi & Alspaugh, 2012a)
	• Identifying relationship	• (Angeren, Kabbedijk, Jansen, & Popp, 2011; Fricker, 2009, 2010; Handoyo, Jansen, & Brinkkemper, 2013; Slinger Jansen et al., 2009)
	• Policies	• (Angeren et al., 2011; Slinger Jansen et al., 2009; Schultis, Elsner, & Lohmann, 2014) • (Scacchi & Alspaugh, 2012b)
Analysis	• Requirements communication or Negotiation	• (Fricker, 2009, 2010; Knauss et al., 2014; Valenca, 2013)
	• Conflict management	• (Fricker, 2010; Schultis et al., 2014; Valença et al., 2014)
	• Conflict analysis • Requirements prioritization	• (Christensen, Hansen, Kyng, & Manikas, 2014; Fricker, 2009; Schultis, Elsner, & Lohmann, 2013)
	• Requirements selection	• (Jansen, Finkelstein, & Brinkkemper, 2009; Knauss et al., 2014) • (Valença et al., 2014)
Specification	• Notation semantics	• (Boucharas et al., 2009)
	• Modelling approaches	• (Boucharas et al., 2009; Pettersson et al., 2010; Sadi & Yu, 2017; Santos, 2014; R. P. d Santos & C. M. L. Werner, 2012; Van den Berk et al., 2010b; Yu & Deng, 2011)
Validation	• Model formalism	• (Boucharas et al., 2009)
	• Requirements verification, validation and testing	• (Soltani & Knauss, 2015)
Management	• Global RE	• (R. P. d Santos & C. M. L. Werner, 2012)
	• Management practices	• (Knauss, Yussuf, Blincoe, Damian, & Knauss, 2016)

Our mapping review in P3 indicates that there was not yet a standardized process for RE in SECOs, which is in line with previous claims by Immonen et al. (2016). Notably, there had been little research on the overall process or method, procedures, techniques, and tools for requirements engineering in SECOs, as observed in (R. P. d. Santos & C. M. L. Werner, 2012). On the other hand, there had been somewhat more research looking at specific phases of RE in the development of SECOs, e.g., requirements elicitation -identifying stakeholder roles and relationships or requirements analysis – selection and prioritization, verification as indicated in Table 5.

Another finding from the results shown in Table 5 is that there are several examples of research on the transition between requirements and architecture for software ecosystems, which is also similar to previous findings by Immonen et al. (2016). For instance, particularly on architectural design, discussing different views - architectural, business, social, SECOs engineering and management (Campbell & Ahmed, 2010; R. P. d. Santos & C. M. L. Werner, 2012). Several papers tried to represent architecture in various ways, e.g., as a reference model or framework, as design to show participating components, interfaces, relations between components, APIs, providing overall software configuration layout (Pettersson et al., 2010; Van den Berk et al., 2010b).

Several of the mapped studies indicate that generic modelling approaches from the software engineering field, e.g., goal models, social-network models, and supply chain models, are also useful for requirements elicitation and analysis in SECOs (Fricker, 2009; Jansen & Cusumano, 2013; Jansen, Handoyo, & Alves, 2015; Sadi & Yu, 2017; Yu & Deng, 2011). However, the extent to which these modelling approaches are appropriate is a question of debate, as Jansen, Cusumano, and Popp (2019) find that modeling approaches used in the software engineering field do not scale upward to model SECOs - when many actors are involved, the models tend to become too complex.

Table 6. Research topics across software quality attribute in SECOs

Quality Attribute	*Type*	*Topics*	*Papers*
Reliability	• In-operation	• Reliability concerns	• (Bosch, 2010)
	• In-operation	• Dependencies between unreliable components in CRAN ecosystem	• (Claes, Mens, & Grosjean, 2014)
Maintainability	• In-development	• Dependency problems in CRAN software ecosystem	• (Claes et al., 2014)
Safety	• In-operation	• Architecture for embedded open software ecosystems, Safety of automotive	• (Eklund & Bosch, 2014)

		software	
Security	• In-operation	• Certification and policy management	• (Bezzi, Damiani, Paraboschi, & Plate, 2013)
	• In-operation	• Security patterns	• (Fernandez, Yoshioka, & Washizaki, 2015; Fernandez, Yoshioka, Washizaki, & Syed, 2016)
	• In-operation	• Software redundancy	• (Gherbi, Charpentier, & Couture, 2011)
	• In-operation	• Authentication, accountability, transparency	• (Fahl, Dechand, Perl, Fischer, Smrcek, & Smith, 2014)
Performance	• In-development	• Developer performance	• (Goldbach & Benlian, 2015)
	• In-development	• Performance measurements of eclipse software ecosystem	• (Hansen & Zhang, 2013)
	• In-development	• Ecosystem healthiness: robustness, productivity, interoperability, stakeholder satisfaction and creativity	• (Mhamdia, 2013)
Testability	• In-development	• Software testing requirements for mobile applications	• (Dantas, Marinho, Costa, & Andrade, 2009)
	• In-development	• Static analysis (or Code reviews) of vulnerabilities in web applications and their plugins	• (Walden, Doyle, Lenhof, Murray, & Plunkett, 2010)
	• In-development	• Review and approval methods for platform extensions	• (Jansen & van Capelleveen, 2013)
Transparency	• In-development	• Requirements engineering	• (Leite & Cappelli, 2010)
	• In-operation	• confidentiality	• (Knauss et al., 2016)
Usability	• In-operation	• Software platform usability and extendibility	• (Jansen, 2013)

Table 6 shows papers that discussed topics related to quality attributes in SECOs. Our review identified eight quality attributes that had been focused on in research on RE for SECOs: usability, transparency, testability, performance, security, safety, maintainability, reliability. Among these, most research was performed on security, performance, and testability. Lima et al. (2019) provided a list of 64 quality attributes, including several quality attributes and properties specific to SECOs context, their list contains all attributes addressed in our review. They evaluated attributes comparing to the ISO/IEC 25000 standard and performed a questionnaire survey with experts in industry and academia. Their survey results indicate that respondents felt proposed quality attributes were highly relevant for SECOs context. The results from our mapping review on quality attributes provided domain specific research for some of the attributes listed in previous findings of Lima et al. (2019).

As for the investigation about quality attributes for software ecosystems, our mapping study is closely related to two other mapping studies that looked especially at this (Axelsson & Skoglund, 2016; Lima et al., 2019). Both these studies cover many more quality attributes than our study. The mapping study by Axelsson and Skoglund (2016) looked directly at mapping many publications about quality attributes in software ecosystems, whereas our study primarily mapped publications about requirements engineering in software ecosystems, and then within that body of work found a group of papers addressing quality attributes. Hence, Axelsson and Skoglund (2016) will found papers also covering quality attributes mainly relevant for the design, implementation, or operation of software ecosystems, while we will have identified discussion of quality attributes only to the extent that it took place within papers focussing on requirements engineering. Thus, the findings of these papers complement each other. Axelsson and Skoglund (2016) showed what quality attributes are generally considered relevant for software ecosystems, while our paper shows which of these quality attributes are receiving much attention in published research within the field of requirements engineering.

P3 has clear limitations in being just a systematic mapping study, rather than a systematic literature review. The latter could have provided even more detailed insights, with more quality evaluation and synthesis of the published works in the area but was considered too demanding. In spite of the mentioned limitations, P3 has identified clusters and gaps in the research about requirements engineering for software ecosystems.

5.1.3. Requirements process for e-exam systems

C3. *Empirically grounded descriptions of the requirements process surrounding acquisition and development of e-exam systems in Norwegian higher education* This contribution extends P3 with the empirical study presented in P7, which provides qualitative interviews with vendors of e-exam systems and with managers of such systems in Norwegian higher education institutions (HEI) concerning requirements for e-exam systems in the digital ecosystems perspective.

Our study in P7 indicates that the requirements process for e-exam systems acquisition in Norway is still somewhat ad hoc. This is in line with previous findings

by Foss-Pedersen and Begnum (2017), though they indicated an even more ad hoc process - the division of responsibility among various actors seems to have become clearer when we studied the process in 2019/2020 than it was a couple of years earlier. Vendors stated in the interviews that they could have given more emphasis to interoperability if customers had prioritized that, but the vendors had the impression that other issues, e.g., functional features and security against cheating, had higher priority. This resembles findings of Foss-Pedersen and Begnum (2017) with respect to universal access that vendors responded that they could have been able to improve the software more in that respect if that had been given a high priority from customers.

Especially, the arrangement to have one national organization – Unit – oversee the requirements and integration process seemed to be satisfactory to the participants. No participant indicated a wish to go back to the previous situation when each institution would run separate procurement and integration projects. Unit maintains a requirements specification for e-exam systems for Norwegian HE institutions, negotiates with vendors on universities' behalf, and takes responsibility for developing integration software to make the e-exam systems fit into the common system infrastructure. A similar approach is used for Sweden, as mentioned by vendors and process managers. We also found evidence in literature of such an approach in the higher education sector in the Netherlands (Boezerooy, Cordewener, & Liebrand, 2007). An interesting question is whether a similar collaboration with joint acquisition could be made to work across universities in several countries. As indicated by stakeholders, some requirements will differ from country to country such as those relating to the national IT infrastructure and those relating to laws, regulations, and student rights concerning exams and grade appeals. On the other hand, many requirements will be common across universities, such as mitigating typical cheating threats, and many requirements will vary between disciplines rather than between countries. For instance, math exams may have a joint need irrespective of country of an easier way for candidates to write equations during e-exams, which is currently cumbersome with the keyboard and requires scanning if using paper. Programming, likewise, may need support for question genres checking code against test suites, or with smooth support for Parsons problems or other question types that are popular in introductory programming courses, as discussed in P5.

Vendors in our study indicated that difference in the requirements from country to country, and even between universities in the same country, would cause challenges with balancing the needs of various customers. This resembles the challenges with coordination of a large number of stakeholders discussed in the study by Bosch and Bosch-Sijtsema (2010), and they recommended an architecture-centric approach as a coordination mechanism, claiming that a process-centric approach (face-to-face communication) may not effectively manage large-scale software development. In our interview material, we could not find any clearly expressed stakeholder preference with respect to architecture-centric or process-centric coordination. On the other hand, it seemed from our findings that the approach actually used by universities appeared more architecture-centric than process-centric. For instance, in our study, universities in the Norwegian higher education sector ended up choosing different solutions for integrations between e-assessment system and their local

systems even though universities have similar requirements, causing e-assessment system development delay. However, this was a deliberate choice by universities, not an effect of poor coordination and an ad hoc RE process.

5.1.4. Key features for e-exam software

C4: *Description based on empirical evidence of key features for e-exam software according to vendors, process managers, and higher education institutions in Norway.* This contribution discusses results (P5 and P7) concerning key features for e-exam software widely used in Norway through a case study with vendors, process managers, and system managers from the Norwegian HE sector.

During interviews (P7), participants discussed what they considered significant functional features of e-exam systems, such as authoring, logistics support, question analytics, grading, explanation of grades, appeals and complaint grading. They also gave their opinions about various quality attributes such as scalability, usability, integration and interoperability, security, and reliability. Our findings indicate that stakeholders spent more time discussing recently added and perhaps challenging features than more straightforward features. For instance, some obvious functional features of an e-exam system would be the ability to present questions to candidates during the exam, and to receive and store the answers. These are features no sensible e-exam system could do without. Still, respondents hardly talked about them, which could be plausibly explained by their obviousness. It is a well-known phenomenon in the field of requirements engineering, as for instance observed by Firesmith (2003), that assumedly obvious features will tend to be omitted by interviewees, especially if they are experts on the technology in question.

Of the features that participants did talk about, there was comparatively more mention of question authoring, explanation of grades, and appeals grading, less about the logistic system, question analytics and ordinary grading. Of course, ordinary grading (for all students) is a much more used functionality than appeals grading (for the smaller fraction of students who dispute their grades). When interview participants talked more about appeals grading, this again is likely due to the tendency to omit the obvious. Moreover, appeals grading had probably gained extra attention from many participants in our study because it was initially unsupported by the e-exam systems used in Norway, addressed by a cumbersome manual workaround that both teachers and students had been dissatisfied with. Hence, it had been included as a feature of the e-exam system just recently before the interviews. An over-focus on complaints grading by respondents compared to ordinary grading could be a kind of recency bias (respondents focusing more on the problems they experienced most recently), this is in line with findings by Pitts and Browne (2007) suggesting that stakeholders in requirements elicitation tend to over-focus on their most recent information system usage or development experiences.

Our results on key features are in line with previous findings by Kuikka et al. (2014) and Striewe (2019). The features highlighted by our participants relate to all four major components proposed in the framework by Striewe (2019): front-end components (user interface), educational components (underlying logic of tests and

question types, pedagogical modules), knowledge representation and storing components (e.g., question pools, tests, exam answers and results), and connector components (related to integration and interoperability). Admittedly, the development of features for knowledge representation is currently much less advanced than suggested by Striewe. Kuikka et al. noted that none of their compared systems satisfied all the needs of teachers. Likewise, in our study, participants indicated that features with less backing from customers would end up not being prioritized. Also, teachers would sometimes have to make trade-offs, such as only being allowed to have open book exams if an e-exam were to include usage of third-party tools. This was not a limitation decided by the vendors, but by system managers at the university, due to concern that usage or third-party tools would open up a potential vulnerability allowing candidates to access forbidden information during the exam. Hence, demanding that e-exams including third party tools must be open book, this risk of information access would be short-circuited.

The participants in our study felt that authoring and interoperability were important features, this is similar to the findings from an expert survey by Isaias et al. (2019) that found authoring, interoperability and feedback features to be important, but participants in our study talked less about feedback. Different from the situation in Norway, Kuikka et al. suggested that it is an advantage to have the same software product work both as LMS and e-exam system. This may be true, but some of our system manager participants believed otherwise, thinking that too many features in the same system would confuse some teachers.

Our findings show that stakeholders perceived security as the most important non-functional feature for e-exam system, to restrict students accessing sites and other materials outside the exam system. The main countermeasures against cheating built into the e-exam systems are the lockdown browser, log tests, and monitoring of students' devices. This is in line with the previous findings by Hillier and Fluck (2013), who found similar countermeasures against cheating in other e-exam software that they were investigating. There were relatively fewer findings for scalability and reliability in our study, this again is likely due to the tendency to omit the obvious features for e-exam systems. However, vendors perceived scalability and reliability as very important features for their customers, as the BYOD exams were mainly run on the internet. Fluck, Pálsson, et al. (2017) similarly found that scalability and reliability were important qualities for e-exam systems.

5.1.5. Enablers and barriers for achieving open digital ecosystems for e-exams

C5. *Identification of enablers and barriers for achieving open digital ecosystems for e-exams within a larger ecosystem of e-learning.* This contribution explores the concepts and issues discussed in the systematic mapping study (P3) within a larger ecosystem of e-learning, primarily focusing on enablers and barriers for achieving open digital ecosystems for e-exams as presented in P5 and P7.

Our findings indicate much support for the idea of digital ecosystems, both in the literature about e-learning and e-assessment systems (as reviewed, e.g., in related work for P5, and in related work for the thesis as a whole), and among the stakeholders interviewed in the study reported in P7. Both vendors, process and system managers seemed to agree that organizing tools and resources for e-assessment in a digital ecosystem would be the way to go for the future. However, our findings show that the implementation of open digital ecosystems for e-exams still seems to be at an immature stage. Based on the findings from P5 and P7 in this thesis, in this section, we propose enablers and barriers for open digital ecosystems for e-exams.

Enablers for achieving open digital ecosystems for e-exams:

Enablers for achieving open digital ecosystems for e-exams from our findings in P7 are categorized as below:

- Availability of Open APIs
- Use of open standards for development
- Increased reusability
- Culture

Availability of open APIs. Stakeholders felt that the e-assessment platform's APIs from vendors would help universities to customize the platform according to their needs, e.g., making a new question type. Both the e-exam systems we studied have APIs but there is little published research about their usage, an exception being Fitzharris and Kent (2020) who did a case study on one of the two systems we studied (i.e., WISEflow), however their study was about analytics, not question types. These are two rather different use cases for an API. Data analytics (typically of exam results) is a backstage activity performed after the exam, so it does not need a nice robust UI and still could be enough to be able to extract data as a file and then analyse in another tool. New question types on the other hand would be used live by students during exams, thus need much more robustness, better UI, and careful testing. Hence, the successful usage of API to support data analytics does not guarantee that it also gives good possibilities for e.g., universities themselves to make new question types.

Use of open standards for development. System managers felt that the open standards (e.g., IMS PCI) used by Inspera Assessment vendors would facilitate the niche development, e.g., to develop new question types, to the e-exam system platforms from universities. This complements previous findings (Pettersson, 2009; Uden et al., 2007) that investigated the importance of standardisation for e-learning systems and specifically about whether e-learning systems incorporate the notion of ecosystems in their development.

Increased reusability. One of the key advantages mentioned by stakeholders in our study was that using e-learning/e-assessment systems than monolithic system would increase reusability of contents (e.g., questions) and improve the quality of contents. This is in line with findings by (Pettersson, 2009) who initiated principles for e-learning ecosystems more than a decade ago, discussing the notion of reusability, claiming that reusability of contents would increase with the use of e-learning ecosystems.

Culture. Our empirical findings in P7 show that stakeholders thought the culture of the universities would be one of the main factors that would encourage collaboration between universities for joint development of exam questions through e-learning ecosystems. However, our findings have not indicated what was the issue with joint development of exam questions as discussed by Sanders, Ahmadzadeh, Clear, Edwards, Goldweber, Johnson et al. (2013) that a problem with the question bank, perhaps, is that the format of the information is somewhat old-fashioned, ordinary text on a web page, so the reuse approach would be sort of copy-paste and somewhat time consuming to get into an e-exam systems.

Barriers for achieving open digital ecosystems for e-exams:

Barriers for achieving open digital ecosystems for e-exams from our findings in P7 are categorized as below:

- Lacking openness
- Lacking interoperability between components
- Lacking priority for interoperability from vendors and customers
- Non-uniformity in choosing solutions for integrations

Lacking openness: Respondents in our study did not explicitly mention lack of openness in accessing APIs. However, our results indicate trade-offs between openness and security. Vendors mentioned that lack of centralised control on their APIs could enable the users of their systems to pull the data that they are not allowed to use. Thus, the access control on their APIs was strictly monitored. Still, system managers thought that access control would not be a problem for them as they will conduct only verification testing at universities while integration would be primarily performed at the vendors' site.

Lacking interoperability between components: In our study, the interoperability challenge has been mentioned by participants in various ways. The major challenge they brought up was concerning *lack of content sharing between components.* System managers in our study felt that the lack of compatibility between and across e-learning/e-assessment systems for sharing questions would be a burden for teachers in creating exams. Hence, they suggest that more sharing within and between universities would be possible with the integration of question banks. Similarly, Laine, Sipilä, Anderson, and Sydänheimo (2016) and Chituc et al. (2019) also suggest that sharing is possible with the integration of question bank. However, interoperability and IPR issues remain potential obstacles for developing shared item banks (JISC, 2007).

Another issue reported was related to *lack of support in the e-exam systems for using third-party tools during exams.* Stakeholders asserted that both Inspera Assessment and WISEflow support minimal external software during BYOD exams, requiring whitelisting or integration. The support for third-party access was still in the pilot stage. Thus, access to third-party tools would have been feasible only at the cost of reduced security from the lock-down browsers. Hence, when e-exams require third-party tools, e.g., Microsoft Excel or Matlab, universities demand that exams must be open-book, as students would too easily cheat by accessing content on the PC. The limitations surrounding usage of third-party tools in e-exams has also been observed

by Fluck, Pálsson, et al. (2017), where they discussed the lack of third-party access in one of the e-assessment supporting tools (i.e., SEB) used by participants in our case study.

Lacking priority for interoperability from vendors and customers: While agreeing in principle that digital ecosystems were a good idea, several respondents did not feel that it was the highest priority for the moment to achieve the interoperability needed to support an open ecosystem architecture. Moreover, some system managers were also worried that seamless integration of several systems, e.g., e-exam system and LMS, could confuse end-users by the increased number of functional features available. The skepticism of some system managers in this respect may however not be supported by empirical data about teachers' own preferences, as Kuikka et al. (2014), where teachers instead preferred using the same system for e-exams and LMS. The potential differences that may have led to different preferences was that maybe the system managers were thinking in particular about teachers who may not be so technically skilled (e.g., teachers in humanities and social sciences, not just teachers in tech subjects), while Kuikka et al. surveyed a different population of teachers (environment technologies, health care, life sciences, ICT and business-related subjects).

There was functional level integration between a few systems, e.g., the Inspera Assessment exam system and the SEB lockdown browser. Although there were no issues explicitly mentioned with this particular integration in our interview study, vendors felt this specific integration specification was not sufficiently based on standards. The in-house team of vendor organizations mainly performed integrations. However, system managers did not raise any issue with this during our study.

Non-uniformity in choosing solutions for integrations: System managers in our study felt that universities had chosen different solutions for integrations that had caused integrations and implementation delays.

Given that it is more than a decade ago that digital ecosystems were claimed to be the approach of the future within e-learning, the situation in Norway (and elsewhere, e.g., Chituc and Rittberger (2019)) indicates that the development in this direction has so far been slow, it seems at least for the time being the barriers are stronger than the enablers.

5.2. Implications for Research

This thesis is primarily targeted at researchers, educators (i.e., teachers), and technology developers in the field of higher education. In the previous section, we provided our contributions in relation to existing literature. In this section, we discuss implications for future research.

5.2.1. Threat analysis for exams

Threat modeling has been used in previous research, in general, related to software security (Scandariato et al., 2015), but not so much for systems concerning e-learning and e-assessments. We contributed to the body of knowledge on cheating threats towards exams through threat analysis of *paper exams vs. e-exams (C1)*. In previous research, Dawson (2016) asserted that BYOD e-exams are less secure than both pen-and-paper exams and examinations held in the computer laboratory, whereas Küppers et al. (2020) claim that e-exams may be more secure than paper exams. Our analysis shows that both Dawson and Küppers may be correct to some extent. Alike Dawson, we found that BYOD e-exams have additional threats for cheating via the PC - while, as Dawson argues, still having very much the same threats as pen and paper exams when it comes to old-fashioned cheating like peeking, whispering, cheat notes. On the other hand, alike Küppers, our analysis indicates that at the same time, e-exams enable new countermeasures which are impossible or infeasible for paper exams. Our studies indicate that it is hard to make general statements that one will be more secure than the other – it depends on the security of the e-exam system and surrounding phyiscal and organizational set-up, the type of examination and questions, etc.

The key conclusion from this thesis added to the existing research is that threat modeling could enable the analysis of practical countermeasures against cheating. In addition, we tested the validity of threat models in representing the attacks by performing penetration testing of e-assessment supporting tools. Further, we added empirical evidence for the validity of threats and countermeasures identified in our previous analysis through a survey with students and teachers in Norway. Our work has proposed initial threat analysis for e-exam technology. Researchers and technology developers in e-learning could borrow ideas from our work to pursue more detailed threat analysis.

5.2.2. Requirements engineering for e-learning ecosystems

So far, there has been little published research specifically on requirements engineering in e-learning ecosystems. Due to the limited amount of research in RE specifically for e-learning ecosystems, we chose to pursue a systematic mapping review with the broader scope of RE for software ecosystems in general – to see what processes and techniques are suggested, and to what extent this could inform our further studies of requirements processes for e-exam systems.

The research on software ecosystems is generally progressing, but more about software development, especially on how developers communicate. But relatively little research had been done on the requirements engineering perspective at the time of our mapping study. The systematic mapping review resulted in a review of issues and potential research gaps in requirements engineering, thus adding to the body of knowledge concerning the understanding of the state-of-the-art in requirements engineering activities and non-functional requirements for software ecosystems (C2). We further extended this work by performing a case study with experts within

vendor companies and universities to provide empirically grounded descriptions of the requirements process surrounding the acquisition and development of e-exam systems in Norwegian higher education (C3). Previously published work about the requirements process for e-exams in Norway by Foss-Pedersen and Begnum (2017) was looking specifically at universal access requirements in 2017, describing a situation with less synchronization of requirements of various universities. Our study supports the findings by Foss-Pedersen and Begnum (2017) that the requirements process is somewhat ad hoc, though our study conducted in 2019/2020 showed that the coordination between stakeholders had been improved during the period between their study and ours, especially with a much clearer role assigned to the public acquisition organization Unit. Also, there is little empirical work published about such requirements processes internationally. This work provides researchers, educators, and technology developers with the improved understanding of a requirements engineering process for e-learning ecosystems concerning joint software acquisition by several universities on a national level. Further our results could help tool developers and universities in better decision making for using requirements process during new e-exam systems acquisitions or transitioning to new e-exam systems.

5.2.3.Requirements and key features for e-exam systems

There has been a lot of published research on key requirements and features for e-exam systems. However, requirements would always differ from country to country, such as those relating to the national IT infrastructure and those relating to laws, regulations, and student rights concerning exams and grade appeals, and even some requirements would be specific for universities in country. So far, much of the research has been design-oriented, researchers proposing their own prototypes or architectures for e-exam systems, there has been less empirical research on mass-market e-exam systems from commercial vendors. This thesis has contributed to the scientific knowledge about requirements and features for e-exam systems, based on empirically investigated viewpoints of vendors, process managers and system managers in higher education institutions in Norway (C4). The previous literature study by (Striewe, 2019) proposed four essential components for e-assessment systems together with the features that the components would provide. This thesis complemented findings by Striewe (2019), by providing empirical results relevant for four components and features suggested by Striewe. Our study also complements previous survey findings by Kuikka et al. (2014) on teachers perceptions about features for e-exam systems through a qualitative study. However, admittedly, our study did not directly indicate teachers perceptions of the features. We heard about teachers experiences with e-exam systems through system managers.

5.2.4. Digital ecosystems for e-exams

The research on digital ecosystems has received little attention in the e-exams domain though this is a growing trend for learning management systems in e-learning. Although the principles of ecosystems have been studied for e-learning

more than a decade ago (Pettersson, 2009; Uden et al., 2007), much of the work has been on the idea and architecture level, and there has so far been little empirical research on digital ecosystems for e-exam systems, and how digital ecosystems might impact security. We added two contributions to the existing research by (Pettersson, 2009) and (Uden et al., 2007). The first contribution is the empirical investigation in P7 about the views of key stakeholder groups (vendors, managers) on digital ecosystems as an ideal solution for e-exam systems – as well as the degree of practical progress towards this ideal – where it was found that there was great agreement about the ideal, but the practical progress was limited and not highly prioritized. Second, from our empirical investigation, we presented enablers and barriers for achieving open digital ecosystems for e-exams within a larger ecosystem of e-learning (C5). This work has added to the body of knowledge on digital ecosystems for e-exams in terms of providing an empirically based understanding of the discrepancy between ideal goals and practical progress.

5.3. Evaluation of Validity Threats

Credibility of findings is of utmost importance for empirical research. The research reported here focuses on evidence from case studies, surveys, and literature reviews. We will structure this discussion of threats to validity according to categories of such threats often found in the literature, namely internal validity, external validity, reliability, construct validity, conclusion validity. Figure 7 shows the hierarchical tree of validity threats observed in this PhD research project.

5.3.1. Internal Validity

Internal validity is determined by how well a study can systematically rule out alternative explanations for its findings (Fraenkel et al., 2011). Also, it refers to the extent to which evidence supports a claim about cause and effect within the context of a particular study so that the findings can be trusted. For this thesis work, the three most relevant categories of internal validity threats are as follows:

- Subject characteristics, i.e., validity threats related to human informants in the empirical studies. Two subcategories of such threats are especially relevant to us, selection bias and participants bias.
- Location, i.e., validity threats related to the physical or virtual location where data were collected.
- Instrumentation, i.e., validity threats related to the instruments used for collecting and analyzing data. Four subcategories of such threats are especially relevant to us, researcher bias, limited respondent knowledge, sloppy responding in surveys, varied question scaling in surveys

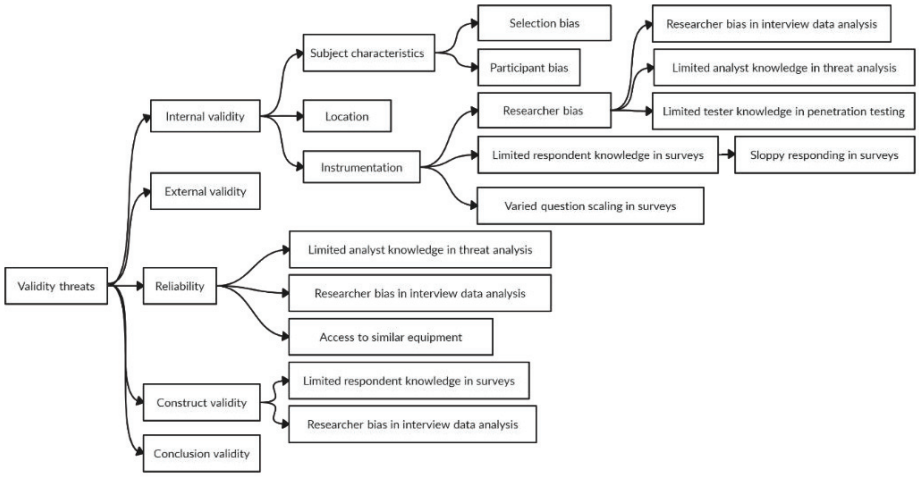

Figure 7. Hierarchical tree of validity threats observed in this PhD research

5.3.1.1. Subject Characteristics

Selection bias: The selection of students and teachers for surveys in P6 and the selection of interview participants in P7 may have affected the results. For instance, in P6, engineering and technology students were heavily represented. These may be more technically skilled than the average user when it comes to IT, hence they may be more positive towards e-exams than the average student. We mitigated this threat by approaching subjects from various departments.

In P7, there were more respondents from NTNU than from other universities, and there were more system managers than vendors or process managers. This situation emerged because there are many universities in Norway, but only two vendors of e-exam systems in use (Inspera, UNIwise), and only one organization taking the coordinating role (Unit). Moreover, people in the companies were very busy so could not spend a lot of time on interviews. But, vendors and Unit provided the persons whom they thought would be most able to discuss the topics at hand, i.e., those centrally placed in the requirements process which was the target of investigation. So, we could be able to interview most relevant candidates for our study, including system managers from different universities rather than only from NTNU. Through this, we have mitigated selection bias. However, adding more participants just to add more participants (but then getting some who were less knowledgeable about the topic) might not have given much added value to the study.

Participants bias: May the participants knowingly or unknowingly have given inaccurate information during interviews (P7)? Some very obvious features of e-exam systems were hardly mentioned in P7 – omitting information that was taken

for granted is a well-known phenomenon. There may also be other reasons to reply inaccurately, such as memory, embarrassment (if something went wrong with the system or project), or secrecy (e.g., provider representatives not wanting to reveal business secrets). Again, interviewing several persons will reduce this threat. Moreover, we have explicitly reported cases where participants were reluctant to answer about cheating vulnerabilities and concrete ways to utilize them.

5.3.1.2. Location

Among the student survey participants (P6), some participants (n = 84) filled the survey by direct administration in the classroom. To avoid the fear that respondents admitting to cheating might be identified in the answers, we have chosen an appropriate questionnaire and did not collect demographic information such as age, gender, and nationality. We removed incomplete questionnaire responses from the data analysis process, but it did not affect the study results. Still, the location of filling the survey might have affected answering the questions.

Table 7 shows the opinions of students directly administered vs web on ease of cheating in paper exams and BYOD exams. A value smaller than 3 in the column 'Mean' would indicate the first examination type (e.g., paper exams in Table 7) enables easier cheating, whereas a value larger than three would indicate the second type (e.g., BYOD in Table 7) enables easier cheating. We also conducted Mann Whiteney U-test to compare the mean ranks between two groups. To be statistically significant in the difference (i.e., $p < .05$), z value should either be less than -1.96 or greater than 1.96. There was a significant difference in responses of students directly administered and students answered surveys via web for cheating using impersonation, forbidden aids, peer collaboration and outside assistance. Of which, students who answered the survey via web perceived BYOD exams are easier to cheat than paper using impersonation, forbidden aids, peer collaboration, and outside assistance namely with highest mean ranks.

Table 7. Opinions on ease of cheating in paper exams and BYOD e-exams

Type of Cheating threat	Students-direct administered (n=84)		Students-Web (n=64)		Students-direct administered (A) and web (B) (Mann Whitney U-test)			
	Mean	SD	Mean	SD	A Mean Rank	B Mean Rank	Z	p-value (Sig.)
Impersonation	2.87	.555	3.03	.397	70.39	79.89	-2.048	.041*
Forbidden aids	3.08	1.132	3.48	1.069	68.58	82.27	-1.995	.046*
Peeking	3.45	1.057	3.42	1.005	75.00	73.84	-.172	.864
Peer collaboration	3.23	.683	3.52	.836	68.73	82.08	-2.115	.034*
Outside assistance	3.33	.646	3.73	.877	66.79	84.63	-2.875	.004**
Student-staff collusion	2.95	.463	2.92	.482	75.27	73.48	-.415	.678

*($p < 0.05$), ** ($p < 0.01$), and ***($p < 0.001$).

Table 8. Opinions on effectiveness of countermeasures for paper exams and e-exams

Type of Cheating threat	Students-direct administered			Students-Web			Students-direct administered (A) and web (B) (Mann Whitney U-test)			
	N	Mean	SD	N	Mean	SD	A Mean Rank	B Mean Rank	Z	p-value (Sig.)
Proctors	84	2,79	,517	57	2,77	,627	71.20	70.70	-.097	.923
Biometry	84	3,08	,542	56	3,13	,662	70.23	70.91	-.138	.890
Mingling	84	2,99	,768	57	3,07	,704	69.83	72.72	-.503	.615
Shuffling	84	3,11	,695	57	3,16	,774	70.93	71.11	-.029	.977
Random drawing	84	3,11	,621	57	3,30	,680	68.49	74.70	-.1.184	.236
Sequencing	84	3,40	,696	57	3,49	,710	69.66	72.97	-.547	.584
Broadcasting	83	3,13	,488	56	3,12	,605	69.68	70.47	-.173	.863

*($p < 0.05$), ** ($p < 0.01$), and ***($p < 0.001$).

Table 8 shows opinions on the effectiveness of countermeasures against cheating threats mentioned above in Table 7. Results indicate that both groups felt that except proctors, all other countermeasures would be easier to implement in e-exams. However, there were no significant differences reported between the groups.

5.3.1.3. Instrumentation

Researcher bias: Researchers may tend to interpret interview data in ways that confirm their preconceived ideas. Various measure taken to mitigate this threat during interviews includes avoiding leading questions, not pushing participants in any particular direction, following a well-defined protocol for analysing the data (P6 and P7). Participant checking, method triangulation, and investigator triangulation of analyses were used during analysis, which are the best practice guidelines for implementing and reporting qualitative research (Twining et al., 2017). Transcriptions were sent to participants before analysis to verify whether they indicate what participants intended to say. After analysis, a draft of article P7 was sent to interview participants for comment before journal submission to let them point out any cases where their statements may have been misinterpreted. Their suggestions have been accommodated in the article. Method triangulation was used for P6, where we collected data using both quantitative surveys and qualitative interviews. For investigator triangulation of analysis, interview data were analysed together with co-authors (P6 and P7).

Researcher bias can happen during threat analysis since it is a subjective method based on the abilities of the analysts to imagine relevant threats. As such, it is of course vulnerable to any bias by the researchers who conducted the threat analysis. For instance, researchers might have had a favourable attitude towards e-exams versus paper exams from the outset, thus tended to exaggerate the problems with paper exams and underestimate problems with e-exams. Such threats cannot entirely

be mitigated, but we have tried in the papers using threat analysis (P1, P2, P5) to reduce the possible impact of such threats by arguing as clearly as possible for the inclusion of various threats in the model, and for the comparisons made between e-exams and paper exams. In addition, limited analyst knowledge may have affected the threat analysis, and limited tester knowledge may have affected the penetration tests. Here, investigator triangulation of analyses was used to mitigate these threats.

Limited respondent knowledge. Many questions (P6) in the survey were such that respondents were unlikely to know the precise answers and had to guess (e.g., percentage of delivered exams which have used cheating). Moreover, at the university where the questionnaire study was performed (NTNU), the teachers are not directly involved in cheating prevention in the exam venue, which is done by administrative employees and part-time invigilators hired short-term for the exam period. Thus, the teacher's involvement in cheating mitigation would be through designing tasks to make cheating more difficult, or during the grading process if particular answers contain evidence of cheating. Teachers do, however, visit the venue during the exam to respond to clarification issues or corrections to exam questions, so they will be familiar with the typical seating arrangements and density of invigilators per student, which was relevant knowledge for a question in the teacher survey. It must be acknowledged that this may also threaten validity due to limited respondent knowledge. Especially if some of the responding teachers were recently hired at the time of data collection, e.g., from abroad where there might be different standards concerning exam proctoring, they would not have a clear idea about the typical density of invigilators per student.

Sloppy responding in surveys. A threat related to limited respondent knowledge in the survey would be sloppy responding, i.e., respondents just answering questions quickly to get it done, without reading the text carefully enough. Especially in combination with some variation in question design (e.g., some questions having 3 as the neutral mid-point, while others were on a low to high scale), this could have led to unreliable responses, for instance, if the respondent answered a question according to a wrong assumption about its content or scaling. The typical way to mitigate sloppy responses is to have many questions for the same variables, to be able to check whether respondents have answered consistently. This will, however, cause questionnaires to be much longer. Thus, we chose not to do this, as it might dramatically have reduced the number of respondents.

Varied question scaling in surveys. The scaling of questions may have caused our results to miss nuances. In the survey for P6, Q3 (likelihood of getting caught cheating) was designed with a uniform 5-step scale divided at 20-40-60-80%-100%, whereas Q2 (percentage of delivered exams that used cheating) had another division: 1-5-10-20%. In the questionnaire survey, it was good that Q2 did not choose the same uniform scale as Q3 since the uniform scale would have placed almost all responses at 1, (cf. Figure 4 in P6). Indeed, it might have been better if Q3 had been more similar to Q2 since its current uniform scale yielded results strongly skewed towards the low end, with hardly anybody choosing alternatives 3-5 – thus losing granularity.

5.3.2. External Validity

External validity refers to the extent that the results of a study can be generalized to the overall population and other settings (Wohlin, Runeson, Höst, Ohlsson, Regnell, & Wesslén, 2012).

One notable limitation is that some of the research focused on the situation in only one university (NTNU) in Norway. The students and teachers surveyed (SQ1) were from the STEM study programmes at the researcher's own university NTNU. Hence, findings may not be representative and generalizable to other countries, universities, or disciplines. Yet, there is no specific reason to assume that Norwegian students are more or less honest than other students. Higher education is increasingly global, so the findings should have relevance for research related to cheating in other countries, too.

Case studies were conducted (SQ2 and SQ3) at Norwegian universities and vendor companies. The two vendor companies have customers in several countries, and one of the companies (WISEflow) is Danish, so the vendors dealing with required features for their products will have had a somewhat more international perspective, also exemplified by specific statements from participants in P7 that requirements would be different from country to country. Still, the context of the case is specifically the situation in Norway, and a study including universities from other countries might have come up with different findings. The two companies involved in the study were both vendors of dedicated e-exam software, hence catering to universities who use different products for high stakes e-exams than what they use for e-learning in general. Many universities around the world may be using the same system (e.g., Canvas, Blackboard, Moodle) both for e-learning and for high stakes tests, which may lead to differences in expectations towards the products. So, further work is needed to take a more international approach and to get findings covering a broader spectrum of educational software products. Nevertheless, challenges such as security and interoperability are key to e-exams in many countries – as indicated by related work – so findings are believed to be of interest also outside the specific Norwegian context.

5.3.3. Reliability

Reliability refers to the consistency and repeatability of the findings with procedures and instruments used in similar settings (Creswell, 2013). Reliability and validity are bound together. If the process and instrument are reliable, then the results of the study would be consistent and valid.

Threats to this aspect of validity are in survey and interview questionnaire, either if the questions themselves are unclear or if it is unclear how to code the collected data. To minimize errors in questionnaire instruments, the consistency was checked with colleagues of the researcher before the distribution of questionnaires to participants. This helped us to ensure understandability and estimate the time needed to respond, which we wanted to keep within reasonable limits to have a chance of getting enough answers. To overcome researcher bias, data analysis was performed together with

one more researcher (investigator triangulation). For the interview study reported in P7, researchers were also in prolonged engagement with participants (participant checking), to ensure whether the analysis was consistent with collected data rather than researchers' own imaginations. Also, the analysis process has been constantly verified with the study instrument and research approach. So, the research approach and instrument (i.e., interview questionnaire) can be used in a similar group of subjects and research settings.

Threats to reliability would also happen in threat analysis since threat analysis is a subjective process depending on the experience and imagination of the participating analysts. There is no guarantee that other persons doing the same kind of threat analysis would arrive at the same results. It might easily happen that the new analysts ignore some threats identified by previous analysts or find new threats that were not identified by previous analysts. However, we have tried to mitigate the subjective aspects by undertaking the threat analysis in a structured manner and checking with literature on cheating threats to see if there were threats not covered by the analysis.

For the penetration tests, full repeatability would require that the new researchers have access to similar equipment, with the exact same versions of software as was used by us. Especially important would be to have the same version of Safe Exam Browser (i.e., SEB 2.0), Windows 10 OS, and USB rubber ducky to perform key injection. Historical versions of SEB are available from their web page, so in principle, it should be possible for other researchers to repeat the same penetration tests. However, admittedly, the penetration tests were done in 2016, so it would likely be more of interest to new researchers to do penetration testing on recent versions of the product, where many of the vulnerabilities that we found could have been fixed.

The counterpart to reliability is conclusion validity, see further Section. 5.3.5.

5.3.4. Construct Validity

Construct validity concerns the relationship between theory and observation (Wohlin et al., 2012). This work has mainly aimed for exploratory and descriptive, rather than theory testing, so the construct validity is not that relevant to discuss. However, construct validity also concerns using the right tools and metrics for gathering the data, e.g., to what extent the research methods measure what the researcher intended to measure.

In the quantitative study, some surveyed questions were hypothetical, thus hard for respondents to answer accurately. For instance, concerning the effectiveness of countermeasures, if a student has not sat any exam where a certain countermeasure was used (e.g., mixed seating of candidates), and a teacher has never given such an exam, answers would reflect qualified guesses by the respondents rather than experiences. However, the paper only claims findings of respondents' beliefs about the amount of cheating, ease of cheating, and effectiveness of countermeasures, not about the real amount and ease of cheating, nor the real effectiveness of countermeasures.

In qualitative studies, a threat to construct validity occurs when the constructs (or questions) discussed in the interviews are not interpreted in the same way by the data collector. This threat was mitigated with investigator triangulation and participant checking (cf. Section 5.3.3).

5.3.5. Conclusion Validity

Conclusion validity concerns with the issues that affect the ability to draw the correct conclusion about relations between treatment and the outcome of an experiment or the independent and dependent variables (Wohlin et al., 2012). Conclusion validity is sometimes referred to as statistical conclusion validity and concerns, e.g., choice of statistical tests, care taken in the implementation, and measurementation of the experiment (Wohlin et al., 2012).

Our data collection for some of the questions in surveys in P6 was done based on five-point Likert scale, which consists of ordinal data (See questionnaire from, https://www.dropbox.com/sh/8h0wgzusbx1vd9e/AACnkL47EHZR4f2IgxYGFecLa?dl=0). To test for the significant difference between paper exams and e-exams, we used one-sample t-tests (with the neutral alternative 3 as test value) on students and teachers samples and independent t-tests between both groups. Generally, parametric t-tests assume that the data is usually normally distributed, thus may not be appropriate for ordinal data that are not normally distributed. There exists disagreement amongst scholars about whether Likert data should be analyzed with parametric, e.g., t-tests, or non-parametric, e.g., rank-based Mann-Whitney-Wilcoxon tests (Carifio & Perla, 2008; Jamieson, 2004). However, research shows that t-tests will tend to work fine as long as the data are unimodal and the N is larger than 40 (Lumley, Diehr, Emerson, & Chen, 2002). A total of 212 students and 162 teachers participated in the questionnaire surveys in P6. Thus, our data met the size criterion specified by Lumley et al. (2002). Also, the literature shows that the non-parametric Mann-Whitney tests can be alternative to independent t-tests as both have equivalent power (false-positive type 1 error, false-negative type II error) for most of the data (De Winter & Dodou, 2010). Hence, to mitigate the threat to conclusion validity from the choice of t-tests, the same data have also been analyzed by non-parametric Mann-Whitney U tests (as an alternative to independent t-tests) and Wilcoxon signed-rank test (as an alternative to one-sample t-tests, with the neutral alternative 3 as hypothesized median value).

The comparison was made mainly on the tests where we got significant results in P6. The results indicated similarities in significance from both tests but with slight variation in effect size. For instance, Table 9 and Table 10 show similar statistical significance between students and teachers responses for ease of cheating for forbidden aids, peeking, outside assistance, and student-staff collusion with both t-tests and Mann Whitney U-test. Similarly, non-parametric Wilcoxon signed-rank tests indicated similarities in significance with one-sample t-tests on students and teachers responses for the comparison between paper exams and BYOD exams. Table 11 shows the results for ease of cheating for Paper vs Univ PC exams and BYOD vs Univ PC exams.

Table 9. Parametric tests on opinions on ease of cheating in paper exams and BYOD e-exams

Type of Cheating threat	Students (t-test)				Teachers (t-test)				Students and Teachers (independent t-test)	
	Mean	SD	P-value	Mean Diff.	Mean	SD	P-value	Mean Diff	Sig.	Mean Diff
Impersonation	2.94	.498	.140	-.061	3.08	.679	.230	.084	.056	-.145
Forbidden aids	3.26	1.120	.006**	.257	3.68	1.187	.000**	.684	.005*	-.427
Peeking	3.44	1.032	.000**	.439	3.07	.948	.451	.074	.005*	.366
Peer collaboration	3.35	.764	.000**	.351	3.56	.908	.000**	.558	.068	-.207
Outside assistance	3.51	.778	.000**	.507	3.77	.944	.000**	.766	.027*	-.259
Student-staff collusion	2.94	.470	.118	-.061	3.15	.699	.043*	.147	.012*	-.208

*($p < 0.05$), ** ($p < 0.01$), and ***($p < 0.001$).

Table 10. Non-parametric tests on opinions on ease of cheating in paper exams and BYOD e-exams

Type of Cheating threat	Students (wilcoxon signed-rank test)			Teachers (wilcoxon signed-rank test)			Students and Teachers (Mann Whitney U-test)	
	Test Statistic	Z	Sig.	Test Statistic	Z	Sig.	Z	Sig.
Impersonation	111.500	-1.480	.139	138.000	1.274	.203	-1.617	.106
Forbidden aids	3895.000	2.658	.008**	1836.000	4.823	.000**	-2.815	.005*
Peeking	4516.000	4.644	.000**	679.000	.700	.484	-3.126	.002*
Peer collaboration	1659.000	5.089	.000**	1396.000	5.111	.000**	-1.908	.056
Outside assistance	1657.000	6.430	.000**	1695.000	5.936	.000**	-2.581	.010*
Student-staff collusion	73.500	-1.560	.119	175.000	2.135	.033*	-2.277	.023*

*($p < 0.05$), ** ($p < 0.01$), and ***($p < 0.001$).

Table 11. Opinions on ease of cheating for Paper vs Univ PC exams and BYOD vs Univ PC exams

Type of Cheating threat	Students and Teachers Sig. from independent t-test	Students and Teachers Sig. from Mann-Whitney U tests
Forbidden aids in Paper vs Univ PC exams	.003**	.005**
Peeking in Paper vs Univ PC exams	.009**	.003**
Peeking in BYOD vs Univ PC exams	.015*	.014*
Peer collaboration in BYOD vs Univ PC exams	.026*	.009**
Student Staff collusion in BYOD vs Univ PC exams	.030*	.010*

6. Conclusion and Future Work

6.1. Conclusion

The research aim of this paper was to investigate ***RQ: How can e-exam systems contribute to achieving an effective digital ecosystem for e-learning?*** The digital ecosystem phenomenon within e-exam systems has been addressed by investigating how two dedicated e-exam systems – Inspera Assessment and WISEflow become key parts of an effective digital ecosystem for e-exam systems. This section will wrap up this thesis by summarizing the main conclusions drawn for each sub question and further provide concluding remark for main research question towards the end of this section.

SQ1. To what extent is the risk of cheating an obstacle to the adoption of e-exams, and how do e-exams compare to traditional pen and paper exams when it comes to cheating risks?

Our literature review and studies indicated that cheating had been a big concern towards the adoption of e-exams in the higher education sector, not only in Norway but in many universities around the world. There were a number of concerns and issues reported in our studies, especially for exams using student-owned devices (i.e., BYOD e-exams), which – in spite of increased cheating threats – are preferred in many universities for scalability advantages. While many technological solutions are available to mitigate cheating, with the advance in technology, students' use of cheating technology is also rising in exams. However, the problem here is not about these traditional ways of cheating (collusion through signals and codes, whispering, use tiny wireless earpiece, hidden cameras, smart glasses, smartwatches, etc.), more about cheating through the e-exam system itself. While e-exam systems come with several countermeasures to avoid cheating, still they need human effort to mitigate cheating.

SQ1 provided comparisons between traditional exams and e-exams through threat analysis, penetration testing, empirical findings from mixed-method research. We focused our investigation mainly on cheating during exams, not before or after the exams. The comparison between traditional exams and e-exams was mainly focused on cheating through impersonation, forbidden aids, and assistance/collaboration. The comparison was done using threat analysis, penetration testing and empirical research through surveys and interviews. Our threat analysis indicated that e-exams have additional cheating threats. Penetration tests further showed that some of the threats identified in threat analysis were viable in practice. Our empirical study with students and teachers further strengthens the results from threat analysis and penetration testing that cheating can be easier with e-exams, especially with student-owned devices. The key finding drawn from this research question was that most of the countermeasures, including mixed seating, variation in question sets, moving calculators and books into the exam system, strict question/answer sequence, automated plagiarism, and biometric authentication, would be effective against cheating during e-exams. Moreover, our empirical findings with vendors and system managers indicate that, although acknowledging that e-exams were not 100% secure,

vendors and managers felt that there was much more cheating taking place outside of the electronic exam infrastructure (e.g., using old-fashioned cheat notes or concealed mobile phones) than what was done inside this infrastructure (e.g., breaking lockdown and using forbidden aids via the exam PC itself).

SQ2. What are the key requirements for e-exam systems, how are such requirements established, and how does the requirements process for acquisition and development of e-exam systems relate to approaches used for requirements in the field of software ecosystems?

This research question was investigated via a systematic mapping review and a case study with vendors, process managers and system managers in the Norwegian higher education sector.

Vendors, system managers and process managers considered authoring, logistics support, question analytics, grading, explanation of grades, appeals and complaint grading as key functional features for e-exam systems. As for key quality attributes, they focused on scalability, usability, integration and interoperability, security, and reliability. There was comparatively more mention of question authoring, explanation of grades, appeals grading, integration and interoperability and security.

As for requirements establishment, our study indicated that the requirements process for e-exam systems acquisition in Norway is still somewhat ad hoc. However, it shows some improvement compared to other researchers' observations about the process a couple of years earlier. Especially, the arrangement to have one national organization – Unit – oversee the requirements and integration process seemed to be satisfactory to the participants. No participant indicated a wish to return to the previous situation when each institution would run separate procurement and integration projects. Our study indicated that a similar approach has been used also in Sweden and Netherlands.

The requirements engineering process in both the software ecosystems field and in e-learning are still appears to be somewhat immature. However, our findings indicate that the requirements engineering process used for acquisition and development of the e-exam systems has similarities with requirements engineering processes proposed for software ecosystems. Role identification is one of the important properties for requirements engineering process of software ecosystems. Our participants indicated a clearer assignment of responsibilities between vendors, process managers, and system managers. For instance, vendors implemented e-exam software to mass-market needs and opened their APIs to universities to customize their platform. Unit as process manager used the Mulesoft ESB framework to do integrations between the interfaces of the e-exam systems with supporting systems. There was also a clear collaboration reported for requirements elicitation, prioritization, and negotiation.

SQ3. What are key obstacles towards achieving the interoperability needed for a digital ecosystem for e-exams and e-assessment?

This research question provided findings from two case studies on e-assessment systems. While our first case study was based on our own experience as users of the

tools, the second case study reported findings from vendors, system managers, and process managers from higher education. The key obstacles towards achieving the openness and interoperability needed for a digital ecosystem for e-exams and e-assessment identified from our case studies were security, as well lacking priority for interoperability from vendors and customers.

As for key obstacles to achieving openness were strict access control to APIs from vendors and limited access to application API. For instance, system managers can access only a few parts of APIs mainly being used for administration workflows, e.g., registering users. Vendors felt that lack of centralised control on their APIs could enable the users of their systems to pull the data that they are not allowed to use. On the other hand, system managers asserted that it would not affect their workflows since they would only perform verification rather than integration at the university site. Vendors addressed key obstacle for achieving openness in relation to security, they felt that open sharing of questions within the exam system would allow users to share the data they are not allowed to share.

As for key obstacles towards interoperability, lacking priority for interoperability from vendors and customers was reported as a main cause for lacking interoperability between systems. Our findings show that most of the system managers did not feel that it was the highest priority for the moment to achieve the interoperability needed to support an open ecosystem architecture. Vendors mentioned that they would prioritise interoperability higher if they get clear requirements from customers in that direction.

Overall, from SQ1-SQ3, we found many examples of features and developments that the stakeholders were satisfied with and which they considered a substantial improvement over the previous situation with more manual work processes. At the same time, we also found examples of poor interoperability between various systems, and although stakeholders responded positively about the idea of taking an ecosystem perspective in principle, it seemed more difficult for interoperability to be highly prioritized in actual development of products, where the addition of more functional features tended to receive more attention. Hence, we must acknowledge that the main research question could not be fully answered. The area of digital exams and their tool support is still somewhat immature, and in spite of some progress, the effective ecosystem is far from being achieved. To progress along that way, it will be important that the acquisition and development process has a key stakeholder involved who understands the need to prioritize interoperability and has the power to push that perspective.

6.2. Future Work

The work presented in this thesis could be further extended in several ways.

One possible extension could be more detailed threat analyses, considering the latest developments in e-exam technology, or inviting more persons to participate in the threat analysis (e.g., e-exam security experts), thus getting higher credibility for the results. Broader threat analysis would also be possible, e.g., looking more at other forms of assessment (e.g., group projects, semester-long course-work), not just the

typical individual exams that have been the main focus in this research, or also looking at cheating before and after the exam, and cheating by employees, not just by students.

As for penetration testing, one possibility for further work would be to use newer versions of the target software and more complete solutions and testing a wider range of technology, e.g., new version of SEB, penetration testing of IA, WF and other e-exam tools, and of LMSs used for exam purposes.

When it comes to the empirical investigation, our investigations had clear limitations in the number and selection of informants. Our questionnaire investigations only targeted students and teachers from one university (NTNU), so one possible avenue for further work could be to do similar investigations with students and teachers in other universities, possibly also in other countries. However, it could be questioned whether simply redoing the same questionnaire survey with a larger group of informants would be the best way to continue, as more interesting results might be achieved if also trying to improve the questionnaire instrument itself. The survey in P6 was very much targeted towards a comparison between paper exams and e-exams. As e-exams gradually replace paper exams in most universities, this comparison will likely appear less relevant, and other comparisons might be more relevant (e.g., comparing cheating risks of various types of e-assessment, rather than with paper).

Considering the interview investigations, an interesting follow-up could be to interview more stakeholders, e.g., including more system managers from different universities, and vendors of additional e-exam products. Also, it would be interesting to include interviews with students, and teachers, to find out how well (or not) the managers' views on requirements for e-exam systems are aligned with the views of students and teachers as key end-user groups.

Further empirical research investigating in more detail the enablers and barriers for open digital ecosystems for e-exams would also be interesting. Due to the early stage in integration between the tools we studied in this thesis, we were unable to find more empirical evidence on the security aspects concerning e-exam ecosystems. Thus, it would be interesting to research in particular how the integrations affect tools, whether changes in one tool would affect other tools related to security, and to what extent cheating concerns and related security requirements form an obstacle towards achieving an open digital ecosystem around e-assessment.

Design research could be another possibility for follow-up research to this thesis, e.g., coming up with frameworks for how to achieve better interoperability, sharing of huge question banks between universities worldwide, making prototypes of plugins for various needs to demonstrate the viability of a plugin approach or if it does not work, document precisely current shortcomings of current e-exam platforms for supporting plugins.

As indicated by this thesis, the progress towards open digital ecosystems in e-assessment is still limited. Vendors, educators, and researchers should make more strategic decisions towards progressing interoperability between tools through increased openness, not only nationally but also internationally. The possibility to share the labour-intensive effort of developing high-quality tests and other learning

resources across universities can be beneficial both to teachers and their students. In that regard, ecosystem thinking will be important solution for the future of e-assessment, thus to the future of higher education.

Appendix

Information letter and consent form

Are you interested in taking part in the research project "Digital Exams Security"?

This is an inquiry about participation in a research project where the primary purpose is to investigate the participants' experiences on digital exams. In this letter, we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

The main focus of the project is to investigate different stakeholders (i.e. Vendors, Customers, Purchasers, Teachers, Students, System managers) experiences on digital exams to improve the implementation and security of exams. These exams could be traditional paper-based exams, e-exams and BYOD (Bring Your Own Devices). This study is part of empirically-oriented research by a PhD candidate, and the responses gathered from participants can be analyzed to 1) identify the key requirements, challenges of digital exam systems 2) compare whether some way of cheating is more difficult with paper exams vs e-exams, 3) compare whether some type of mitigation is easier or more difficult with paper exams vs e-exams and 4) stakeholders envision a move towards a more open digital ecosystem for e-exams.

Sample Selection

The participants will be randomly selected voluntarily. For the survey, if participants have some knowledge on digital exams, it is a bonus but not necessarily require more knowledge on digital exams. For interviews, the participant should have supervised or authored or involved in the implementation or administration of the digital exam. The participants need to be people who work with the implementation and risk analysis of exams, people involved in organizing exams. The study will take place both on the web, and the various buildings at NTNU and the participants will be recruited via emails and printed advertisements.

What does participation involve for you?

The participants will be engaged in interviews or group interviews. During the interviews, participants will be asked regarding their experiences with digital exams. Interviews will be conducted at the institutions and through online if the participant stays far from NTNU. The total duration of the interview will take approx. 40 minutes.

Participation is voluntary

Participation in the project is voluntary. If you chose to participate, you could withdraw your consent at any time without giving a reason. All information about you will then be made anonymous.

Your personal privacy – how we will store and use your personal data

We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially under data protection legislation (the General Data Protection Regulation and Personal Data Act).

- The data will not be handed to other researchers or professors at NTNU or outside of the institution.
- If we collect IP addresses, they will be removed before we start analyzing the data.

What will happen to your personal data at the end of the research project?

The project is scheduled to end 2020. The data will be stored in a university computer, and it will be kept until the end of the project after that data will be deleted.

Your rights

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?

We will process your personal data based on your consent.

Based on an agreement with Dept. of Computer Science - NTNU, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is following data protection legislation.

Where can I find out more?

If you have questions about the project, contact:

- Supervisor: Professor Guttorm Sindre at guttorm.sindre@ntnu.no
- PhD candidate: Aparna Chirumamilla at aparna.vegendla@ntnu.no

Yours sincerely,
Aparna Chirumamilla

Consent form

I hereby confirm that I have been fully informed about the aims and purposes of the study and the project in general. I understand that my participation is entirely

voluntary and, if I no longer wish to participate, I may at any stage withdraw my participation. I have been informed and understand that my participation in the research involves use of surveys, interviews, focus group interviews and that aggregated data will be stored and analyzed for the purposes of the project.

I have received and understood information about the project "Digital Exams Security" and have been given the opportunity to ask questions. I give consent:

- ☐ to participate in Focus group interview
- ☐ to participate in interview
- ☐ to participate in survey
- ☐ for audio recording

I give consent for my personal data to be processed until the end date of the project - 31/12/2020

(Signed by participant, date)

NSD Approval

Det innsendte meldeskjemaet med referansekode 218037 er nå vurdert av NSD.

Følgende vurdering er gitt:

Our assessment is that the processing of personal data in this project will comply with data protection legislation, presupposing that it is carried out in accordance with the information given in the Notification Form and attachments, 08.11.2018, as well as dialogue with NSD. Everything is in place for the processing to begin.

NOTIFY CHANGES

If you intend to make changes to the processing of personal data in this project it may be necessary to notify NSD. This is done by updating the Notification Form. On our website we explain which changes must be notified. Wait until you receive an answer from us before you carry out the changes.

TYPE OF DATA AND DURATION

The project will be processing general categories of personal data until 31.12.2020.

LEGAL BASIS

The project will gain consent from data subjects to process their personal data. We find that consent will meet the necessary requirements under art. 4 (11) and 7, in that

it will be a freely given, specific, informed and unambiguous statement or action, which will be documented and can be withdrawn. The legal basis for processing personal data is therefore consent given by the data subject, cf. the General Data Protection Regulation art. 6.1 a).

PRINCIPLES RELATING TO PROCESSING PERSONAL DATA

NSD finds that the planned processing of personal data will be in accordance with the principles under the General Data Protection Regulation regarding:

- lawfulness, fairness and transparency (art. 5.1 a), in that data subjects will receive sufficient information about the processing and will give their consent

- purpose limitation (art. 5.1 b), in that personal data will be collected for specified, explicit and legitimate purposes, and will not be processed for new, incompatible purposes

- data minimisation (art. 5.1 c), in that only personal data which are adequate, relevant and necessary for the purpose of the project will be processed

- storage limitation (art. 5.1 e), in that personal data will not be stored for longer than is necessary to fulfil the project's purpose

THE RIGHTS OF DATA SUBJECTS

Data subjects will have the following rights in this project: transparency (art. 12), information (art. 13), access (art. 15), rectification (art. 16), erasure (art. 17), restriction of processing (art. 18), notification (art. 19), data portability (art. 20). These rights apply so long as the data subject can be identified in the collected data.

NSD finds that the information that will be given to data subjects about the processing of their personal will meet the legal requirements for form and content, cf. art. 12.1 and art. 13.

We remind you that if a data subject contacts you about their rights, the data controller has a duty to reply within a month.

FOLLOW YOUR INSTITUTION'S GUIDELINES

NSD presupposes that the project will meet the requirements of accuracy (art. 5.1 d), integrity and confidentiality (art. 5.1 f) and security (art. 32) when processing personal data.

Selectsurvey is a data processor for the project. NSD presupposes that the processing of personal data by a data processor meets the requirements under the General Data Protection Regulation arts. 28 and 29.

To ensure that these requirements are met you must follow your institution's internal guidelines and/or consult with your institution (i.e. the institution responsible for the project).

FOLLOW-UP OF THE PROJECT

NSD will follow up the progress of the project underway (every other year) and at the planned end date in order to determine whether the processing of personal data has been concluded/is being carried out in accordance with what is documented.

Good luck with the project!

Contact person at NSD: Kajsa Amundsen

Data Protection Services for Research: +47 55 58 21 17 (press 1)

References

Abad-Segura, E., González-Zamar, M.-D., Infante-Moro, J. C., & Ruipérez García, G. (2020). Sustainable Management of Digital Transformation in Higher Education: Global Research Trends. *Sustainability, 12*(5), 2107.

Adebayo, O., & Abdulhamid, S. M. (2014). E-exams system for Nigerian universities with emphasis on security and result integrity. *International Journal of the Computer, the Internet and Management (IJCIM), 18*(2), 1-12.

Alharthi, A. D., Spichkova, M., & Hamilton, M. (2019). Sustainability requirements for eLearning systems: a systematic literature review and analysis. *Requirements Engineering, 24*(4), 523-543.

Angeren, J. v., Kabbedijk, J., Jansen, S., & Popp, K. M. (2011). A survey of associate models used within large software ecosystems. In *Proceedings of the Third International Workshop on Software Ecosystems(IWSECO2011),* (pp. 27-39). Brussels, Belgium.

Aparicio, M., Bacao, F., & Oliveira, T. (2016). An e-learning theoretical framework. *An e-learning theoretical framework*(1), 292-307.

Appiah, M., & Van Tonder, F. (2018). E-Assessment in Higher Education: A Review. *International Journal of Business Management & Economic Research, 9*(6).

Arkin, B., Stender, S., & McGraw, G. (2005). Software penetration testing. *IEEE Security & Privacy*, 84-87.

Axelsson, J., & Skoglund, M. (2016). Quality assurance in software ecosystems: A systematic literature mapping and research agenda. *Journal of Systems and Software, 114*, 69-81.

Baró-Solé, X., Guerrero-Roldan, A. E., Prieto-Blázquez, J., Rozeva, A., Marinov, O., Kiennert, C., Rocher, P. O., & Garcia-Alfaro, J. (2018). Integration of an adaptive trust-based e-assessment system into virtual learning environments—The TeSLA project experience. *Internet Technology Letters, 1*(4), e56.

Bernardi, R. A., Baca, A. V., Landers, K. S., & Witek, M. B. (2008). Methods of Cheating and Deterrents to Classroom Cheating: An International Study. *Ethics & Behavior, 18*(4), 373-391.

Bezzi, M., Damiani, E., Paraboschi, S., & Plate, H. (2013). Integrating Advanced Security Certification and Policy Management. In *Felici M. (eds) Cyber Security and Privacy. CSP* (Vol. 182, pp. 55-66): Springer Berlin Heidelberg.

Bianco, A., De Marsico, M., & Temperini, M. (2005). Standards for e-Learning. *The TISIP Foundation, Trondheim, Norway*.

Biggs, J. (2003). Aligning teaching for constructing learning. *Higher Education Academy, 1*(4).

Biggs, J. B. (2011). *Teaching for quality learning at university: What the student does*: McGraw-hill education (UK).

Bilen, E., & Matros, A. (2021). Online cheating amid COVID-19. *Journal of Economic Behavior & Organization, 182*, 196-211.

Bird, C. (1929). An improved method of detecting cheating in objective examinations. *The Journal of Educational Research, 19*(5), 341-348.

Bishop, M. (2007). About Penetration Testing. *IEEE Security & Privacy, 5*(6), 84-87.

Boeije, H. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity, 36*(4), 391-409.

Boezerooy, P., Cordewener, B., & Liebrand, W. (2007). Collaboration on ICT in Dutch Higher Education: The SURF Approach. *Educause Review, 42*(3).

Bond, M., Marín, V. I., Dolch, C., Bedenlier, S., & Zawacki-Richter, O. (2018). Digital transformation in German higher education: student and teacher perceptions and usage of digital media. *International Journal of Educational Technology in Higher Education, 15*(1), 48.

Borcan, O., Lindahl, M., & Mitrut, A. (2014). The impact of an unexpected wage cut on corruption: Evidence from a "Xeroxed" exam. *Journal of Public Economics, 120*, 32-47.

Bosch, J. (2010). Architecture challenges for software ecosystems. In *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume* (pp. 93-95). Copenhagen, Denmark: ACM.

Bosch, J., & Bosch-Sijtsema, P. (2010). Coordination between global agile teams: from process to architecture. In *Agility Across Time and Space* (pp. 217-233): Springer.

Boucharas, V., Jansen, S., & Brinkkemper, S. (2009). Formalizing software ecosystem modeling. In *Proceedings of the 1st international workshop on Open component ecosystems* (pp. 41-50). Amsterdam, The Netherlands: ACM.

Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: what are they and what are their benefits? *Assessment & Evaluation in Higher Education, 34*(3), 305-319.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Brink, R., & Lautenbach, G. (2011). Electronic assessment in higher education. *Educational Studies, 37*(5), 503-512.

Burrus, R. T., McGoldrick, K., & Schuhmann, P. W. (2007). Self-reports of student cheating: Does a definition of cheating matter? *The Journal of Economic Education, 38*(1), 3-16.

Buzzetto-More, N. A., & Alade, A. J. (2006). Best practices in e-assessment. *Journal of Information Technology Education: Research, 5*(1), 251-269.

Campbell, P. R. J., & Ahmed, F. (2010). A Three-dimensional View of Software Ecosystems. In *Proceedings of the Fourth European Conference on Software Architecture* (pp. 81-84): ACM.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical education, 42*(12), 1150-1152.

Carpenter, D. D., Harding, T. S., Finelli, C. J., Montgomery, S. M., & Passow, H. J. (2006). Engineering students' perceptions of and attitudes towards cheating. *Journal of Engineering Education, 95*(3), 181-194.

Carter, T. M. (1928). What College Students Think with Respect to Cheating in Examination. *The Phi Delta Kappan, 11*(1), 3-10.

Cerimagic, S., & Hasan, M. R. (2019). Online exam vigilantes at Australian universities: Student academic fraudulence and the role of universities to counteract. *Universal Journal of Educational Research, 7*(4), 929-936.

Chandra, P., Wohleber, T., Feragamo, J., & Williams, J. (2007). CLASP v1.2: comprehensive, lightweight application security process. In. OWASP.

Chang, V., & Guetl, C. (2007). E-Learning Ecosystem (ELES) - A Holistic Approach for the Development of more Effective Learning Environment for Small-and-Medium Sized Enterprises (SMEs). In *2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference* (pp. 420-425).

Chang, V., & Uden, L. (2008). Governance for E-learning ecosystem. In *2nd IEEE International Conference on Digital Ecosystems and Technologies, 2008. DEST 2008* (pp. 340-345).

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*: sage.

Chirumamilla, A., Sindre, G., & Nguyen-Duc, A. (2020). Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway. *Assessment & Evaluation in Higher Education, 45*(7), 940-957.

Chituc, C.-M., Herrmann, M., Schiffner, D., & Rittberger, M. (2019). Towards the Design and Deployment of an Item Bank: An Analysis of the Requirements Elicited. In (pp. 155-162). Cham: Springer International Publishing.

Chituc, C.-M., & Rittberger, M. (2019). Understanding the Importance of Interoperability Standards in the Classroom of the Future. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society* (Vol. 1, pp. 6801-6806).

Chituc, C. (2020). Interoperability Standards in the IoT-enabled Future Learning Environments: An analysis of the challenges for seamless communication. In *2020 13th International Conference on Communications (COMM)* (pp. 417-422).

Christensen, H. B., Hansen, K. M., Kyng, M., & Manikas, K. (2014). Analysis and design of software ecosystem architectures – Towards the 4S telemedicine ecosystem. *Information and Software Technology, 56*(11), 1476-1492.

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*: Routledge.

Cizek, G. J. (2004). Cheating in academics. *Encyclopedia of applied psychology, 1*, 307.

Claes, M., Mens, T., & Grosjean, P. (2014). On the maintainability of CRAN packages. In *2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering, CSMR-WCRE 2014 - Proceedings* (pp. 308-312).

Cluskey Jr, G., Ehlen, C. R., & Raiborn, M. H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics, 4*(1).

Colnerud, G., & Rosander, M. (2009). Academic dishonesty, ethical norms and learning. *Assessment & Evaluation in Higher Education, 34*(5), 505-517.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology, 13*(1), 3-21.

Creswell, J. W. (2013). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Harlow, United Kingdom: Pearson Education Limited.

Creswell, J. W. (2014). *A concise introduction to mixed methods research*: SAGE publications.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*: Sage publications.

Crisp, G. (2011). Teacher's Handbook on e-Assessment. *Transforming Assessment-An ALTC Fellowship Activity, 18*.

Daclin, N., Daclin, S. M., Chapurlat, V., & Vallespir, B. (2016). Writing and verifying interoperability requirements: Application to collaborative processes. *Computers in Industry, 82*, 1-18.

Dagger, D., Connor, A. O., Lawless, S., Walsh, E., & Wade, V. P. (2007). Service-Oriented E-Learning Platforms: From Monolithic Systems to Flexible Services. *IEEE Internet Computing, 11*(3), 28-35.

Dantas, V. L. L., Marinho, F. G., Costa, A. L. d., & Andrade, R. M. C. (2009). Testing requirements for mobile applications. In *2009 24th International Symposium on Computer and Information Sciences* (pp. 555-560).

Dawson, P. (2014). Our anonymous online research participants are not always anonymous: Is this a problem? *British Journal of Educational Technology, 45*(3), 428-437.

Dawson, P. (2016). Five ways to hack and cheat with bring-your-own-device electronic examinations. *British Journal of Educational Technology, 47*(4), 592-600.

Dawson, P. (2020). *Defending assessment security in a digital world: preventing e-cheating and supporting academic integrity in higher education*: Routledge.

De Lambert, K., Ellen, N., & Taylor, L. (2006). Chalkface challenges: a study of academic dishonesty amongst students in New Zealand tertiary institutions. *Assessment & Evaluation in Higher Education, 31*(5), 485-503.

De Winter, J., & Dodou, D. (2010). Five-point likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). *Practical Assessment, Research, and Evaluation, 15*(1), 11.

Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology, 40*(2), 203-214.

Dick, J., Hull, E., & Jackson, K. (2017). *Requirements engineering*: Springer.

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring Relations Between Formative and Summative Assessment. In J. Dolin & R. Evans (Eds.), *Transforming Assessment: Through an Interplay Between Practice, Research and Policy* (pp. 53-80). Cham: Springer International Publishing.

Dong, B., Zheng, Q., Yang, J., Li, H., & Qiao, M. (2009). An E-learning Ecosystem Based on Cloud Computing Infrastructure. In *Ninth IEEE International Conference on Advanced Learning Technologies, 2009. ICALT 2009* (pp. 125-127).

Downing, S. M. (2002). Threats to the validity of locally developed multiple-choice tests in medical education: Construct-irrelevant variance and construct underrepresentation. *Advances in Health Sciences Education, 7*(3), 235-241.

Dreier, J., Giustolisi, R., Kassem, A., Lafourcade, P., & Lenzini, G. (2015). A Framework for Analyzing Verifiability in Traditional and Electronic Exams. In (pp. 514-529). Cham: Springer International Publishing.

Dyba, T., Dingsoyr, T., & Hanssen, G. K. (2007). Applying systematic reviews to diverse study types: An experience report. In *First international symposium on empirical software engineering and measurement (ESEM 2007)* (pp. 225-234): IEEE.

Eklund, U., & Bosch, J. (2014). Architecture for embedded open software ecosystems. *Journal of Systems and Software, 92*, 128-142.

Ettien, A. (2018). A Study of Proctors' Involvement in National Examination Cheating: The Case of" Collège Privé MBF d'Abobo" Exam Center. *Journal of Education and e-Learning Research, 5*(1), 22-27.

Fahl, S., Dechand, S., Perl, H., Fischer, F., Smrcek, J., & Smith, M. (2014). Hey, NSA: Stay Away from My Market! Future Proofing App Markets Against Powerful Attackers. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1143-1155): ACM.

Fendler, R. J., Yates, M. C., & Godbey, J. M. (2018). Observing and Deterring Social Cheating on College Exams. *International Journal for the Scholarship of Teaching and Learning, 12*(1), 4.

Fernandez, E. B., Yoshioka, N., & Washizaki, H. (2015). Patterns for security and privacy in cloud ecosystems. In *Evolving Security and Privacy Requirements Engineering (ESPRE), 2015 IEEE 2nd Workshop on* (pp. 13-18).

Fernandez, E. B., Yoshioka, N., Washizaki, H., & Syed, M. H. (2016). Modeling and security in cloud ecosystems. *Future Internet, 8*(2).

Firesmith, D. (2003). Specifying good requirements. *Journal of Object Technology, 2*(4), 77-87.

Fitzharris, R., & Kent, S. (2020). Adoption of Bring-Your-Own-Device Examinations and Data Analytics. In D. Ifenthaler & D. Gibson (Eds.), *Adoption of Data Analytics in Higher Education Learning and Teaching* (pp. 327-348). Cham: Springer International Publishing.

Fluck, A., Adebayo, O. S., & Abdulhamid, S. i. M. (2017). Secure e-examination systems compared: Case studies from two countries. *Journal of Information Technology Education: Innovations in Practice, 16*(1), 107-125.

Fluck, A., Pálsson, H., Coleman, M., Hillier, M., Schneider, D., Frankl, G., & Uolia, K. (2017). eExam symposium: design decisions and implementation experience. In *11th IFIP TC 3 World Conference on Computers in Education*.

Fluck, A. E. (2019). An international review of eExam technologies and impact. *Computers & Education, 132*, 1-15.

Foss-Pedersen, R. J., & Begnum, M. E. N. N. (2017). Universell utforming og digital eksamen i UH-sektoren: 5 anbefalte tiltakspunkter. In *Norsk konferanse for organisasjoners bruk at IT* (Vol. 25).

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education*: New York: McGraw-Hill Humanities/Social Sciences/Languages.

Frankl, G., Schartner, P., & Zebedin, G. (2012). Secure online exams using students' devices. In *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-7).

Fricker, S. (2009). Specification and Analysis of Requirements Negotiation Strategy in Software Ecosystems. In *in IWSECO@ ICSR*.

Fricker, S. (2010). Requirements Value Chains: Stakeholder Management and Requirements Engineering in Software Ecosystems. In R. Wieringa & A. Persson (Eds.), *Requirements Engineering: Foundation for Software Quality* (pp. 60-66): Springer Berlin Heidelberg.

García-Holgado, A., & García-Peñalvo, F. J. (2014). Knowledge management ecosystem based on drupal platform for promoting the collaboration between public administrations. In *Proceedings of the*

Second International Conference on Technological Ecosystems for Enhancing Multiculturality (pp. 619-624).

García-Holgado, A., & García-Peñalvo, F. J. (2016). Architectural pattern to improve the definition and implementation of eLearning ecosystems. *Science of Computer Programming, 129*, 20-34.

García-Holgado, A., & García-Peñalvo, F. J. (2018). Human Interaction in Learning Ecosystems Based on Open Source Solutions. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Design, Development and Technological Innovation* (pp. 218-232). Cham: Springer International Publishing.

García Peñalvo, F. J., Conde García, M. Á., Alier Forment, M., & Casany Guerrero, M. J. (2011). Opening learning management systems to personal learning environments. *Journal of universal computer science: J. UCS, 17*(9), 1222-1240.

Geraci, A., Katki, F., McMonegal, L., Meyer, B., Lane, J., Wilson, P., Radatz, J., Yee, M., Porteous, H., & Springsteel, F. (1991). *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*: IEEE Press.

Gherbi, A., Charpentier, R., & Couture, M. (2011). Software diversity for future systems security. *CrossTalk, 24*(5), 10-13.

Gilliom, J., & Monahan, T. (2012). *SuperVision: An introduction to the surveillance society*: University of Chicago Press.

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The discovery of grounded theory; strategies for qualitative research. *Nursing research, 17*(4), 364.

Goldbach, T., & Benlian, A. (2015). Understanding informal control modes on software platforms -The mediating role of third-party developers' intrinsic motivation. In *2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015*.

Grünigen, D. v., Souza, F. B. d. A. e., Pradarelli, B., Magid, A., & Cieliebak, M. (2018). Best practices in e-assessments with a special focus on cheating prevention. In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 893-899).

Guetl, C., & Chang, V. (2008). Ecosystem-based Theoretical Models for Learning in Environments of the 21st Century. *International Journal of Emerging Technologies in Learning (iJET), 3*(2008).

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17-27.

Hammersley, M., & Traianou, A. (2012). *Ethics and educational research*: British Educational Research Association London.

Handoyo, E., Jansen, S., & Brinkkemper, S. (2013). Software ecosystem roles classification. In *Lecture Notes in Business Information Processing* (Vol. 150 LNBIP, pp. 212-216).

Hansen, K. M., & Zhang, W. (2013). Towards structure-based quality awareness in software ecosystem use. In *International Conference on Service-Oriented Computing, ICSOC*. Berlin, Germany.

Harmon, O. R., & Lambrinos, J. (2008). Are Online Exams an Invitation to Cheat? *The Journal of Economic Education, 39*(2), 116-125.

Hellas, A., Leinonen, J., & Ihantola, P. (2017). 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* (pp. 238–243). Bologna, Italy: Association for Computing Machinery.

Hillier, M., & Fluck, A. (2013). Arguing again for e-exams in high stakes examinations. *Electric dreams. proceedings ascilite*, 385-396.

Hillier, M., & Fluck, A. (2017). Transforming Exams-How IT works for BYOD e-Exams. Australasian Society for Computers in Learning in Tertiary Education conference (pp. 100-105). Toowoomba, 4-6 December. In.

Howard, M., & Lipner, S. (2006). *The security development lifecycle* (Vol. 8): Microsoft Press Redmond.

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice, 16*(3), 269-290.

Huszti, A., & Petho, A. (2010). A secure electronic exam system. *Publicationes Mathematicae Debrecen, 77*(3-4), 299-312.

Ifijeh, G., Michael-Onuoha, H. C., Ilogho, J. E., & Osinulu, I. (2015). Emergence of hi-tech examination malpractices in Nigeria: issues and implications. *International Journal of education and Research, 3*(3), 113-122.

Immonen, A., Ovaska, E., Kalaoja, J., & Pakkala, D. (2016). A service requirements engineering method for a digital service ecosystem. *Service Oriented Computing and Applications, 10*(2), 151-172.

IMS. (2021). IMS Global Learning Consortium. Retrived from http://www.imsglobal.org/.

Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology, & Human Values, 41*(1), 17-49.

Isaias, P., Miranda, P., & Pífano, S. (2019). Framework for the analysis and comparison of e-assessment systems. In *ASCILITE 2017-Conference Proceedings-34th International Conference of Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 276-283): Australasian Society for Computers in Learning in Tertiary Education (ASCILITE).

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics, 118*(3), 843-877.

Jakimoski, K. (2016). Challenges of interoperability and integration in education information systems. *International Journal of Database and Theory and Application, 9*(2), 33-46.

James, N., & Busher, H. (2015). Ethical issues in online research. In: Taylor & Francis.

James, R. (2016). Tertiary student attitudes to invigilated, online summative examinations. *International Journal of Educational Technology in Higher Education, 13*(1), 1-13.

Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical education, 38*(12), 1217-1218.

Jamil, M., Tariq, R., & Shami, P. (2012). Computer-Based vs Paper-Based Examinations: Perceptions of University Teachers. *Turkish Online Journal of Educational Technology-TOJET, 11*(4), 371-381.

Jansen, S. (2013). How quality attributes of software platform architectures influence software ecosystems. In *Proceedings of the 2013 International Workshop on Ecosystem Architectures* (pp. 6-10). Saint Petersburg, Russia: ACM.

Jansen, S., Brinkkemper, S., & Finkelstein, A. (2009). Business Network Management as a Survival Strategy: A Tale of Two Software Ecosystems. In *First International Workshop on Software Ecosystems (IWSECO)* (pp. 34-38).

Jansen, S., & Cusumano, M. A. (2013). Defining software ecosystems: a survey of software platforms and business network governance. In *Software ecosystems*: Edward Elgar Publishing.

Jansen, S., Finkelstein, A., & Brinkkemper, S. (2009). A sense of community: A research agenda for software ecosystems. In *31st International Conference on Software Engineering - Companion Volume, 2009. ICSE-Companion 2009* (pp. 187-190).

Jansen, S., Handoyo, E., & Alves, C. (2015). Scientists' Needs in Modelling Software Ecosystems. In *Proceedings of the 2015 European Conference on Software Architecture Workshops* (pp. 44:41-44:46). Dubrovnik, Cavtat, Croatia: ACM.

Jansen, S., & van Capelleveen, G. (2013). Quality review and approval methods for extensions in software ecosystems. In *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry* (pp. 187-217): Edward Elgar Publishing.

Jesson, J., Matheson, L., & Lacey, F. M. (2011). Doing your literature review: Traditional and systematic techniques.

JISC. (2006). e-Assessment Glossary (Extended). Retrived May 18, 2021 from https://www.plymouth.ac.uk/uploads/production/document/path/2/2555/eAssess-Glossary-Extended-v1-01.pdf.

JISC. (2007). Effective Practice with e-Assessment: An overview of technologies, policies and practice in further and higher education. In: Joint Information Systems Committee.

Jones, L. R. (2011). Academic integrity & academic dishonesty: A handbook about cheating & plagiarism.

Kaiiali, M., Ozkaya, A., Altun, H., Haddad, H., & Alier, M. (2016). Designing a Secure Exam Management System (SEMS) for M-Learning Environments. *IEEE Transactions on Learning Technologies, 9*(3), 258-271.

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *MIS Q., 37*(2), 357-370.

Kallio, H., Pietila, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing, 72*(12), 2954-2965.

Karpati, P., Opdahl, A., & Sindre, G. (2013). HARM: Hacker Attack Representation Method. In J. Cordeiro, M. Virvou & B. Shishkov (Eds.), *Software and Data Technologies* (Vol. 170, pp. 156-175): Springer Berlin Heidelberg.

Karpati, P., Opdahl, A. L., & Sindre, G. (2015). Investigating security threats in architectural context: Experimental evaluations of misuse case maps. *Journal of Systems and Software, 104*, 90-111.

Kassem, A., Falcone, Y., & Lafourcade, P. (2015). Monitoring Electronic Exams. In (pp. 118-135). Cham: Springer International Publishing.

Katta, V., Karpati, P., Opdahl, A. L., Raspotnig, C., & Sindre, G. (2010). Comparing Two Techniques for Intrusion Visualization. In P. v. Bommel, S. Hoppenbrouwers, S. Overbeek, E. Proper & J. Barjis (Eds.), *The Practice of Enterprise Modeling* (pp. 1-15): Springer Berlin Heidelberg.

Kelly, B., Wilson, S., & Metcalfe, R. (2007). Openness in Higher Education: Open Source, Open Standards, Open Access. In *ELPUB* (pp. 161-174).

Khlifi, Y. (2020). An Advanced Authentication Scheme for E-evaluation Using Students Behaviors Over E-learning Platform. *International Journal of Emerging Technologies in Learning, 15*(4).

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology, 51*(1), 7-15.

Kitchenham, B. A., Budgen, D., & Pearl Brereton, O. (2011). Using mapping studies as the basis for further research – A participant-observer case study. *Information and Software Technology, 53*(6), 638-651.

Knauss, E., Damian, D., Knauss, A., & Borici, A. (2014). Openness and requirements: Opportunities and tradeoffs in software ecosystems. In *2014 IEEE 22nd International Requirements Engineering Conference (RE)* (pp. 213-222).

Knauss, E., Yussuf, A., Blincoe, K., Damian, D., & Knauss, A. (2016). Continuous clarification and emergent requirements flows in open-commercial software ecosystems. *Requirements Engineering*, 1-21.

Kocdar, S., & Dirkx, K. (2017). Innovative practices in e-Assessment: the TeSLA project. *The Envisioning Report for Empowering Universities*, 39.

Kordy, B., Mauw, S., Radomirović, S., & Schweitzer, P. (2014). Attack–defense trees. *Journal of Logic and Computation, 24*(1), 55-87.

Kordy, B., Piètre-Cambacédès, L., & Schweitzer, P. (2014). DAG-based attack and defense modeling: Don't miss the forest for the attack trees. *Computer Science Review, 13-14*, 1-38.

Kuikka, M., Kitola, M., & Laakso, M.-J. (2014). Challenges when introducing electronic exam. *2014, 22*.

Kurniawan, O., Lee, N. T. S., & Poskitt, C. M. (2020). Securing Bring-Your-Own-Device (BYOD) Programming Exams. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 880-886).

Küppers, B., Eifert, T., Zameitat, R., & Schroeder, U. (2020). EA and BYOD: Threat Model and Comparison to Paper-based Examinations. In *CSEDU (1)* (pp. 495-502).

Laine, K., Sipilä, E., Anderson, M., & Sydänheimo, L. (2016). Electronic exam in electronics studies. In *SEFI Annual Conference 2016*.

Lancaster, T., & Cotarlan, C. (2021). Contract cheating by STEM students through a file sharing website: a Covid-19 pandemic perspective. *International Journal for Educational Integrity, 17*(1), 1-16.

Lauesen, S. (2006). COTS tenders and integration requirements. *Requirements Engineering, 11*(2), 111-122.

Leite, J. C. S. d. P., & Cappelli, C. (2010). Software Transparency. *Business & Information Systems Engineering, 2*(3), 127-139.

Lima, T., Werner, C., & Santos, R. (2019). Identifying Architecture Attributes in the Context of Software Ecosystems Based on a Mapping Study. In (pp. 55-70). Cham: Springer International Publishing.

Llorens, F., Molina, R., Compañ, P., & Satorre, R. (2014). Technological Ecosystem for Open Education. In *IDT/IIMSS/STET* (pp. 706-715).

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual review of public health, 23*(1), 151-169.

Luo, M.-Y., & Lin, S.-W. (2013). From monolithic systems to a federated e-learning cloud system. In *2013 IEEE International Conference on Cloud Engineering (IC2E)* (pp. 156-165): IEEE.

Macnish, K., & van der Ham, J. (2020). Ethics in cybersecurity research and practice. *Technology in Society, 63*, 101382.

Manikas, K., & Hansen, K. M. (2013). Software ecosystems – A systematic literature review. *Journal of Systems and Software, 86*(5), 1294-1306.

Martin, S., Lopez-Martin, E., Moreno-Pulido, A., Meier, R., & Castro, M. (2019). A Comparative Analysis of Worldwide Trends in the Use of Information and Communications Technology in Engineering Education. *IEEE Access, 7*, 113161-113170.

McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (1999). Academic Integrity in Honor Code and Non-Honor Code Environments. *The Journal of Higher Education, 70*(2), 211-234.

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics &Behavior, 11*(3), 219-232.

McGraw, G. (2004). Software security. *IEEE Security & Privacy, 2*(2), 80-83.

Meland, P. H., Bernsmed, K., Frøystad, C., Li, J., & Sindre, G. (2019). An experimental evaluation of bow-tie analysis for security. *Information & Computer Security*.

Mellar, H., Peytcheva-Forsyth, R., Kocdar, S., Karadeniz, A., & Yovkova, B. (2018). Addressing cheating in e-assessment using student authentication and authorship checking systems: teachers' perspectives. *International Journal for Educational Integrity, 14*(1), 1-21.

Melve, I., & Smilden, B. (2015). ICT Architecture for Digital Assessment. In *UNINETT* (pp. 102).

Mhamdia, A. B. H. S. (2013). Performance measurement practices in software ecosystem. *International Journal of Productivity and Performance Management, 62*(5), 514-533.

Moskal, B. M., Leydens, J. A., & Pavelich, M. J. (2002). Validity, Reliability and the Assessment of Engineering Education. *Journal of Engineering Education, 91*(3), 351-354.

Myagmar, S., Lee, A. J., & Yurcik, W. (2005). Threat modeling as a basis for security requirements. In *Symposium on requirements engineering for information security (SREIS)* (Vol. 2005, pp. 1-8): Citeseer.

Nguyen-Duc, A., Cruzes, D. S., & Conradi, R. (2015). The impact of global dispersion on coordination, team performance and software quality–A systematic literature review. *Information and Software Technology, 57*, 277-294.

Noesgaard, S. S., & Ørngreen, R. (2015). The Effectiveness of E-Learning: An Explorative and Integrative Review of the Definitions, Methodologies and Factors that Promote e-Learning Effectiveness. *Electronic Journal of e-Learning, 13*(4), pp277-289-pp277-289.

Noguera, I., Guerrero-Roldán, A.-E., & Rodríguez, M. E. (2016). Assuring authorship and authentication across the e-assessment process. In

International Computer Assisted Assessment Conference (pp. 86-92): Springer.

Oates, B. J. (2005). *Researching information systems and computing*: Sage.

Okada, A., Noguera, I., Alexieva, L., Rozeva, A., Kocdar, S., Brouns, F., Ladonlahti, T., Whitelock, D., & Guerrero-Roldán, A. E. (2019). Pedagogical approaches for e-assessment with authentication and authorship verification in Higher Education. *British Journal of Educational Technology, 50*(6), 3264-3282.

Okada, A., Whitelock, D., Holmes, W., & Edwards, C. (2019). e-Authentication for online assessment: A mixed-method study. *British Journal of Educational Technology, 50*(2), 861-875.

Omonijo, D. O. (2012). A Study of E-Cheating Habit of Students in Three Selected Universities in Nigeria. *WUFENIA Journal, 19*(9), 387-402.

Ouf, S., Abd Ellatif, M., Salama, S. E., & Helmy, Y. (2017). A proposed paradigm for smart learning environment based on semantic web. *Computers in Human Behavior, 72*, 796-818.

OWASP. (2021). Threat_Modeling. Retrived from https://owasp.org/www-community/Threat_Modeling.

Palvia, P., Leary, D., Mao, E., Midha, V., Pinjani, P., & Salam, A. F. (2004). Research methodologies in MIS: an update. *The Communications of the Association for Information Systems, 14*(1), 58.

Pettersson, O. (2009). Software Ecosystems and e-Learning: Recent Developments and Future Prospects. In (pp. 64:427-464:431): ACM.

Pettersson, O., Svensson, M., Gil, D., Andersson, J., & Milrad, M. (2010). On the Role of Software Process Modeling in Software Ecosystem Design. In *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume* (pp. 103-110): ACM.

Piotrowski, M. (2011). QTI: A Failed E-Learning Standard? In L. Fotis, G. Steve & P. Elaine (Eds.), *Handbook of Research on E-Learning Standards and Interoperability: Frameworks and Issues* (pp. 59-82). Hershey, PA, USA: IGI Global.

Pitts, M. G., & Browne, G. J. (2007). Improving requirements elicitation: an empirical investigation of procedural prompts. *Information Systems Journal, 17*(1), 89-110.

Queirós, R., Leal, J. P., & Paiva, J. C. (2016). Integrating Rich Learning Applications in LMS. In (pp. 381-386). Singapore: Springer Singapore.

Robinson, D. G., & Halderman, J. A. (2011). Ethical issues in e-voting security analysis. In *International Conference on Financial Cryptography and Data Security* (pp. 119-130): Springer.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*: Wiley-Blackwell.

Roman, G.-C. (1985). A taxonomy of current issues in requirements engineering. *IEEE Annals of the History of Computing, 18*(04), 14-23.

Rosmansyah, Y., Hendarto, I., & Pratama, D. (2020). Impersonation Attack-Defense Tree. *International Journal of Emerging Technologies in Learning (iJET), 15*(19), 239-246.

Rosmansyah, Y., Ritonga, M., & Hardi, A. (2019). An Attack-Defense Tree on e-Exam System. *International Journal of Emerging Technologies in Learning (iJET), 14*(23), 251-260.

Ross, J., & Macleod, H. (2018). Surveillance,(dis) trust and teaching with plagiarism detection technology. In *Proceedings of the 10th international conference on networked learning*.

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering, 14*(2), 131-164.

Sadi, M. H., & Yu, E. (2017). Modeling and analyzing openness trade-offs in software platforms: A goal-oriented approach. In *Grünbacher P., Perini A. (eds) Requirements Engineering: Foundation for Software Quality, REFSQ* (Vol. 10153 LNCS, pp. 33-49).

Saini, S., Grispos, G., Liu, C. Z., & Choo, K.-K. R. (2017). Back to Pen and Paper: Recovering Assessment Questions from Computer-Based Examination Applications.

Saini, V., Duan, Q., & Paruchuri, V. (2008). Threat modeling using attack trees. *Journal of Computing Sciences in Colleges, 23*(4), 124-131.

Sanders, K., Ahmadzadeh, M., Clear, T., Edwards, S. H., Goldweber, M., Johnson, C., Lister, R., McCartney, R., Patitsas, E., & Spacco, J. (2013). The Canterbury QuestionBank: building a repository of multiple-choice CS1 and CS2 questions. In *Proceedings of the ITiCSE working group reports conference on Innovation and technology in computer science education-working group reports* (pp. 33–52). Canterbury, England, United Kingdom: Association for Computing Machinery.

Santos, R. P. d. (2014). ReuseSEEM: an approach to support the definition, modeling, and analysis of software ecosystems. In *Companion Proceedings of the 36th International Conference on Software Engineering* (pp. 650-653). Hyderabad, India: ACM.

Santos, R. P. d., & Werner, C. M. L. (2012). ReuseECOS: An Approach to Support Global Software Development through Software Ecosystems. In *2012 IEEE Seventh International Conference on Global Software Engineering Workshops (ICGSEW)* (pp. 60-65).

Santos, R. P. d., & Werner, C. M. L. (2012). ReuseECOS: An Approach to Support Global Software Development through Software Ecosystems. In *2012 IEEE Seventh International Conference on Global Software Engineering Workshops* (pp. 60-65).

Scacchi, W., & Alspaugh, T. A. (2012a). Designing Secure Systems Based on Open Architectures with Open Source and Closed Source Components. In I. Hammouda, B. Lundell, T. Mikkonen & W. Scacchi (Eds.), *International Conference on Open Source Systems* (pp. 144-159): Springer Berlin Heidelberg.

Scacchi, W., & Alspaugh, T. A. (2012b). Understanding the role of licenses and evolution in open architecture software ecosystems. *Journal of Systems and Software, 85*(7), 1479-1494.

Scandariato, R., Wuyts, K., & Joosen, W. (2015). A descriptive study of Microsoft's threat modeling technique. *Requirements Engineering, 20*(2), 163-180.

Schneider, D. (2014). Safe Exam Browser 2.0 How To (install, configure, deploy and use SEB 2.0). Retrived from http://safeexambrowser.org/presentations/HowTo_SEB2.0.pdf.

Schneier, B. (1999). Attack trees. *Dr. Dobb's journal, 24*(12), 21-29.

Schultis, K.-B., Elsner, C., & Lohmann, D. (2014). Architecture Challenges for Internal Software Ecosystems: A Large-scale Industry Case Study. In *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering* (pp. 542-552): ACM.

Schultis, K. B., Elsner, C., & Lohmann, D. (2013). Moving towards industrial software ecosystems: Are our software architectures fit for the future? In *2013 4th International Workshop on Product Line Approaches in Software Engineering (PLEASE)* (pp. 9-12).

Schultze, U. (2015). Skirting SLR's language trap: reframing the 'systematic' vs 'traditional' literature review opposition as a continuum. *Journal of Information Technology, 30*(2), 180-184.

Sclater, N. (2007). The Demise of eAssessment Interoperability? *WWWrong, 317*.

Selander, L., Henfridsson, O., & Svahn, F. (2013). Capability search and redeem across digital ecosystems. *Journal of Information Technology, 28*(3), 183-197.

SelectSurvey. (2008). SelectSurvey .NET. Retrived from http://selectsurvey.net/.

Selwyn, N. (2014). *Digital technology and the contemporary university: Degrees of digitization*: Routledge.

Severance, C., Hanss, T., & Hardin, J. (2010). Ims learning tools interoperability: Enabling a mash-up approach to teaching and learning tools. *Technology, Instruction, Cognition and Learning, 7*(3-4), 245-262.

Sheard, J., & Dick, M. (2011). Computing student practices of cheating and plagiarism: a decade of change. In *Proceedings of the 16th annual joint conference on Innovation and technology in computer science education* (pp. 233–237). Darmstadt, Germany: Association for Computing Machinery.

Sheard, J., Dick, M., Markham, S., Macdonald, I., & Walsh, M. (2002). Cheating and plagiarism: Perceptions and practices of first year IT students. In *Proceedings of the 7th annual conference on Innovation and technology in computer science education* (pp. 183-187).

Shostack, A. (2014). *Threat modeling: Designing for security*: John Wiley & Sons.

Sim, G., Holifield, P., & Brown, M. (2004). Implementation of computer assisted assessment: lessons from the literature. *ALT-J, 12*(3), 215-229.

Sindre, G., & Opdahl, A. L. (2005). Eliciting Security Requirements with Misuse Cases. *Requirements Engineering, 10*, 34-44.

Sindre, G., & Vegendla, A. (2015). E-exams versus paper exams: A comparative analysis of cheating-related security threats and countermeasures. In *Norwegian Information Security Conference (NISK)* (Vol. 8, pp. 34-45).

Soltani, M., & Knauss, E. (2015). Cross-organizational challenges of requirements engineering in the AUTOSAR Ecosystem: An exploratory case study. In *2015 IEEE Fifth International Workshop on Empirical Requirements Engineering (EmpiRE)* (pp. 41-48).

SPSS, I. (2017). IBM SPSS statistics for Windows, version 25.0. Retrived from https://www.ibm.com/support/pages/node/618179#en.

Stockley, D., & Balkwill, L.-L. (2013). Raising awareness of research ethics in SoTL: The role of educational developers. *Canadian Journal for the Scholarship of Teaching and Learning, 4*(1), 7.

Striewe, M. (2019). Components and Design Alternatives in E-Assessment Systems. In (pp. 220-228). Cham: Springer International Publishing.

Tang, A. (2014). A guide to penetration testing. *Network Security, 2014*(8), 8-11.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British journal of educational studies, 53*(4), 466-478.

Teixeira, A. A., & Rocha, M. F. (2010). Cheating by economics and business undergraduate students: an exploratory international assessment. *Higher Education, 59*(6), 663-701.

Trost, K. (2009). Psst, have you ever cheated? A study of academic dishonesty in Sweden. *Assessment & Evaluation in Higher Education, 34*(4), 367-376.

Tuli, F. (2010). The basis of distinction between qualitative and quantitative research in social science: Reflection on ontological, epistemological and methodological perspectives. *Ethiopian Journal of Education and Sciences, 6*(1).

Twining, P. (2010). Educational information technology research methodology: looking back and moving forward. In *Researching IT in Education* (pp. 169-184): Routledge.

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education, 106*, A1-A9.

Uden, L., Wangsa, I. T., & Damiani, E. (2007). The future of E-learning: E-learning ecosystem. In *Digital EcoSystems and Technologies Conference, 2007. DEST '07. Inaugural IEEE-IES* (pp. 113-117).

Ullrich, M., Forell, M., Houy, C., Pfeiffer, P., Schüler, S., Stottrop, T., Willems, B., Fettke, P., & Oberweis, A. (2021). Platform Architecture for the Diagram Assessment Domain. In *Software Engineering*.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory'back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal, 20*(4), 357-381.

Valenca, G. (2013). Requirements negotiation model: A social oriented approach for software ecosystems evolution. In *Requirements Engineering Conference (RE), 2013 21st IEEE International* (pp. 393-396).

Valença, G., Carina, A., Virgínia, H., Slinger, J., & Sjaak, B. (2014). Competition and collaboration in requirements engineering: A case study of an emerging software ecosystem. In *Requirements Engineering Conference (RE), 2014 IEEE 22nd International* (pp. 384-393).

Van den Berk, I., Jansen, S., & Luinenburg, L. (2010a). Software Ecosystems: A Software Ecosystem Strategy Assessment Model. In (pp. 127-134): ACM.

Van den Berk, I., Jansen, S., & Luinenburg, L. (2010b). Software ecosystems: a software ecosystem strategy assessment model. In *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume* (pp. 127-134). Copenhagen, Denmark: ACM.

Vegendla, A., Nguyen-Duc, A., Gao, S., & Sindre, G. (2018). A Systematic Mapping Study on Requirements Engineering in Software Ecosystems. *Journal of Information Technology Research (JITR), 11*(1), 49-69.

Vegendla, A., & Sindre, G. (2019). Mitigation of Cheating in Online Exams: Strengths and Limitations of Biometric Authentication. In A. V. S. Kumar (Ed.), *Biometric Authentication in Online Learning Environments* (pp. 47-68). Hershey, PA, USA: IGI Global.

Vegendla, A., Søgaard, T. M., & Sindre, G. (2016). Extending HARM to make test cases for penetration testing. In *International Conference on Advanced Information Systems Engineering* (Vol. 249, pp. 254-265): Springer, Cham.

Volante, L. (2004). Teaching to the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy*.

Walden, J., Doyle, M., Lenhof, R., Murray, J., & Plunkett, A. (2010). Impact of Plugins on the Security of Web Applications. In *Proceedings of the 6th International Workshop on Security Measurements and Metrics* (pp. 1:1-1:8). Bolzano, Italy: ACM.

Walker, J. (2010). Measuring plagiarism: Researching what students do, not what they say they do. *Studies in Higher Education, 35*(1), 41-59.

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education, 39*(3), 235-274.

Wibowo, S., Grandhi, S., Chugh, R., & Sawir, E. (2016). A pilot study of an electronic exam system at an Australian University. *Journal of Educational Technology Systems, 45*(1), 5-33.

Williamson, M. H. (2018). Online exams: The need for best practices and overcoming challenges. *The Journal of Public and Professional Sociology, 10*(1), 2.

Wills, G. B., Davis, H. C., Gilbert, L., Hare, J., Howard, Y., Jeyes, S., Millard, D., & Sherratt, R. (2009). Delivery of QTIv2 question types. *Assessment & Evaluation in Higher Education, 34*(3), 353-366.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*: Springer Science & Business Media.

Woldeab, D., & Brothen, T. (2021). Video Surveillance of Online Exam Proctoring: Exam Anxiety and Student Performance.

Yin, R. (2002). *Case Study Research: Design and Methods* (Vol. 5): SAGE Publications, Inc.

Yin, R. K. (2017). *Case study research and applications: Design and methods* (6th ed.): Sage publications.

Yu, E., & Deng, S. (2011). Understanding software ecosystems: A strategic modeling approach. *proc of 3rd IWSECO*, 65-76.

Part II: Research papers

Paper 1:

E-exams versus paper exams: A comparative analysis of cheating-related security threats and countermeasures

Guttorm Sindre, Aparna Vegendla

In Proceedings of Norwegian Information Security Conference (NISK)

E-exams versus paper exams: A comparative analysis of cheating-related security threats and countermeasures

Guttorm Sindre, Aparna Vegendla
Department of Computer and Information Science (IDI), NTNU

Abstract

E-exams can have a lot of advantages over traditional paper-based exams, and if using a BYOD approach (Bring Your Own Device) they can also scale to large classes and peak exam days. At the same time, BYOD adds extra security challenges by using student-controlled equipment. To be viable, BYOD e-exams need not have perfect security, only about the same level of security as paper-based exams have. This article uses attack-defense trees to provide an analysis comparing the threats and countermeasures against cheating at controlled exams with paper-based exams versus BYOD e-exams. The conclusion is that neither has a clear advantage from a security perspective.

1. Introduction

A number of advantages have been reported for computer-based assessments over traditional paper-based examinations [1-7], both in terms of computer support for question development, reduced cost of test distribution and administration, reduced cost of distributing answers to graders, and possible automated support for grading. Additionally, new types of test items (e.g., based on audio, video, 3D engineering models, industrial tools, and interaction) can be enabled, thus giving a test with much higher validity vs. professional work practice.

At the same time, there has been reluctance towards introducing e-exams in many universities. This is for instance the case with our own university (NTNU) where most proctored school exams are still traditional paper exams, except a just beginning practice of having e-exams in a limited number of courses with a limited number of students. Two main concerns towards replacing paper-based exams with e-exams are *scalability* and *security*. Scalability is a major challenge especially if tests are to be performed on university equipment. With a Bring-Your-Own-Device (BYOD) approach, where students use their own portable PCs, scalability will improve, but at the cost of added security challenges since the client device can more easily have been tampered with before the exam to facilitate cheating. However, paper-based exams do not have perfect security either [8-9]. Hence, if e-exams have advantages in other respects they need not have better security than traditional paper-based exams, only a similar level of security. So, it is interesting to compare BYOD e-exams and paper-based exams with respect to cheating-related security. In this paper, we will use Attack-Defense Trees (ADTrees) [10] in this analysis, these are an extension of traditional attack trees [11] where there are defense nodes in addition to attack nodes.

For reasons of space and focus, we limit the scope of our paper in several ways. We do not discuss other aspects than security in any detail, though in another paper we have looked more at didactic issues and process improvement potential of e-exams [12]. The focus on this paper is on the cheating challenge, so we only consider *cheating-related* security threats to exams. For instance, this means that denial of service attacks against exam servers will not be discussed in this paper. A successful DoS thwarting an entire exam would be very serious and necessary to protect against, but it can hardly be categorized as cheating since nobody achieves an unfair grade advantage. Also, we limit our discussion to cheating during the exam, not before or after. The rationale for this choice is that cheating before or after the exam is not so dependent on the choice of

Presented at the Norwegian Information Security Conference (NISK 2015)

paper or PC as the medium for the student's answer. Moreover, we focus on written school exams, not other types of assessment tasks. Some types of cheating are even easier for uncontrolled home exams (e.g., impersonation, undue assistance), but using uncontrolled home exams is an implicit choice not to mitigate such cheating threats, and therefore less interesting to analyze in this context.

With a quick intuitive take, BYOD e-exams may seem obviously less secure than paper exams. In a consultancy report in connection with the ongoing project to digitize the exam process at the NTNU, it is said that it is relatively easy to enforce strict rules to prevent cheating with traditional paper exams, while it is much more difficult for digital exams [13]. Similar views can be found in international academic literature, most precisely stated by Dawson: *"The BYOD eExam is by definition less secure than both pen-and-paper examinations, and examinations held in a computer laboratory, as it has all the vulnerabilities of both environments, as well as some of its own."* [14] (p.7).

In our opinion, this claim is exaggerated, although Dawson is obviously right that e-exams may have several serious cheating threats that paper exams do not have. For instance, the following threats can easily be envisioned related to the PC:

- Electronic communication between candidates, or with assisting outsiders.
- Copy-paste plagiarism of allowed or non-allowed sources.
- Peeking at neighbor answers might be easier due to the upright angle of screens.
- The PC can contain materials or tools not allowed for the exam.
- Bigger amounts of information can be crammed into smaller objects (e.g., memory sticks rather than paper), yielding more effective cheating by object passing, either directly between candidates in the exam room, or by using the restroom as a mailbox.

It is also easy to agree with Dawson that many of these threats are worse for BYOD e-exams than for e-exams using university equipment, because students could more easily have tampered with their own PC (e.g., installing materials or tools not allowed, or rigging the PC to circumvent security functions of the exam system). However, what seems to be ignored in the statement that BYOD e-exams are "by definition" less secure than paper exams, is that e-exams − in addition to introducing new threats − also enable many countermeasures against cheating. Some of these countermeasures are effective not only against the threats particular to BYOD e-exams, but also against a wide variety of traditional cheating threats. This complicates the picture and means that it is not at all obvious that BYOD e-exams will generally be less secure than paper exams. Rather, this will depend on the exact implementation of the paper exam, and of the e-exam.

Our research questions for this paper are as follows:

- RQ1: What cheating threats exist for the typical school exams of Norwegian Universities like the NTNU?
- RQ2: What are the main differences in possible threats and countermeasures for paper exams and e-exams.
- RQ3: What requirements are important for secure BYOD e-exams?

The rest of the paper is structured as follows: Section 2 gives a quick introduction to some countermeasures that are better enabled with e-exams than with paper exams, to justify early on our disagreement with arguments that BYOD e-exams are necessarily less secure. Section 3 then gives a more systematic discussion of cheating threats and countermeasures, comparing paper exams and BYOD e-exams by means of ADTrees. Section 3 then makes a more systematic comparison of threats and countermeasures, to see which risks increase by e-exams and which decrease. Section 4 discusses related work, whereupon section 5 concludes the paper with some ideas for further work.

2. Improved countermeasures in e-exams

A key countermeasure in school exams is the use of proctors to oversee the candidates during the exam and catch candidates who cheat. This is the main defense against cheating in paper exams, and could be used similarly for e-exams. However, research indicates that students may be able to cheat in spite of the presence of proctors [15]. An important insight in security research is defense in depth. For exams this would motivate other barriers towards cheating in addition to proctors, and some of these are much easier to implement with e-exams than with paper exams, as will be argued in the following paragraphs:

Mixed seating. Instead of having students of the same course seated row by row in the same exam room, mixed seating of many different courses could effectively mitigate cheating by close range collaboration (whispering, peeking at neighbor's answers, passing answer sheets or other information objects). For paper exams, large scale mixing is hard due to the increased workload of sorting question sets into room piles before the exam day, and complicated distribution of questions on the day. Mixing would also give teachers a hard time if they need to come to the exam room to clarify issues with the exam questions. For e-exams, distribution of questions and collection of answers can be fully automated. Clarifications to questions could be done online, which not only makes this task easier for the teachers, but also improves fairness as all candidates of an exam could get the same information at the same time.

Non-uniform questions. Identical question sets to all examinees makes collaborative cheating much easier, since it suffices to communicate the answers. For instance, the solution to a 100 item multiple choice test can be communicated with 100 letters A/B/C/D, 50 HEX symbols, or even fewer signs using various compression techniques. This is easily within reach of what can be communicated by SMS, code signals, or simply written on a piece of paper to be dropped in a previously agreed WC trash bin to be picked up by somebody else. Randomizing the order of questions makes such cheating much harder for multiple choice and short answer questions, since it increases the cheaters' communication burden when question information must also be included. Such randomization is easy for e-exams, while much harder for paper exams due to more cumbersome printing and copying of question sets.

Moving calculators and books into the exam system. Calculators and allowed books are well known vulnerabilities for cheaters, who might hide forbidden information in calculators or books, or even transfer information if allowed to share resources. With e-exams, the calculator could be an app provided by the exam system, obliterating brought calculator devices. Allowed written resources could be provided digitally through the exam system, this obliterating the need to bring books. For paper exams, allowed written materials can of course be printed as an attachment to the question set, but this only works in cases where it is just a couple of pages (e.g., some few formulas), otherwise it will be way to expensive in terms of copying costs.

Strict question/answer sequence. If you want to cheat by getting assistance from one or more outsiders, it is often preferable to use as few communications acts as possible, ideally just two. You export the questions just after the start of the exam (e.g. photographing with a smart phone and send as MMS, or smuggle questions to agreed WC trash bin for pick-up by the accomplice). Then you get answers back later (again various options for how to do this). One way to mitigate such cheating is to reveal questions in strict sequence (i.e., Q2 is only shown to the candidate after a no-return response has been submitted to Q1, etc.). Then, cheater and accomplice would need much more frequent communication, which would be riskier. Whether to use this mitigation or not, could however depend also on other factors. For some types of tests,

the teacher might want the students to be able to revisit previous questions and improve their answers for the duration of the exam.

Automated plagiarism checking. With digital answers, tools for plagiarism checking can be used effectively, not just for direct copy paste, but gradually also for various rephrasing tricks. This kind of cheating then becomes much more risky than it were for paper exams.

Biometric authentication. Impersonation - i.e., having somebody else sit the exam for you - is a rare cheating threat, but potentially the most effective of all if uncaught, since even an F candidate can get a perfect A if the impostor has strong subject knowledge. The current approach with student ID cards is insecure in case a candidate knows a willing impostor with a quite similar face. Also, there are services on the internet for buying custom fake ID cards. Biometrics is assumed to give far better security [16]. Of course, this could be employed both for paper and e-exams, but e-exams have an advantage that infrastructure for the authentication is then anyway in place, e.g., using recognition of face, voice, and keystroke dynamics via the same PC that is used for answering the questions. For a paper exam, equipment for such authentication would instead have to be provided by the university and e.g. carried around by the proctors, giving extra cost.

Above we have mentioned six important countermeasures against cheating. All of them are in theory possible also for paper exams, but will be more costly or cumbersome than for e-exams. This should serve as sufficient illustration of our point that BYOD e-exams are not necessarily less secure than paper exams. This should also justify that a more detailed comparative analysis of cheating threats and countermeasures of e-exams and paper exams might be of interest. Such a comparison will be provided in the next section.

3. Detailed comparison of cheating threats and defenses

Cheating can be defined as behavior which is against the regulations of the university or of the particular exam, and which may give some candidates an unfair advantage over others. A detailed treatment of cheating in legal terms is beyond this paper (and these authors), the reader can consult [17] for a discussion related to Norwegian law.

From how-to pages on the web it can be inferred that there is a wide variety of cheating methods, even in invigilated school exams[1], all the way from old-fashioned crib-notes and peeking at the answers of others to high-tech cheating with smart phones and miniature spy equipment. Hetherington & Feldman [18] present a taxonomy of cheating with four categories: individualistic-opportunistic, individualistic-planned, social-active, and social-passive. Björklund & Wenestam list 23 ways of cheating in the appendix of their article [19]. In our paper, we will use a list with somewhat broader categories still thought to cover those of the referenced works. New threats specifically occurring with e-exams are in italics in the list below:

1. *Impersonation*: Having your exam answered by somebody else.
2. *Assistance / collaboration*: Candidates get assistance from other candidates, employees, or outsiders, or collaborate in a way not allowed for the exam.
3. *Plagiarism:* Presenting somebody else's words or ideas as one's own, i.e., without proper referencing and quotation marking.

[1] Some illustrative web pages: http://www.wikihow.com/Cheat-on-a-Test-Using-School-Supplies, http://www.gsm-earpiece.com/howto/tips-on-cheating-exam/, http://bozgo.com/how-to-cheat-on-exams/ http://www.complex.com/pop-culture/2012/09/50-ways-technology-can-help-you-cheat/, http://www.learning-mind.com/7-best-ways-to-cheat-on-exam-without-being-noticed/,

4. *Using aids not allowed for the exam*: Most exams have restrictions on usage of materials (e.g., textbooks) and tools (e.g., types of calculators).
5. *Timing violation*: The candidate starts to work on the exam before allowed, or continues to work after the exam inspector has declared that time is out.
6. *Lying to proctors* to achieve some favorable outcome. One example of a favorable outcome could be to get extra time on the exam or leniency in grading, due to a claimed (but not real) problem with the exam. Another example could be to have an exam attempt cancelled rather than failed, for instance by faking disease during the examination.
7. *Smuggling out the exam questions after the exam*. Some universities consider this a serious offense, typically because the same questions may be reused in subsequent tests. In Norway, this is seldom the case. At the NTNU previous exams are normally publicly available documents and students are allowed to keep the questions when the exam time is out.

Discussing all these threats in detail will be too much for this one paper. Hence, we decide to drop the least important ones. #7 is not so relevant in Norway. #6 is believed to be rare, and the ability to fake disease is not affected by the choice of paper or e-exam. #5, although relevant, will seldom give a huge advantage in grade, e.g., getting a couple of minutes extra is not likely to help much if you were not able to answer questions during the 3 hours that the exam really lasted. Finally, plagiarism (#3) tends to be more relevant for home exams and term papers than for school exams. Thus, our further analysis here will focus on the three threats of impersonation, assistance / collaboration, and using aids not allowed for the exam, to be analyzed in the next subsections.

3.1 Impersonation

An attack-defense tree for impersonation is shown in Figure 1. The attack nodes (red ovals) indicate what a cheater tries to achieve. In this particular diagram, all decompositions of attacks into sub-attacks are OR-relationships, as no arc is connecting the lines. Laptop and paper icons are not part of the original ADTree notation but are used for illustration here. Having such an icon placed next to a green rectangular defense node indicates a potential advantage for that type of exam, i.e., the defense is more feasible. Having it next to a red node indicates a disadvantage: that type of exam is more vulnerable to the particular attack. Hence, it indicates two main impersonation attacks:

Since the rightmost branch ("Label swapping") is the simplest, we discuss this first. This would entail that two students collaborate, and each identify correctly as themselves. However, upon delivery of the answers, candidate X instead labels his answer with Y's candidate number, and vice versa. Mitigations would be either than the proctor checks that the correct number has been written, or that even the proctor writes the number (the candidate not knowing it in advance). For an e-exam, explicit labeling might not at all be needed, as this could be done automatically and internally in the system based on the authenticated identity of each candidate.

To prevent spoofing one must verify the candidate's identity. The traditional approach is to require a picture ID of the candidate, but this is vulnerable both to look-alikes and fake ID cards (left sub-tree). With a BYOD e-exam it would be possible to authenticate the candidate by username + password instead. However, username + password would be much less secure than even the picture ID, since the candidate could simply give his login credentials to the impostor before the exam. Hence, this option is crossed out (this cross not a part of the original ADTree notation, but used for

illustration purposes here). Biometrics seems to be the better choice, and as argued earlier e-exams can enable this more cheaply. However, like passwords, biometrics can also be vulnerable to replay attacks. With proctor-provided equipment, the only feasible attack would be an external replay (e.g., wearing a fake fingerprint, pretending to speak into a device for voice-recognition while really playing a recording or live streaming of the real candidate from a small hidden device, holding a picture of the real candidate's face in front of the camera). From these examples, it seems that face recognition would be the most secure of these three options. It is easy to produce fake fingerprints, but very hard to produce masks to fool face recognition, and it is harder to hold up a picture unnoticed by the proctors than to play audio from a hidden device.

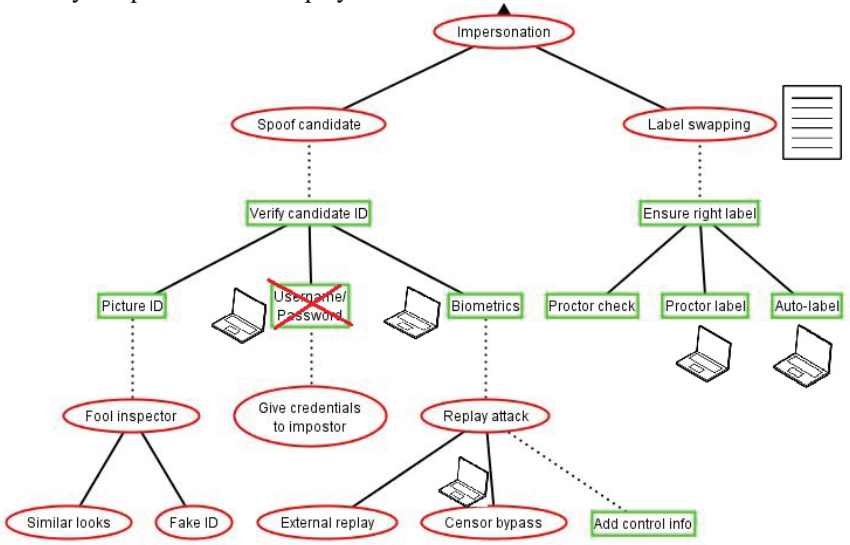

Figure 1 ADTree for Impersonation, made with the ADTool [20]

Using the BYOD laptop, which the candidate and impostor might have rigged for the purpose beforehand, an additional attack of censor bypass might be possible (e.g., the laptop pretends to be sending the server live video of the examinee from its camera, while it really streams recorded video from a file). Thus any advantage of cheaper biometrics for BYOD e-exams might be partially dissipated by the additional attack available. However, both types of replay attacks can be mitigated by the same means, namely adding some control information not known to the examinees beforehand. Hence, if using voice recognition, you would not only demand candidates to state their names (which could easily be pre-recorded) but also to add a phrase provided by the proctors just at authentication time. If using face recognition, you might demand recognition of the background of the exam room, or of a proctor-provided visual cue, in addition to the examinee's face. Hence, the threat of censor bypass does not introduce the need for a lot of new defenses. All in all, therefore, e-exams seems to have the potential to be more secure than paper exams versus impersonation because they can easily mitigate the Label swapping threat, and may offer biometrics at lower cost than what is the case for paper exams.

3.2 Assistance / collaboration, and Unallowed aids

While impersonation might be rare, collaboration and usage of forbidden materials are much more common cheating practices. As argued in Section 2, e-exams can offer a

lot of countermeasures which are practically infeasible for paper exams, and some of these countermeasures are effective against several different types of cheating, both collaborative (getting assistance from other examinees or outsiders) and individual (e.g., using unallowed materials during the exam). An ADTree for Assistance / Collaboration is shown in Figure 2, and an ADTree for Unallowed Aids is shown in Figure 3. As can be seen from these figures, the laptop is associated with a number of countermeasures already mentioned, like Mixed seating, Non-uniform questions, and enforcing a Strict Q/A sequence. These all contribute to mitigating attacks related to traditional in-room communication between candidates, as well as peeking at neighbor answers (which might be done even if the other does not knowingly collaborate). Non-uniform questions also mitigates collaboration via the toilet, since answers hidden there by one candidate might not be relevant to an accomplice retrieving them. A strict Q/A sequence mitigates both collaborative and individual cheating (e.g. hidden resources) in the toilet, as a candidate would need to make suspiciously many toilet visits, whereas with a paper exam one visit to the restroom could be used to check many questions.

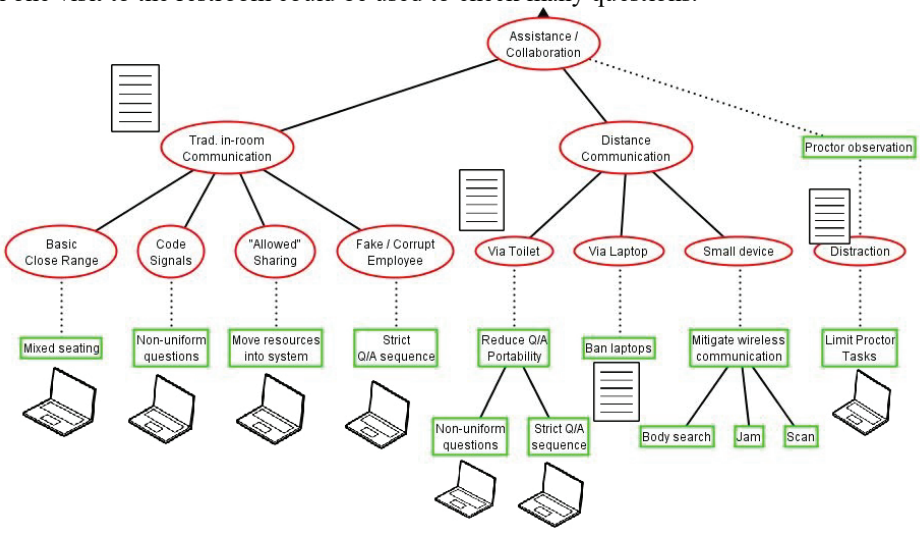

Figure 2 ADTree for Assistance / Collaboration.

Looking specifically at Figure 3, we have already touched upon peeking (leftmost subtree) and cheating in WC (middle). In the exam room, cheating can otherwise be done with concealed aids (e.g., small, hidden cheat note) or with aids that can safely be put on the desk. For concealed aids, it is hard to see any advantage for either type of exam. However, there might be a difference for desk aids. In a paper exam, a cheat sheet which is or looks sufficiently similar to an official drafting sheet for exam usage, can safely be put on the desk as long as the cheater waits some necessary minutes after the start of the exam time, so that the contents is something he could plausibly have written after the exam started. Then, the only really risky moment for the cheater is the second he spends pulling the sheet from a hidden location (e.g., under T-shirt) and placing it on the desk. Thereafter it can safely reside on the desk and help him throughout the entire exam. Having a paperless exam, however, would suddenly make the possession of such a sheet risky all way through. This could be possible with BYOD e-exams, though not all, as the candidates may need to use paper for drafting. For instance, in a math exam with multiple choice questions, candidates would likely need

to make the calculations to know which answer alternative would be the correct one. Hence, the paperless exam mitigation would only be reasonable in courses where no drafting would be needed, or where drafting could best be done on the PC anyway.

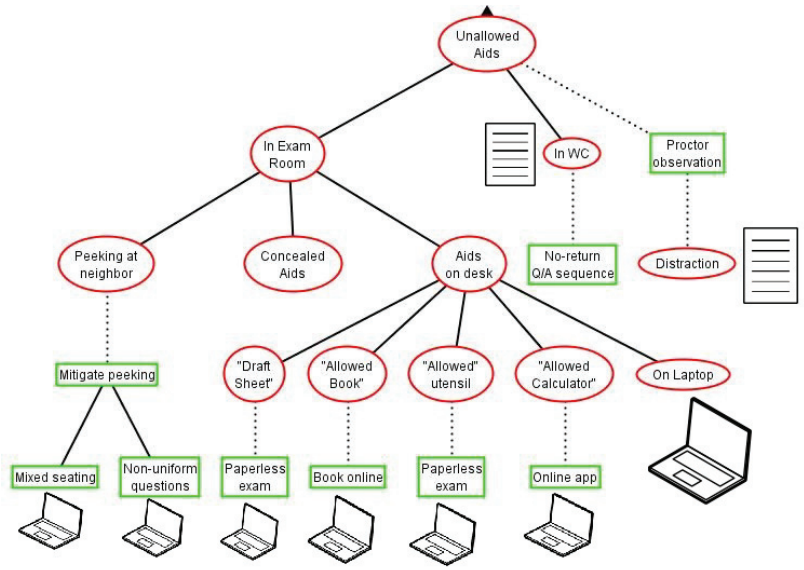

Figure 3 ADTree for Unallowed aids

In addition to the possible advantage of a paperless exam that cheat sheets are no longer plausible items to have on the desk, it also means that a number of other utensils like pencils, erasers, rulers etc. become obsolete, and there are known cheating methods related to all of these. Doing away with calculators and brought books opens up a further advantage, both versus individual cheating (as these resources may also contain other hidden material) and collaborative cheating (if a proctor allows two candidates to share such a resource, for instance because one had forgotten his book or suffered a calculator malfunction). Another potential advantage of e-exams is that they limit proctor tasks. In paper exams, proctors are responsible for supplying candidates with more answer sheets, collecting answers when candidates deliver, and requesting the presence of the teacher if candidates suspect errors in the question set. With e-exams the latter task can go online, and the two former disappear, hence proctors can concentrate better on their primary task of detecting cheating.

The Fake/Corrupt employee threat (Fig 2) is assumed to be rare, though one cannot totally exclude collusion between employees and candidates for reasons such as bribery, blackmail, or relationships (family, love, cults…). The Strict Q/A sequence countermeasure does not prevent a fake employee (e.g., posing as a teacher) or similarly bribed one to come to the candidate's desk, but it makes the attack less effective. If the candidate can only get help with the one question he is currently at, the accomplice might need to make many visits to the candidate's desk for the help to be substantial in terms of grade improvement. This would make the operation much more suspicious.

Indeed, strict Q/A sequence would also mitigate some threats of communication via laptop or small wireless devices. Sending questions to an accomplice early on (email with question file, MMS with photo of questions,…) and getting answers back later would be hard, instead one would be left with the need to communicate more frequently which would likely be riskier. To prevent against continuous communication between

candidates, or with outsiders (e.g., using a background Skype call with shared screen as exemplified by Dawson [14], or using a wireless mini camera to export questions, a wireless earplug to import answers, both connected to a small GSM box to facilitate communication with a person outside the exam room), one must either rely on prevention or detection. For the laptop, prevention approaches would typically include forcing the candidates to boot using a proctor-provided USB memory stick, and/or using a lockdown browser, so that candidates are prevented from opening files or programs not allowed during the exam. Detection approaches might include monitoring of the screens of the laptop (e.g., by taking screen shots of student laptops at regular intervals, as is done at the University of Southern Denmark (SDU)[2], and by installing cameras in the exam room), as well as the monitoring of laptop audio (to prevent usage of the laptop for e.g. background Skype conversations), keystrokes (e.g., to discover if a large chunk of text suddenly finds its way into the answer by other means than being typed by the candidate) and running programs. Also, one could monitor the network traffic related to each PC, as for instance a background skype call or sending questions to an outsider would have a quite different communication profile from the normal question/answer work through the e-exam tool.

Of course, there may be legal implications of electronic surveillance of the students during the exam, and there are several questions concerning this that are unclear in Norwegian law [21]. Hence, although a certain kind of surveillance is used at the SDU or other places, this does not guarantee that the same type of surveillance is possible in Norway. If using their own laptops for the exam, students of course also have a lot of personal information on these devices, possibly also of a sensitive nature. So it could expected that, e.g., the exam system taking a copy of all files on every student's PC, would not be allowed. From the web page of the University of Southern Denmark[2] it seems that the system only inspects active files and programs and communication in/out from the PC, but not passive files on the computer. A legal discussion whether this would be legal in Norway is beyond the scope of this paper, but will likely be clarified if a Norwegian university takes a similar approach. The legal argument in favor of some surveillance would be that, unlike traditional cheating, electronic activity inside the laptops cannot possibly be seen by proctors, so there needs to be some other kind of mitigation in place to avoid easy cheating.

As is suggested by the attack node "Small device" second to the right in Figure 2, the laptop is not the only possible means for electronic communication among candidates or with outsiders. Cell phones and other small devices can also be used. To mitigate against this, one could try to jam such communications if this is legal (prevention), or to scan for them (detection). Although a deep comparative analysis of the PC threat versus the small devices threat is beyond the scope of this paper, it seems that using the laptop for cheating might in many ways be riskier than using small devices. There are more effective monitoring approaches available for the laptop, especially if collaborative cheating is attempted through the established network. With small devices it is more difficult to discover, and even if some bluetooth, radio, or other signals are discovered, it is hard to know exactly what candidate it came from.

4. Related work

Sheard et al. [22] presents the survey on cheating and plagiarism with undergraduate IT students in two universities. The survey was conducted by a questionnaire based on 18 scenarios. The students were asked to rate their perception on

[2] https://em.sdu.dk/

scenarios on a 5-point Likert scale. The survey was performed on cheating, but without any comparison of paper vs. e-exams. Hillier [23] has made a questionnaire survey comparing student perceptions of e-exams versus paper exams. The study is not particularly about security, but touches upon security, too. The students on average thought that e-exams would be less secure than paper exams, but the difference was not massive. King et al. [24] made a survey with 121 business students, where students believed it was easier to cheat with online exams.

Sclater and Howie [7] present a number of requirements to e-exam systems, a couple of these are security requirements. Other works have looked at specific protection mechanisms such as cryptographic schemes [25] and security protocols [26].

Clariana and Wallace [27] compared paper based versus computer based assessments, but not particularly security. Instead they focused on student performance, and concluded that higher attaining students benefit more from computer-based assessments compared to paper-based ones. Another comparison by Jamil et al. [28] focuses on teachers' perceptions, concluded that attitude among their informants was quite positive towards computer-based examinations, though in some cases, teachers preferred paper as well.

Except for the quite brief and crude statement by Dawson that BYOD e-exams are less secure than paper exams [14], and some surveys and general investigations touching upon security issues, this paper is to our knowledge the first more detailed comparison of cheating-related security and countermeasures of paper vs. e-exams.

5. Discussion and conclusion

If BYOD e-exams fail to utilize the several new countermeasures that become feasible in a paper-less process, one could agree with Dawson that BYOD e-exams would by definition be less secure than paper-based exams. Even if using lock-down browsers like Safe Exam Browser[3], as well as monitoring the screens, keyboards, camera, audio, and network traffic of the laptops during the exam, one must imagine that some students may be able to perform cheats that go under the radar − for instance by having installed software on their laptops to conceal illegal activities as legal ones.

On the other hand, the case for e-exams is that they offer a wide range of countermeasures that mitigate a number of cheats, such as mixed seating, non-uniform questions, strict Q/A sequence, and reducing the need for brought items. Hence, a number of known cheating techniques are suddenly much less effective. Moreover, for some types laptop-enabled cheating, such as electronic communication between candidates or with outsiders, the laptop might not be the device of choice for the potential cheater. It is much easier for the exam organizers to monitor a laptop which they know is there, and which is hooked up with their network, than to monitor the usage of small devices which they do not know about. Also, if detecting suspicious network traffic from a laptop, one can quickly determine which candidate it is associated with, but this might be much more difficult if the candidate is using some type of wireless communication from a miniature device. Hence, even for the new threats that it brings, the laptop might not be the weakest link in the chain, as spy equipment like wireless earplugs and micro cameras are becoming ever smaller and cheaper, and partly advertised specifically as exam cheating equipment.

To avoid becoming less secure than paper-based exams, it is however important that BYOD e-exams utilize the countermeasures it has as its disposal. Hence, exam organizers should utilize mixed seating whenever feasible. If teachers are required to

[3] http://safeexambrowser.org/news_en.html

provide clarifications to question on the exam day, the exam system must support requests and clarifications online. Otherwise, mixed seating is difficult. Online clarifications also improve fairness, as all get the same information at the same time. Non-uniform questions should be utilized whenever there is otherwise a big cheating threat in answers that can be quickly communicated or seen by peeking at neighbors (e.g., multiple choice or short answers). Calculators, dictionaries and other allowed books should as much as possible be provided as digital resources in the exam system rather than as traditional brought objects, as this reduces some well-known cheating vulnerabilities. A strict Q/A sequence might feel inappropriate in some types of exams, but if it is used, it certainly makes several types of cheating much harder, especially related to communication among candidates or with outsiders (via electronic equipment, or using the restroom as a mailbox). Hence, digital exam tools should have a setting for the teacher to decide whether questions must be answered in strict sequence or not.

As a final conclusion, this paper of course has a number of limitations. It is far from a complete and systematic treatment of every possible cheating threat, and there is little technical detail about particular threats relating to student laptops. Yet, it has pointed out some possible countermeasures that e-exams have against cheating which are infeasible for paper exams. While this certainly does not show that BYOD e-exams are more secure than paper exams, but at least demonstrates that they need not be less secure, as the level of security will depend on the actual implementation of each exam type, what countermeasures are in place, the skills of the proctors, and the types of questions asked on the exam. Further work is needed to arrive at precise security requirements for e-exam systems, and this would have to include both technical requirements for the e-exam tool and infrastructure, and organizational requirements concerning training and awareness of proctors to handle this new mode of examination.

References

1. Conole, G. and B. Warburton, *A review of computer-assisted assessment.* Research in Learning Technology, 2005. **13**(1).
2. Sim, G., P. Holifield, and M. Brown, *Implementation of computer assisted assessment: lessons from the literature.* Research in Learning Technology, 2004. **12**(3).
3. Csapó, B., J. Ainley, R.E. Bennett, T. Latour, and N. Law, *Technological issues for computer-based assessment*, in *Assessment and teaching of 21st century skills*. 2012, Springer. p. 143-230.
4. Gipps, C.V., *What is the role for ICT-based assessment in universities?* Studies in Higher Education, 2005. **30**(2): p. 171-180.
5. Kuo, C.-Y. and H.-K. Wu, *Toward an integrated model for designing assessment systems: An analysis of the current status of computer-based assessments in science.* Computers & Education, 2013. **68**: p. 388-403.
6. Terzis, V. and A.A. Economides, *The acceptance and use of computer based assessment.* Computers & Education, 2011. **56**(4): p. 1032-1044.
7. Sclater, N. and K. Howie, *User requirements of the "ultimate" online assessment engine.* Computers & Education, 2003. **40**(3): p. 285-306.
8. McCabe, D.L., L.K. Trevino, and K.D. Butterfield, *Cheating in academic institutions: A decade of research.* Ethics &Behavior, 2001. **11**(3): p. 219-232.
9. Dick, M., J. Sheard, C. Bareiss, J. Carter, D. Joyce, T. Harding, and C. Laxer. *Addressing student cheating: definitions and solutions*. in *ACM SigCSE Bulletin*. 2002: ACM.

10. Kordy, B., S. Mauw, S. Radomirović, and P. Schweitzer, *Attack–defense trees.* Journal of Logic and Computation, 2012: p. exs029.
11. Schneier, B., *Secrets and Lies: Digital Security in a Networked World*. 2000, Indianapolis: Wiley.
12. Sindre, G. and A. Vegendla, *E-exams and exam process improvement*, in *UDIT 2015*. 2015, Bibsys OJS: Ålesund.
13. Hovde, P. and S.O. Olsen, *Utredning - Digital eksamen NTNU 2015-2019*. 2015, NTNU: Trondheim, Norway.
14. Dawson, P., *Five ways to hack and cheat with bring-your-own-device electronic examinations.* British Journal of Educational Technology, 2015.
15. Franklyn-Stokes, A. and S.E. Newstead, *Undergraduate cheating: Who does what and why?* Studies in Higher Education, 1995. **20**(2): p. 159-172.
16. Levy, Y. and M. Ramim. *A theoretical approach for biometrics authentication of e-exams*. in *Chais Conference on Instructional Technologies Research* 2007. Raanana, Israel: The Open University of Israel.
17. Unneberg, I., *Eksamensfusk ved universiteter og høyskoler.* Lov og Rett, 2012. **51**(8): p. 491-510.
18. Hetherington, E.M. and S.E. Feldman, *College cheating as a function of subject and situational variables.* Journal of Educational Psychology, 1964. **55**(4): p. 212.
19. Bjorklund, M. and C.-G. Wenestam, *Academic cheating: frequency, methods, and causes*, in *European Conference on Educational Research*. 1999: Lahti, Finland.
20. Kordy, B., P. Kordy, S. Mauw, and P. Schweitzer, *ADTool: security analysis with attack–defense trees*, in *Quantitative Evaluation of Systems*. 2013, Springer. p. 173-176.
21. Hilding, G., M. Seim, M.L.B. Tho, M. Felton, M.M. Jegersberg, L. Tande, A.G. Fjeldstad, K. Møller, J. Morland, and K. Veium, *Digital vurdering og eksamen - en juridisk vurdering*. 2014.
22. Sheard, J., M. Dick, S. Markham, I. Macdonald, and M. Walsh. *Cheating and plagiarism: perceptions and practices of first year IT students*. in *ACM SIGCSE Bulletin*. 2002: ACM.
23. Hillier, M. *The very idea of e-Exams: student (pre) conceptions*. in *Australasian Society for Computers in Learning in Tertiary Education Conference*. 2014: ascilite.
24. King, C.G., R.W. Guyette Jr, and C. Piotrowski, *Online Exams and Cheating: An Empirical Analysis of Business Students' Views.* Journal of Educators Online, 2009. **6**(1): p. n1.
25. Castella-Roca, J., J. Herrera-Joancomarti, and A. Dorca-Josa. *A secure e-exam management system*. in *Availability, Reliability and Security, 2006. ARES 2006. The First International Conference on*. 2006: IEEE.
26. Giustolisi, R., G. Lenzini, and P.Y. Ryan, *Remark!: A secure protocol for remote exams*, in *Security Protocols XXII*. 2014, Springer. p. 38-48.
27. Clariana, R. and P. Wallace, *Paper–based versus computer–based assessment: key factors associated with the test mode effect.* British Journal of Educational Technology, 2002. **33**(5): p. 593-602.
28. Jamil, M., R. Tariq, and P. Shami, *Computer-Based vs Paper-Based Examinations: Perceptions of University Teachers.* Turkish Online Journal of Educational Technology-TOJET, 2012. **11**(4): p. 371-381.

Paper 2:

Extending HARM to make Test Cases for Penetration Testing

Aparna Vegendla, Thea Marie Søgaard, Guttorm Sindre

In Proceedings of Advanced Information Systems Engineering Workshops: CAiSE

Extending HARM to make Test Cases for Penetration Testing

Aparna Vegendla(✉), Thea Marie Søgaard, and Guttorm Sindre(✉)

Department of Computer and Information Science,
Norwegian University of Science and Technology (NTNU), Trondheim, Norway
{aparnav,guttors}@idi.ntnu.no

Abstract. [Context and motivation] Penetration testing is one key technique for discovering vulnerabilities, so that software can be made more secure. **[Question/problem]** Alignment between modeling techniques used earlier in a project and the development of penetration tests could enable a more systematic approach to such testing, and in some cases also enable creativity. **[Principal ideas/results]** This paper proposes an extension of HARM (Hacker Attack Representation Method) to achieve a systematic approach to penetration test development. **[Contributions]** The paper gives an outline of the approach, illustrated by an e-exam case study.

Keywords: Security · Penetration testing · Misuse cases · Socio-technical systems · e-exams

1 Introduction

The alignment of requirements and testing has been emphasized as an important problem in software development in general [1, 2] and also for security requirements in particular [3], where testing might then be a combination of penetration testing [4] and ethical hacking [5].

Penetration testing is often used for finding security vulnerabilities in software [6]. As observed by [4], it can be effective if combined with security-related findings from earlier lifecycle stages, but less effective if done completely ad hoc. Even with a systematic approach it is important to be aware that there may be other vulnerabilities remaining in addition to those the tests have uncovered [4].

Previously, our research group has been involved in the development of a method called HARM [7], with the purpose of representing hacker attacks in various ways. In the current paper, we explore how this method could be extended to provide a bridge between security requirements and testing. More precisely, our research question is **RQ1:** *How can HARM be extended to support the development of penetration test cases from security requirements?*

The rest of the paper is structured as follows: Sect. 2 provides background on HARM, illustrating the method with a running example related to the case study, as well as discussing related work. Section 3 discusses how HARM can be extended to include manual human attacks in addition to technical attacks, and to support the development of test cases. Section 4 then presents a case study where HARM is used to

J. Krogstie et al. (Eds.): CAiSE 2016, LNBIP 249, pp. 254–265, 2016.
DOI: 10.1007/978-3-319-39564-7_24

capture security requirements, analyze threats and suggest security test cases for a digital exam system. Section 5 concludes the paper and outlines some ideas for further work.

2 Background

2.1 Running Example: BYOD e-exams

Many universities are currently switching from traditional school exams using pen and paper to e-exams, in some cases performed at home (e.g., remote exams), in some cases in a controlled campus environment. For scalability and cost reduction, even the latter type will often require students to use their own laptops (BYOD, Bring Your Own Device), although this gives increased challenges with security [8]. Concentrating here on individual school exams with invigilators, it is typically necessary to ensure the rules/requirements related to cheating security as shown in Table 1.

Table 1. Some rules against cheating during controlled school exams

	Rule/requirement
R1	Only authenticated examinees shall be able to access and respond to exam questions
R2	It shall be possible to respond to exam questions only while seated at one's assigned place in a controlled venue
R3	Examinees are prohibited from communicating with each other or with outsiders during the exam
R4	Examinees are prohibited from using tools or resources other than those listed as being allowed for the specific exam
R5	Examinees are prohibited from peeking at and copying answers of other examinees

Since the focus here is on BYOD, it makes most sense to focus specifically on some key security requirements to prevent cheating via the laptop, such as:

SecR1. It shall be impossible to access other resources on the laptop than those specifically allowed for the exam.

SecR2. It shall be impossible to use the laptop for communication with co-examinees or outsiders during the exam.

A key approach to mitigate cheating with BYOD e-exams is the usage of so-called *lock-down browsers* [9]. By locking the screen in a way that cannot be escaped while connected to the exam server, this technology prevents examinees from starting up other programs, opening documents or accessing other web sites than the exam server. The e-exam application which delivers questions to the students and receives answers will typically be running on top a lockdown browser. By these measures, examinees should be prevented from accessing cheat material and getting illegitimate help from accomplices via their laptops − *if* the technology is a 100 % effective.

However, a number of attacks could circumvent lock-down browsers. One simple example: After starting up the lock-down browser, we may be unable to start up Skype

to communicate with an accomplice. But what if we have Skype running *before* we start the lock-down browser? This is the outset of our running example. The problem of many hands can easily be envisioned here. The invigilators in the exam room − and the university administration, who give instructions for their conduct − might think that skyping via the laptops during the exams is made impossible by some component of the e-exam technology. The developers of the technology might have been thinking that Skype conversations is something that the invigilators should prevent. There could also be dispersion of responsibility between different technology providers. The developers of the e-exam application might believe that the lock-down browser prevents Skype conversations, while the developers of the lock-down browser consider this outside the scope of their tool, rather to be done by the e-exam application or monitoring software that the university should get from yet another vendor.

2.2 HARM (Hacker Attack Representation Method)

HARM [7] is a method for modeling threats and security attacks in combination with the system architecture, so as to better understand the potential attacks. In this section we summarize the method, so that the extensions that will be proposed later will be understandable. HARM combines several different specification formats to give a comprehensive view of the possible attacks. In the following, we will list these and illustrate them by means of our running example.

Attack Sequence Descriptions (ASD): These are simple natural language descriptions of the attack, forming a sequence of actions. An example ASD could be something like *"(1) Start up a Skype call with an outside accomplice, and have it run in the background. (2) Enter the exam venue and begin the exam in the normal way. (3) Communicate questions to the accomplice and get answers back via Skype, using a hidden wireless earpiece. (4) Type the answers into the e-exam system and submit."*

Misuse Sequence Diagrams (MUSD): If preferring a more formal form of expression than the natural language ASD, a similar sequence can be described as a MUSD [10]. This is similar to a UML sequence diagram, but in addition to legitimate objects and message calls, it also contain attacking objects and message calls (having red boxes and red arrows). The diagram in Fig. 1 shows the cheating examinee setting up a Skype call with an accomplice before the start of the exam. Then the examinee starts up the lock-down browser and authenticates with an to get an access code to connect with the exam server. Via the Skype connection, the examinee communicates the questions to an accomplice, and the accomplice replies with answers. The dashed red ovals indicate vulnerabilities that are utilized to make the attack work, and their labels are explained to the right of the diagram.

Misuse Case Maps (MUCM): Like MUSD, MUCM [11] also show an attack sequence. The difference is that Misuse Case Maps put more focus on the relationship between the attack sequence and the architecture, showing each step in its architectural context [12], just like Use Case Maps show how legitimate functionality propagates through the architecture [13]. Figure 2 shows a MUCM for another one of the cheating threats investigated in our study, usage of disallowed material. The naïve approach of

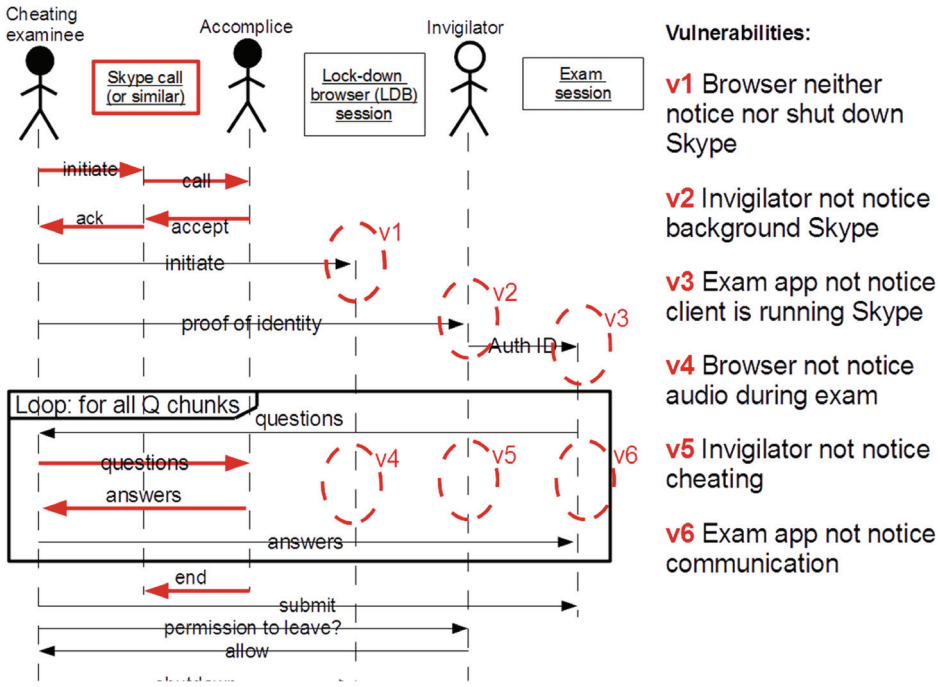

Fig. 1. MUSD for a cheat with pre-connected Skype call (Color figure online)

putting cheat files on the laptop's disk or memory sticks might fail if the lock-down browser prevents the opening of any files during the exam. A more sophisticated approach, as pointed out by Dawson [8], is to use a USB key injector containing the cheat notes. It behaves just like a keyboard, and would thus be unlikely to raise suspicion if there is automated monitoring - as students might be allowed to use external keyboards to their laptops for improved ergonomics of typing a lot of text quickly.

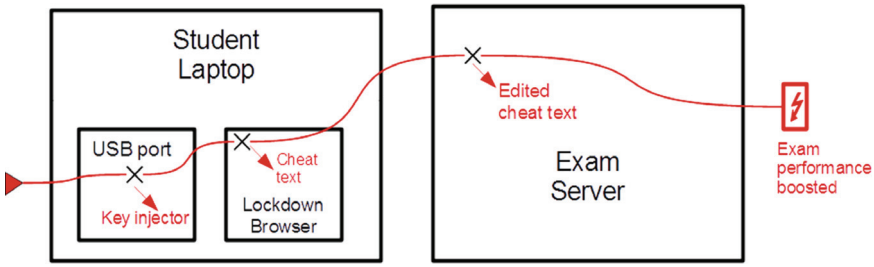

Fig. 2. MUCM for using a key injector with a cheat note

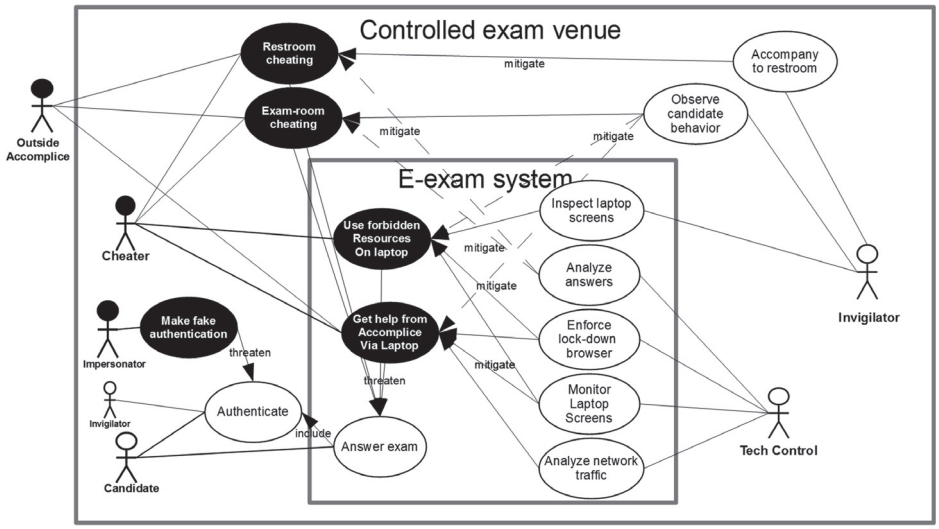

Fig. 3. Misuse case diagram including both electronic and traditional cheating

Misuse Case Diagrams (MUD): MUD extends UML use case diagrams to show how mis-users perform regular as well as irregular activities with the system. Figure 3 shows the MUD for cheating threats studied in our study. Compared to MUSD and MUCM, which show details of one particular type of attack, misuse case diagrams show a broader overview. In the particular diagram in Fig. 3, this overview is made extra broad by showing both the functions and threats particular to the e-exam application (inner system boundary) and cheating threats outside this (e.g., more traditional ways of cheating in the exam room).

Attack Trees (AT): These also show an overview of several threats. Unlike misuse case diagrams, which focus on relationships between threats and legitimate behavior, attack trees focus on the illegitimate behavior alone, breaking high level threats down to more detailed ones. The non-leaf nodes are decomposed into trees of conjunctive ("AND- branch") and disjunctive ("OR-branch") nodes. OR-nodes represent alternatives, while AND nodes represent sub goals where all must be fulfilled to achieve the goal. In Fig. 4, all branches are OR-branches, indicating various ways to perform the high level attack "Cheat during BYOD exams".

3 From Requirements to Penetration Test Cases via HARM

Whereas HARM as illustrated in the previous section has been described in earlier publications, the new contribution of this article is to propose a method to develop penetration test cases aided by HARM. Given some security requirements, like SecR1 and SecR2, there are actually two different approaches that can be used to develop a set of penetration tests:

- **Top down approach:** For each security requirement

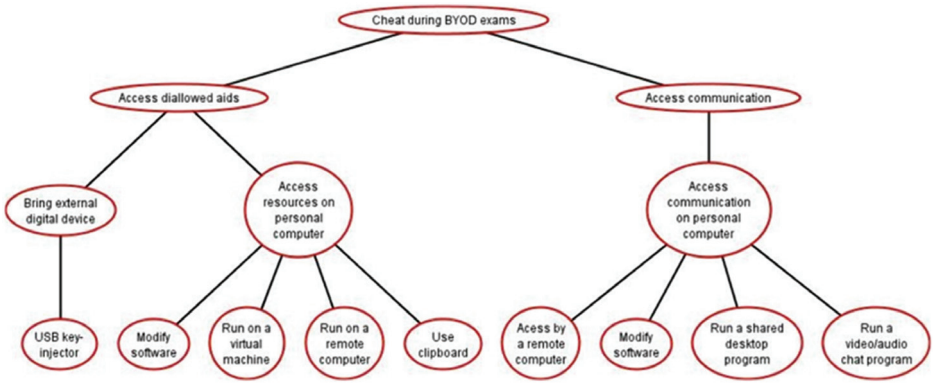

Fig. 4. Attack tree for using a key injector with a cheat note, from [14].

- Make an attack tree, starting with the top level node being a generic violation of that security requirement, then gradually breaking down towards concrete attacks. Brainstorming might be one possible technique to use in developing this tree.
- Make a misuse case diagram relating attacks to relevant legitimate use cases, including mitigations that are known to be in place. This can be used to eliminate from the attack tree those attacks that are not worth trying, or to adjust them to keep them worthwhile. For instance, if one attack is "Open document" with a cheat file during the exam, this should not be possible with the mitigating use case "Enforce lock-down browser" (cf. Fig. 3). So, to keep "Open document" it should have to be in an AND-relation with "Escape lock-down" in the attack tree.
- Make attack sequence descriptions explaining how the attack is going to be executed. If necessary, e.g., to understand a technically complicated attack which can be performed in several different ways, complement the simple textual description of the attack sequence with MUCM (if it is useful to see it in the architectural context) or MUSD (if it is useful to see how the cheat attack propagates via various objects and agents).
- This should be continued until there are attack sequences described for all the leaf nodes of the attack tree.

- **Bottom up approach:** For each security requirement
  - Start with finding some concrete ways of breaking them, and describe these as attack sequence diagrams, possibly also by MUCM and/or MUSD if this is helpful to understand possible attacks and different ways of doing things.
  - When you run out of ideas for concrete attacks, group the similar ones to make the higher level nodes and form the complete attack tree. Make the misuse case diagram to see relationship between attacks and possible countermeasures.
  - It could be a good idea here when the overall attack tree has been formed to work back down in a top down manner, to see if you get any new ideas for possible attacks after seeing the whole picture.

Whatever combination of top-down and bottom-up is chosen, the final step in the planning is to transform the attack sequence descriptions/misuse case maps/misuse sequence diagrams into penetration test scenarios, typically described in tabular form. With a situation similar to the e-exam case, tests would best be developed in two steps:

1. **lab tests**, with the purpose of finding out whether some attack is technically possible or not. Lab tests may investigate small partial attacks one at a time. Table 2 shows a lab penetration test scenario for the cheating via Skype example shown in Fig. 1.
2. **real world tests**, with the purpose of finding out whether attacks are likely to succeed in practice - which may hold bigger challenges than in the relaxed lab setting. Such a test scenario for the Skype example is shown in Table 3.

Since real world tests are more time consuming and expensive than lab tests, it is a good idea to describe the lab tests first. If it turns out that some type of attack was not even possible in the lab, it may be a waste of time to develop a real-life test for it, so resources should rather be spent on other attacks that were more likely of succeeding. (E.g., if we were not even able to have a Skype connection in the lab, there would be little point in trying in the exam-room with the additional challenge of invigilators, etc.). In the planning stage, the rightmost column of Table 2 (Result) would of course be left empty, to be filled in later, while here − to save space, we indicate at once the results that came out of our tests.

Table 2. Penetration test scenario for communicating via Skype

Lab penetration test scenario: *communicate via Skype*			
Step	Action	Success criterion	Result
1	Establish Skype connection between examinee's laptop and accomplice's PC	Connection established	OK
2	Start lock-down browser (SEB) on examinee's laptop	SEB running normally	OK
3	Examinee give info to accomplice	At least one works:	OK
3a	Speak	Accomplice hears	OK
3b	Visual (e.g., blink eyes)	Accomplice sees	OK
3c	Share screen	Accomplice sees	–
4	Accomplice give info to examinee	At least one works:	OK
4a	Speak	Examinee hears	OK
4b	Visual (e.g., blink eyes)	Examinee sees	–
4c	Share screen	Examinee sees	–

It can be noted that the penetration test in Table 2 only explores vulnerability v1 and v4 of the MUSD in Fig. 1, namely those related to the lock-down browser. The other vulnerabilities would be explored in the real-world test as described in Table 3.

Table 3. Test case for cheating during exam through assistance from outsider

Real-world penetration test scenario: *get help during exam via Skype*			
Step	Action	Success criterion	Result
1	Establish Skype connection	Connection established	
2	Start lock-down browser (SEB)	SEB running normally	
3	Authenticate and access e-exam app	E-exam app starting normally	
4	Open exam question	Exam question appearing on screen	
5	Communicate question to accomplice (e.g., quietly speaking w/wireless hidden mic)	Accomplice receives question; No cheating is detected	
6	Receive hints from accomplice (e.g., through wireless earpiece) and type answer into e-exam app	Examinee receives and types info; No cheating is detected	
7	Repeat 4-6 until all questions answered, then submit	Exam answer submitted; No cheating detected	

In Table 3 the Result column is empty because none of the real-world tests have been performed yet. Whereas lab tests will tend to either succeed or fail, the real-world tests will more often have some probability of succeeding. For instance, it may depend on how far the penetration tester is seated from the nearest invigilator, how clever the tester is at speaking so quietly that it is inaudible to others yet comes through clear enough to the accomplice, how good the tester is at appearing calm in spite of cheating, how attentive the invigilator is, and what kind of other mitigations are in place in the exam room, such as monitoring software to discover suspicious communication from laptops, not matching the profile of the typical interaction between the lock-down browser and the e-exam server. Hence, while the lab test in Table 2 may only need to be run once to establish that skyping was actually possible in spite of the lock-down browser, the test in Table 3 would best be run several times, with different testers and invigilators, in rooms with different types of background noise, seated in different positions. This would enable to gather some statistics, like probability of getting caught, or mean time to failure (i.e., getting caught), to rank the attack relative to other attacks to determine which ones are most urgent to deal with.

4 Case-Study: Cheating-Related Exam Security

As part of a student project by the second author (supervised by the first and third author), a number of attacks were tested on a certain lock-down browser, namely Safe Exam Browser [15]. This browser was chosen because it is open source, and because

the e-exam tool that our university is using, partly relies on that browser for security during the exams. It should be noted that the project did not try to cover the complete set of security related to e-exams. The following limitations were chosen:

- only look at threats *during* the exam, not before (e.g., getting premature access to exam questions) or after (e.g., manipulating answers after delivery, or manipulating grades).
- only look at *cheating* threats, not other kinds of security threats (e.g., like sabotage of the exam, denial of service). Although such other threats may also need to be handled, they are not threats that give a grade advantage and thus not classified as cheating.
- due to time and resource limitations, only lab tests were actually executed, while the real-world tests remained at the idea level.

Table 4 sums up results for all the different test cases that were tried in the project. Note that "Success" in the Result column means from the penetration tester's (i.e., attacker's) point of view. From the secure e-exam point of view, then, it is the rows with "Fail" that are the successful ones. So, it can be seen that SEB prevents well against attempts to circumvent it by running on a virtual machine when starting the lockdown browser (if this was not prevented against, the examinee could during the exam shift execution from the virtual to the real machine and then run any forbidden application). It also protects well against attempts to hide cheat text in the clipboard and then try to paste it once the exam has started, and as far as we could find, the examinee would not be able to share her desktop with an accomplice. As the table indicates, however, several other cheating options were available, potentially enabling a candidate with very little subject knowledge to get help from somebody much more clever, in the worst case getting an A where an F would have been the correct account of the examinee's competence. The results of the tests have been communicated to SEB developers, so these weaknesses may likely be mended in future versions of the software. It should also be noted − as pointed out in the previous section - that the success of the four lab attacks in Table 4 does not necessarily mean that the same attacks would be certain to succeed in a real-world exam situation, where there would be a combination of several tools involved, plus human invigilators to oversee the candidates. But some of the attacks do not require much visibly suspicious behavior by the examinee, so could be assumed hard to spot by invigilators.

5 Related Work

Dawson [8] presents five attacks against BYOD e-exams, whereof 4 were tried with various e-exam tools and found successful with at least one tool each. Some of the attacks tried out in our work are inspired by his proposals, especially the key injector attack and the Skype call attack. Dawson, however, does not present any modeling approach or other systematic approach to get from requirements to a test plan.

Cota et al. [16] proposed a framework, RACOON, which is a semi-automatic approach to configure accountability mechanisms (e.g. logging, auditing, monitoring) and reputation mechanisms on the P2P systems. The accountability mechanism helps to

monitor cheating whereas the reputation mechanism helps to punish in case of cheating. The paper also discussed the approach to find cheating in the systems through game based simulations using game theory. Although the approach discussed in their paper useful to find cheating in digital exams, the details of penetration tests were not provided in the paper, which is the main consideration for our paper.

Table 4. Tests completed in the project [14] so far

Attack	Result	Description
Inject notes into exam software with USB key injector	Success	We saved a text on a rubber ducky USB and the string was injected into the web page open in SEB
Run SEB on a virtual machine	Fail	When initiating SEB, a pop-up window appears, stating that SEB has detected a virtual machine and will not work
Run SEB on a remote computer	Success	We managed to control SEB from a remote computer, while using SEB
Use clipboard to import notes into exam software	Fail	We were not able to right click or use CTRL + P to paste the clipboard content into SEB
Get assistance by being accessed from a remote computer	Success	We managed to control and access an SEB exam environment from a remote computer
Get assistance by sharing desktop	Fail	Neither Google Hangout nor Skype showed SEB with remote desktop, when it was initiated
Get assistance by communicating with audio/video	Success	Both examinee and assistant can hear each other and use their microphones. The assistant can also see the examinee on camera during a video conversation, but the examinees only sees the SEB environment

Wang et al. [17] present an approach to security testing based on threat models. Using UML sequence diagrams, there is some similarity with our approach (especially the misuse sequence diagrams), but the approach of Wang et al. is more formal, aiming to support automatic generation of test cases, while our approach aims to support brainstorming of test cases that will be performed manually. Other approaches aiming for partly automated generation of test cases from various types of models can be found in [18, 19], and a review of various model-based security testing techniques can be found in [20]. Agile security testing, proposed in [21], uses abuse stories or misuse cases as a starting point, thus having some resemblance with our approach, and in [22] it is further discussed how this can be fit into Scrum. These approaches have some similarities with ours in the initial part, having misuse cases as a possible starting point. Our approach however lacks the connection to agile/Scrum and does not make any assumption about the process, and instead proposes the choice of several different modeling representations, depending on what is found most fitting in the situation.

6 Conclusions and Further Work

This paper has proposed an approach to using models as a basis for brainstorming possible attacks and developing these into penetration tests. It must be admitted that the validation is so far limited, with only 8 lab tests executed so far. Future work in the investigation about e-exams would be to include a broader range of tests, including real-world. Indeed, real-world testing could also be applied to traditional pencil and paper exams, for instance to create a benchmark to establish if cheating is easier with e-exams than with traditional paper exams, which − although often intuitively assumed − need not be the case [23]. Since paper exams are not 100 % secure against cheating either, e-exams may be preferred even in spite of weaknesses, if they are found to have advantages in other respects [8, 24].

For the validation of the proposed method, future work could include experiments to investigate whether people come up with more or better penetration tests if using these modeling languages than if using other approaches (either completely ad hoc, some of those presented in related work, or other modeling approaches like for instance goal-oriented models). It would also be interesting to see if a top-down or bottom-up process to attack brainstorming is the most effective, as well as whether brainstorming is most effective in groups or individually.

References

1. Barmi, Z.A., Ebrahimi, A.H., Feldt, R.: Alignment of requirements specification and testing: a systematic mapping study. In: 2011 IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops (ICSTW). IEEE (2011)
2. Unterkalmsteiner, M., Feldt, R., Gorschek, T.: A taxonomy for requirements engineering and software test alignment. ACM Trans. Soft. Eng. Method. (TOSEM) **23**(2), 16 (2014)
3. Talukder, A.K., et al. Security-aware software development life cycle (SaSDLC) - processes and tools. In: IFIP International Conference on Wireless and Optical Communications Networks, WOCN 2009 (2009)
4. Arkin, B., Stender, S., McGraw, G.: Software penetration testing. IEEE Secur. Priv. **1**, 84–87 (2005)
5. Palmer, C.C.: Ethical hacking. IBM Syst. J. **40**(3), 769–780 (2001)
6. McDermott, J.P., Attack net penetration testing. In: Proceedings of the 2000 Workshop on New Security Paradigms, pp. 15–21. ACM: Ballycotton, County Cork, Ireland (2000)
7. Karpati, P., Opdahl, A., Sindre, G.: HARM: hacker attack representation method. In: Cordeiro, J., Virvou, M., Shishkov, B. (eds.) Software and Data Technologies, pp. 156–175. Springer, Heidelberg (2013)
8. Dawson, P., Five ways to hack and cheat with bring-your-own-device electronic examinations. Br. J. Educ. Technol. (2015). http://onlinelibrary.wiley.com/doi/10.1111/bjet.12246/epdf
9. Frankl, G., Schartner, P., Zebedin, G.: Secure online exams using students' devices. In: 2012 IEEE Global Engineering Education Conference (EDUCON). IEEE (2012)
10. Katta, V., Karpati, P., Opdahl, A.L., Raspotnig, C., Sindre, G.: Comparing two techniques for intrusion visualization. In: van Bommel, P., Hoppenbrouwers, S., Overbeek, S., Proper, E., Barjis, J. (eds.) PoEM 2010. LNBIP, vol. 68, pp. 1–15. Springer, Heidelberg (2010)

11. Karpati, P., Sindre, G., Opdahl, A.L.: Visualizing cyber attacks with misuse case maps. In: Wieringa, R., Persson, A. (eds.) REFSQ 2010. LNCS, vol. 6182, pp. 262–275. Springer, Heidelberg (2010)
12. Karpati, P., Opdahl, A.L., Sindre, G.: Investigating security threats in architectural context: Experimental evaluations of misuse case maps. J. Syst. Soft. **104**, 90–111 (2015)
13. Amyot, D., et al.: Generating scenarios from use case map specifications. QSIC **3**, 108–115 (2003)
14. Søgaard, T.M.: Cheating Threats in Digital BYOD Exams: A Preliminary Investigation. NTNU, Trondheim (2015)
15. Schneider, D.: Safe exam browser 2.0 how to (Install, Configure, Deploy and Use SEB 2.0) (2014). http://safeexambrowser.org/presentations/HowTo_SEB2.0.pdf
16. Cota, G.L., et al.: A framework for the design configuration of accountable selfish-resilient peer-to-peer systems. In: 2015 IEEE 34th Symposium on Reliable Distributed Systems (SRDS). IEEE (2015)
17. Wang, L., Wong, E., Xu, D.: A threat model driven approach for security testing. In: Proceedings of the Third International Workshop on Software Engineering for Secure Systems. IEEE Computer Society (2007)
18. Xu, D., et al.: Automated security test generation with formal threat models. IEEE Trans. Dependable Secure Comput. **9**(4), 526–540 (2012)
19. Marback, A., et al.: A threat model-based approach to security testing. Soft. Pract. Experience **43**(2), 241–258 (2013)
20. Schieferdecker, I., Grossmann, J., Schneider, M.: Model-based security testing (2012). arXiv preprint arXiv:1202.6118
21. Tappenden, A., et al.: Agile security testing of web-based systems via httpunit. In: Proceedings of the Agile Conference, 2005. IEEE (2005)
22. Erdogan, G., Meland, P.H., Mathieson, D.: Security testing in agile web application development - a case study using the EAST methodology. In: Sillitti, A., Martin, A., Wang, X., Whitworth, E. (eds.) XP 2010. LNBIP, vol. 48, pp. 14–27. Springer, Heidelberg (2010)
23. Sindre, G., Vegendla, A.: E-exams versus paper-based exams: a comparative analysis of security threats and countermeasures. In: Norwegian Information Security Conference (NISK 2015). Bibsys OJS: Ålesund (2015)
24. Sindre, G., Vegendla, A.: E-exams and exam process improvement. In: UDIT 2015. Bibsys OJS: Ålesund (2015)

Paper 3:

A Systematic Mapping Study on Requirements Engineering in Software Ecosystems

Aparna Vegendla, Anh Nguyen Duc, Shang Gao, Guttorm Sindre

Journal of Information Technology Research (JITR)

Paper 4:

Mitigation of Cheating in Online Exams: Strengths and Limitations of Biometric Authentication

Aparna Vegendla, Guttorm Sindre

In Kumar, A. (Ed.), Biometric Authentication in Online Learning Environments

Paper 5:

E-Assessment in Programming Courses: Towards a Digital Ecosystem Supporting Diverse Needs?

Aparna Chirumamilla, Guttorm Sindre

In Proceedings of Conference on e-Business, e-Services and e-Society (I3E)

E-Assessment in Programming Courses: Towards a Digital Ecosystem Supporting Diverse Needs?

Aparna Chirumamilla(✉) and Guttorm Sindre

Department of Computer Science,
Norwegian University of Science and Technology,
Trondheim, Norway
{aparnav,guttors}@ntnu.no

Abstract. While a number of advantages have been discussed on e-learning/e-assessment tools, little research has been reported on programming courses. Today, the different types of questions have been used in exams based on course type, e.g., Text-based questions, mathematical questions, and programming questions. All these question types require supporting plug-ins for e-assessments. In this study, we provide our practical experience on programming exams in Inspera Assessment and Blackboard Learn, especially focusing on Parsons problems (drag-and-drop questions) and code writing questions. Our findings indicate that currently, tools have basic support for programming exams, and also there is a low-level integration between the tools. However, the adaptability of any exam system could depend on the interoperability between the platforms and external plugins. Hence, more improvements can be made with the implementation of e-assessments in digital ecosystems while it requires a lot of changes internally and outside institutions. In the paper, we will explain how a digital ecosystem within e-assessment could improve assessments and how it supports diverse needs of programming exams.

Keywords: Digital ecosystem · e-Assessments · Programming exams · Parson problems · Code writing

1 Introduction

Many universities are transitioning from pen and paper exams to e-exams [1]. At the same time, formative e-assessment is receiving increased attention [2]. With automated self-tests where students can get immediate feedback, it is possible to have rapid feedback cycles scale to large and distributed classes without overloading the teaching staff. However, e-assessment systems need to be well adapted to user needs, supporting appropriate assessment tasks for the intended learning outcomes. The development of good test items is often time-consuming, so universities could save effort and increase quality if tests could be shared across countries and learning institutions [3]. Also, it would be interesting to share data and metadata, e.g., about the performance of various student groups, for benchmarking and adaptive testing.

A digital ecosystem is a business ecosystem based on an organizational network in the context of digital technology [4–6]. Digital ecosystems are formed based on digital

Published by Springer Nature Switzerland AG 2019
I. O. Pappas et al. (Eds.): I3E 2019, LNCS 11701, pp. 585–596, 2019.
https://doi.org/10.1007/978-3-030-29374-1_47

objects (digital content, products, ideas, software, hardware, infrastructure) that are interchanged and shared between independent actors [7]. The potential advantages of digital ecosystems in e-learning were outlined more than a decade ago [8, 9]. An e-learning ecosystem is the learning community, together with the enterprise, united by a learning management system (LMS) and it is formed by three categories of components: content providers, consultants, and infrastructure [8]. For the e-assessment aspects of such an ecosystem, sharing of content (e.g., tests and test items) and metadata (e.g., anonymized student scores on test items, to assess difficulty) would be a key ingredient. In addition, easy development and good availability of plug-ins to support various needs in e-assessment would be essential. Traditional monolithic systems might have the ambition that customers find all the features they require within the system. However, user needs will be quite diverse, related to different disciplines and learning outcomes, pedagogical approaches, assessment types, different devices to be used, students with special needs, languages and cultures, and different national rules and regulations of assessments, grading and collection of personal information. In addition, the system should be able to evolve quickly to cater for new needs [10], e.g., new learning methods, test types, technology.

Although monolithic systems may include many features, these will tend to be features that a sufficient number of mainstream customers require, while more specialized needs will not be supported. Moreover, they tend to become heavy and slow to respond to changes. If an e-learning system has an open, well-documented API, this could allow for plug-ins from other vendors, or from universities themselves, with niche expertise to quickly develop functionality supporting specific needs. Our research questions for this paper are: ***RQ1:*** *To what extent does e-learning/e-assessment tools support e-assessment tasks specifically needed in programming courses?* ***RQ2:*** *In what ways could a digital ecosystem within e-assessment make for improved assessments?*

In the case study performed we look in most detail at the tools used in the authors' own university, which we had the opportunity to try out in detail, whereas other related tools were only studied via documentation available on the internet. The rest of the paper is structured as follows: Sect. 2 provides some background on question types in programming and identifies two question types for which the support (or lack of support) will be specifically investigated in the case study – namely Parsons problems [11] and code writing questions [12]. Section 3 then looks at the support for these question types in typical e-assessment/e-learning tools, with most detailed focus on the tools used in the authors' university, namely Blackboard Learn and Inspera Assessment. Section 4 then discusses whether the progress towards digital ecosystems with open API's could help improve the support for more diverse needs in e-assessment. Finally, Sect. 5 concludes the paper.

2 Question Types for E-Assessment in Programming

Programming exams may contain many different types of questions [13]. The below list provides some broad categories:

- Conceptual questions: These are questions that do not directly involve code, but focus on the recall and understanding of concepts, e.g., "What is a key difference

between a list and a set?" (possibly a multiple choice question) or "Explain the concept of polymorphism and its utility?" (possibly a free text question)
- Code tracing: The code is given, and the candidate's task is to explain what the code does. Within this category, questions may vary from those requiring only brief answers, e.g., "What will be the output of this program?", to more detailed ones, e.g., "Explain what this program does, line by line."
- Code writing: It is explained what a program is supposed to do, and the candidate's task is to write the code.
- Code completion: It is explained what a program is supposed to do, and some code is provided, but not fully complete. The candidate's task is then to fill in or select missing parts, or to rearrange code lines in the correct order.
- Error detection: It is explained what a program is supposed to do, and some faulty code is provided. The candidate's task is then to identify the mistakes, possibly also to propose corrections.

As indicated by Sheard et al. [14], code writing appears to be the most used question type in programming exams, followed by code tracing. Writing and tracing tasks can be seen as opposites, i.e., write all the code vs. write no code (rather understand the code which is given). Completion and error detection tasks as somewhere in between those two extremes, requiring both understanding of the code already given, and ability to write some extra code: the missing parts to be added to completion tasks, the corrections to be proposed for error detection tasks.

A detailed analysis of all possible question types would be prohibitively time-consuming, so here we choose to focus on two specific question types, namely Parsons problems [11] and code writing questions [12]. The reason for choosing these two types is that they are quite specific for the discipline of programming, whereas other question types could more easily be supported by generic question types found in most e-assessment and e-learning systems. For instance, conceptual questions could be implemented as free-text short answer tasks or multiple choice questions. The same applies to code tracing questions, where the brief answer variety might typically be given as multiple choice, fill-in-number or fill-in-text depending on the output, while the longer variety could be a free-text answer or a sequence of fill-in fields showing the changes of variable content during execution. Code completion tasks (other than Parsons problems) could be implemented by e.g. multiple choice, fill-in, or pull-down menus for each missing code fragment, and error detection could again be short answer, fill-in (for proposed corrections) or multiple choice (selecting between real errors and distractors).

What are then the particularities of the two mentioned question types? *Parsons problems* [11] are coding problems where it is explained what some piece of code is supposed to do, and the code lines are given, but in jumbled order. It is then the candidates' task to rearrange them in the right order. This question type has attracted a lot of research interest [15–18] because it reduces cognitive load for the students (e.g., recall of syntax, avoiding typing mistakes), yet still tests their visual-spatial abilities, constructive skills in solving a problem and constructing a solution from available building blocks. Since building blocks are larger (entire code lines rather than character by character on the keyboard), each question can be solved faster, thus potentially

achieving better topical coverage in the exam set as a whole. Also, quick solution and automated feedback make such problems interesting for digital learning resources with self-testing features, for instance, the interactive e-book [19] makes extensive use of such problems among its exercises. Questions in Parsons problems can be made easier by providing hints [20] or more difficult by adding distractors [18], they can be one-dimensional (most common) or two-dimensional [21], the latter relevant with programming languages where indents have semantic significance (e.g., Python).

A common way of implementing Parsons problems digitally would be as drag-and-drop questions – a featured question type in many e-learning/e-exam applications. Drag and drop questions may test students' higher order thinking skills, i.e., algorithmic problem-solving skills [22, 23]. The recent research has been progressed more towards the visual programming language (VPL) that allows users to create programs using drag-and-drop genre [24]. However, its use in e-exam applications will normally not have been made with programming tasks in mind, rather tasks such as placing names in the correct positions on a background picture (e.g., Latin names of body parts for an anatomy exam, names on countries on a map for a primary school Geography exam). Hence, standard tool support for drag-and-drop questions may not be ideal for Parsons problems in programming.

Code writing tends to be a key element of programming exams, and most would agree that doing these tasks with pen and paper is not particularly authentic. Switching to a digital interface will make the task more similar to real work – but not necessarily fully authentic, as there may be various ambition levels to the tool support. For instance, students may be able to type the code in the test interface, but this could be in an editor with specific support for code writing (more authentic) or in a generic text input window with few functional features (less authentic). Also, students might be able to compile and run the code (more authentic), or not (less authentic). Sometimes, the more authentic, the better – but not always. A problem with the ability to compile, run, and test the code during an exam, for instance, is that students will then spend more time on each programming task – due to the need to debug and rerun if something was not working. More time on each task would give poorer coverage of the learning outcomes, especially if tool usage was not among the specified learning outcomes for the course. An ideal e-exam tool should therefore have a wide range of support for code writing tasks, anything from writing in a fairly simple editor without the ability to run, to professional tool support for code editing, testing and debugging.

3 Analysis of Mainstream Tool Support

As shown in [25], there are many tools for e-assessment of programming, but many of these are standalone applications or cloud tools not integrated with official university information systems. This section looks at mainstream tool support for Parsons problems and code writing problems, with special focus on Blackboard Learn and Inspera Assessment, which happen to be the mandatory tools in the authors' university for formative and summative e-assessment, respectively. The first subsection looks at Blackboard Learn, the second at Inspera Assessment, and the third makes a quick review of some other tools.

3.1 Blackboard Learn

Blackboard Learn is the current LMS for the authors' university. It is used for communication between teaching staff and students during the semester, e.g., course info and announcements, learning resources, exercises (if not graded), etc. It is not compulsory to use it for everything, so teaching staff could use supplementary tools, in addition, for instance, for students' automated self-testing. However, it would be convenient both for teachers and students if course tasks are seamlessly supported through Blackboard, so that they avoid confusing and time-consuming switches between tools [26].

Support for Parsons problems in Blackboard turns out to be limited. Drag and drop questions do not exist, so such questions would instead have to be approximated by other question types. Obvious candidates might be *ordering questions* or *jumbled sentence questions*. Ordering questions would show the code lines in a shuffled order, then let the user assign ordinal numbers to each in input fields beside the code lines. This is not entirely ideal for the purpose. For instance, code lines are not repositioned, so the resulting code is not easily read. Reordering requires changing the ordinal numbers of all code lines affected, whereas a modern drag and drop interface might solve this by repositioning fewer lines. Jumbled sentence questions would give a series of input fields, where each would yield a drop-down menu when clicked, with all the code lines as alternatives. The student would then have to make a multiple choice selection for each input field. This would appear somewhat better than the ordering question since at least the code would be shown in the wanted order when selections had been made. However, reordering would have the same issues as with the ordering questions, and if the task contains many code lines, the drop-down menus will be long and clumsy.

Specific support for Code writing problems in Blackboard does not exist, beyond generic essay and short answer question forms meant for natural language text, or using file upload questions (e.g., student could write the code in a separate tool more fit for programming, and then upload the file to Blackboard).

3.2 Inspera Assessment

When it comes to Parsons problems, Inspera Assessment does support drag and drop questions. The resulting interface for the student while solving the task is therefore more elegant than what can be achieved in Blackboard, though there are some issues with the user interface. The task has to be made with separate drop areas for each code line, rather than one big drop area where the order is given by relative positioning. This means that the student still has to reposition several code lines in cases where a better interface might have gotten away with just repositioning one line and having other lines yield place. Especially, if trying to make two-dimensional Parsons problems, the snapping feature may behave a little counter-intuitively, since it is not determined by the position of the mouse pointer, rather the middle of the drag object (mouse pointer would be more natural, or the left edge of the drag object). Parsons problems become very time-consuming for the teacher to develop in Inspera, since all the drag areas must be created manually one by one and filled with solution (and possibly distractor) code

lines, and then linked to the correct drop areas, also manually created one by one. Especially for two-dimensional Parsons problems, this takes quite a lot of time. An illustration of a two-dimensional Parsons problem for Python, as implemented in Inspera, is shown in Fig. 1. For space reasons, the natural language explanation of what the code was supposed to do is omitted, showing only the interactive part of the screen. The candidate's task would be to drag each code line into the correct position in the grid (the function heading `def deriv(poly):` going upper left), both concerning vertical order and horizontal indenting, as indents have semantic significance in Python. In Inspera Assessment, the 28 drop areas must be created one by one, hand positioned in the grid and adjusted for size, hence quite time-consuming for the question author.

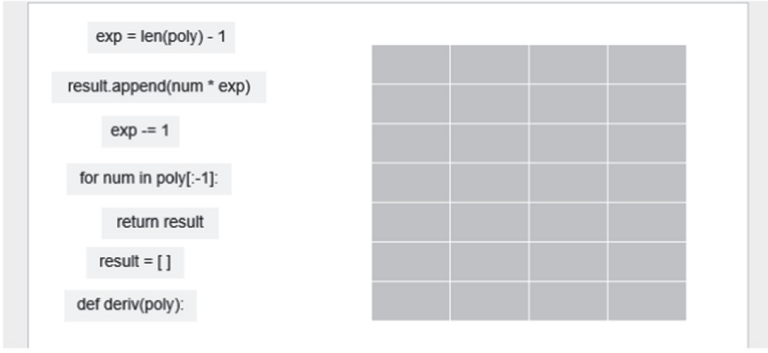

Fig. 1. Two-dimensional Parson problem for Python.

For code writing tasks, Inspera has a dedicated question type called "Programming". Notably, the student is not able to compile and run the code during the exam, nor is staff able to run it afterwards in connection with grading, so this type of task is manually graded. However, it does support the following features:

- A monotype font suitable for code, and syntax highlighting for some much used programming languages
- Other syntax related support, such as automatically giving an end parenthesis for each start parenthesis, and automatically making indents where appropriate, for instance in Python if the previous code line ended with a colon.

All in all, then, Inspera Assessment has better question type support both for Parsons problems and code writing than what Blackboard has, but still with substantial limitations. The user interaction for drag and drop questions is somewhat tedious for students, especially if reordering, and for teacher authoring of questions it is even more tedious. For code writing questions, both have the shortcoming that the code will not run and must be manually graded, and Blackboard does not even have syntactic support. Hence, both Blackboard and Inspera could clearly be made much more usable for handling these question types if there were plugins specifically targeting them.

3.3 Other Tools

Table 1 gives a summary of the possible support for Parsons problems and code writing problems in various tools. In addition to Inspera and Blackboard, other tools worth looking at are the e-exam tool WISEflow (a competitor to Inspera) and general LMS tools Canvas and Moodle (competitors to Blackboard). The authors gathered information about these tools from web-documentation since they do not have direct access for these tools in their institution. Our findings show that Blackboard does not support drag-and-drop functionality while all the other tools support this feature. However, these tools only support the basic functionality of drag-and-drop into text and image, which is not ideal for Parson problems. Code writing is supported in Inspera and Moodle, moreover it seems Moodle has better support than Inspera. Both Moodle and Inspera support code writing with syntactic support (e.g., indentation and code highlighting). In addition, Moodle has an external plugin, Coderunner that allows students to run their programs during exams and teachers to run programs in order to grade student's answers. Limitations of the functionalities in tools can be improved further by third-party extensions and plugins with the adoption of digital ecosystems.

Table 1. Tool support summarized.

Tool	Parsons problems	Code writing	Import/export questions	Plugins
Blackboard	Lacks drag&drop	No specific support (free text)	QTI, LTI	LTI, Google Apps SafeAssign
Inspera	Has drag&drop, but not ideal	Only syntactic support for code [27]	QTI, LTI	Atlassian Jira
Canvas	Hasdrag&drop but not ideal	No specific support (free text)	QTI, LTI	LTI, Facebook, Google Drive, Twitter, Tinychat Google Docs, Kaltura, LinkedIn, Canvasdocs
WISEflow	Hasdrag&drop but not ideal	No specific support (free text)	QTI, Canvas, Moodle XML, Blackboard V6-9	
Moodle	Hasdrag&drop but not ideal	Syntactic support, Code runner support	QTI, LTI, GIFT Moodle XML, XHTML, LTI	SEB Quiz Access, Coderunner Rule, LTI, Turnitin, Plagiarism

4 Towards a Digital Ecosystem

Tools like those discussed in Sect. 3 can import/export questions in the QTI (Question and Test Interoperability) format [28]. So for authoring of drag-and-drop questions (which was somewhat cumbersome in Inspera), a possible way to improve the support would be to make a stand-alone authoring tool that could generate questions as QTI files, then to be uploaded to Inspera, for instance as suggested by [29]. In Blackboard, such an authoring tool would not be of much use, since the question type is not supported. Hence, Blackboard would need an integrated plugin supporting the question type, and an integrated plugin would probably appear better for the user of Inspera, too, especially for students solving the tasks, since the user interface could then be improved with custom features for Parsons problems. A plugin might also be a possible solution for better support of code writing questions in both tools (e.g., for the student, ability to compile and test the code during the exam; for the teacher, support for automated testing and grading of delivered code).

Currently, Inspera offers REST-based APIs to enable the third-party developers to integrate the additional functionalities and a Custom Interaction API that allows customers to build specialized question types. It supports stimuli elements with JavaScript and mathematical tools such as Geogebra and Desmos. These specialized question types can still be exchanged through QTI specification and the IMS Global Assessment Custom Interactions specification.

As the digitization of the exams increased, the need for technology for exams is also rapidly increasing. However, the usability of a digital exam system highly depends on the simplicity of the system. Also, users are sometimes forced to use several systems, not well integrated. For instance, in the authors' own university Blackboard is actively used as an LMS while Inspera is used as an assessment tool. The key requirements from teachers in the computer science department at our university that are ecosystems related include:

- Teachers want to have some exercises using the Inspera UI rather than Blackboard's, to give the students more accurate exam practice. Preferably, students should then be able run Inspera via Blackboard, so that Blackboard could still automatically register who has delivered the exercise.
- Concerning the import and export of contents, teachers may want to use last year's exam questions as exercise questions the next year. However, while Inspera can export questions in QTI 2.1 format, Blackboard (at least the version in our university) for some reason only seems to support the older QTI 1.2 standard.

In a well functioning software ecosystem, the platform system would have open APIs for external third-parties to develop plug-ins on top of the platform. This type of solution has several advantages over monolithic exam systems. García-Holgado and García-Peñalvo [30] explained that technological ecosystems could be considered as a framework to develop technological solutions where information and the human factor are the centre of the system. One of the main advantages with such an ecosystem is the flexibility it provides to institutions to integrate new software components within their workflows to support emerging needs.

The key requirements from teachers could be fulfilled to some extent with the current plug-in support by Inspera: (i) Integrate contents and external tools into LMS. Inspera supports sharing of the contents through the IMS Learning Tools Interoperability™ (LTI) plugin. LTI is an interoperability specification which facilitates full integration between Inspera and Blackboard. With LTI support, Inspera can be launched as a tool from Blackboard, which allows students to take exams directly through Blackboard. This feature is currently supported in Canvas, Blackboard, and Moodle [31]. (ii) Sharing of the contents across e-learning platforms. Issues with import and export questions can be reduced with more updates in the versions of interoperability specifications of platforms and tools [32]. In [30], the authors addressed the problem of sharing questions across departments in university in the e-learning context. They argued that although the technological ecosystem provides tools to facilitate communication between departments, employees are not utilizing the tools.

Presently, Inspera only supports sharing questions among teachers in the same university – for wider sharing, one must export and import. Of course, one deterrent against easier sharing could be increased fear for question leakage, i.e., confidential exam questions being disclosed to candidates before the exam. However, it mostly seems to come down to lacking features, and a natural tendency to prioritize the basic features first: support for each autonomous teaching staff for making the exam in their course, rather than to support a wider community of teachers within a discipline in making larger question bases that can be shared and continuously quality assured and updated.

However, Inspera also has some frustrating shortcomings on the single course level. In Norway, the law says that complaint graders shall not know the grades or viewpoints of the original graders. However, in Inspera it was impossible to hide given scores on the tasks. This meant that complaint graders could not do their grading in Inspera, but instead had to receive pdf screenshots of student answers, and then had to score manually even tasks like multiple choice, that would have been auto-scored in Inspera – with higher work-load and increased risk of error as a result. Fixing such issues will of course have a higher priority for the next release than more ambitious ideas supporting disciplinary communities. In Norwegian universities, Inspera must also be integrated with FS (Common Student System), a legacy system used for the administration of students in universities. Both the LMS and the e-exam system will fetch information from FS (e.g., which students are enrolled, registered for the exam, etc.) and send information back to FS (e.g., grades). The legacy system is not directly seen by students or teachers, but by administrative personnel – for instance it also contains the link between anonymous candidate numbers used during exams and the students' identities.

The implementation of digital exam ecosystems involves a higher degree of complexity due to the integration of different components that should evolve both individually and collectively. Although the REST APIs aids the developer, lack of the framework and design patterns makes the integration with plug-ins more difficult. A framework for technological ecosystems will consider all aspects related to integration, interoperability, and the evolution of the components [33] thus forms the well-developed open ecosystem. Several frameworks and methods were discussed in the literature. For instance, A framework can be designed using architectural patterns using

the Business Process Model and Notation (BPMN) [30]. García et al. proposed a service-based framework connecting Moodle LMS and Basic LTI (BLTI) [33]. Consequently, it could ease commercial vendors and free software developers to make plugins supporting the authoring, solving, and grading of various question types.

5 Conclusion

Several advantages have been discussed in literature about e-assessments, and today many tools are available for course management and assessments. Although many e-learning/e-assessment tools are available, only a few support programming exams. In the paper, we discussed our practical experience with programming questions in Blackboard Learn and Inspera Assessment tools, particularly focusing on Parsons problems (i.e., drag - and–drop questions) and code writing questions. Our observations revealed that currently Inspera, Moodle, Canvas, WISEflow supports drag-and-drop questions but not ideal for programming using Parson problems. Also, there is a low-level integration between Inspera and Blackboard for programming exams. The improvements can be made further with the transition of a monolithic digital exam system to digital exam ecosystem by opening APIs though it requires a lot of changes internally and outside institutions. However, open APIs alone cannot be able to improve e-assessments, without the support of frameworks, and architectural designs that explain software updates, security policies, access permissions etc. Though many papers discussed ecosystem phenomenon in e-learning, its implementation on the digital exam is still in infancy. This paper has initiated the concept of the ecosystem in the digital exams area focusing on programming exams.

The paper still has some limitations: It discussed only details of the tools used in authors' university, Inspera and Blackboard, since they have direct access to only these tools. Currently, there are many tools available for digital assessment; the study of every tool would require more time for research and cost (to buy licenses for tools). Moreover, students and teachers are adapted to the tools they use, so it is more convenient to receive their feedback. The findings from this study are based on the author's practical experience. Hence, this study can be improved in the future by more quantitative and qualitative research in academia and industries, especially on the perspective of a digital ecosystem.

References

1. Fluck, A.: An international review of eExam technologies and impact. Comput. Educ. **132**, 1–15 (2018)
2. Spector, J.M., et al.: Technology enhanced formative assessment for 21st century learning (2016)
3. Veiga, W., et al.: A software ecosystem approach to e-learning domain. In: Proceedings of the XII Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era-Volume 1. Brazilian Computer Society (2016)

4. Stanley, J., Briscoe, G.: The ABC of digital business ecosystems. arXiv preprint arXiv:1005.1899 (2010)
5. Nachira, F., Dini, P., Nicolai, A.: A network of digital business ecosystems for Europe: roots, processes and perspectives. Introductory Paper, 106. European Commission, Bruxelles (2007)
6. Jansen, S., Cusumano, M.A.: Defining software ecosystems: a survey of software platforms and business network governance. In: Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry, p. 13 (2013)
7. Kallinikos, J., Aaltonen, A., Marton, A.: The ambivalent ontology of digital artifacts. MIS Q. **37**(2), 357–370 (2013)
8. Uden, L., Wangsa, I.T., Damiani, E.: The future of E-learning: E-learning ecosystem. In: 2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference, DEST 2007 (2007)
9. Oskar, P.: Software ecosystems and e-learning: recent developments and future prospects. In: Proceedings of the International Conference on Management of Emergent Digital EcoSystems, France, pp. 427–431. ACM (2009). ISBN 978-1-60558-829-2
10. Marti, R., Gisbert, M., Larraz, V.: Technological learning and educational management ecosystems. Thirteen characteristics for efficient design. In: EdMedia+ Innovate Learning. Association for the Advancement of Computing in Education (AACE) (2018)
11. Parsons, D., Haden, P.: Parson's programming puzzles: a fun and effective learning tool for first programming courses. In: Proceedings of the 8th Australasian Conference on Computing Education, vol. 52. Australian Computer Society, Inc. (2006)
12. Sheard, J., et al.: Assessment of programming: pedagogical foundations of exams. In: Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education. ACM (2013)
13. Simon, et al.: Introductory programming: examining the exams. In: Proceedings of the Fourteenth Australasian Computing Education Conference, vol. 123. Australian Computer Society, Inc. (2012)
14. Sheard, J., et al.: Exploring programming assessment instruments: a classification scheme for examination questions. In: Proceedings of the Seventh International Workshop on Computing Education Research. ACM (2011)
15. Denny, P., Luxton-Reilly, A., Simon, B.: Evaluating a new exam question: Parsons problems. In: Proceedings of the Fourth International Workshop on Computing Education Research. ACM (2008)
16. Helminen, J., et al.: How do students solve parsons programming problems?: an analysis of interaction traces. In: Proceedings of the Ninth Annual International Conference on International Computing Education Research. ACM (2012)
17. Ericson, B.J., Margulieux, L.E., Rick, J.: Solving parsons problems versus fixing and writing code. In: Proceedings of the 17th Koli Calling International Conference on Computing Education Research. ACM (2017)
18. Harms, K.J., Chen, J., Kelleher, C.L.: Distractors in Parsons problems decrease learning efficiency for young novice programmers. In: Proceedings of the 2016 ACM Conference on International Computing Education Research. ACM (2016)
19. Guzdial, M., Ericson, B.: CS Principles: Big Ideas in Programming. RuneStone Academy (2014)
20. Morrison, B.B., et al.: Subgoals help students solve Parsons problems. In: Proceedings of the 47th ACM Technical Symposium on Computing Science Education. ACM (2016)
21. Ihantola, P., Karavirta, V.: Two-dimensional parson's puzzles: the concept, tools, and first observations. J. Inf. Technol. Educ. **10**, 119–132 (2011)

22. Kalelioğlu, F.: A new way of teaching programming skills to K-12 students: Code.org. Comput. Hum. Behav. **52**, 200–210 (2015)
23. Lee, Y.Y., Chen, N., Johnson, R.E.: Drag-and-drop refactoring: intuitive and efficient program transformation. In: Proceedings of the 2013 International Conference on Software Engineering, San Francisco, CA, USA, pp. 23–32. IEEE Press (2013)
24. Tsai, C.-Y.: Improving students' understanding of basic programming concepts through visual programming language: the role of self-efficacy. Comput. Hum. Behav. **95**, 224–232 (2019)
25. Gupta, S., Gupta, A.: E-Assessment tools for programming languages: a review. In: Proceedings of the First International Conference on Information Technology and Knowledge Management (2018)
26. Forment, M.A., Guerrero, M.J.C., González, M.Á.C., Peñalvo, F.J.G., Severance, C.: Interoperability for LMS: the missing piece to become the common place for Elearning innovation. In: Lytras, M.D., et al. (eds.) WSKS 2009. LNCS (LNAI), vol. 5736, pp. 286–295. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04754-1_30
27. Inspera Assessment. Programming - Knowledge Base - Inspera. https://inspera.atlassian.net/wiki/spaces/KB/pages/57311556/Programming
28. IMS Global. Question and Test Interoperability (QTI): Overview. https://www.imsglobal.org/question/qtiv2p2/imsqti_v2p2_oview.html
29. Jørgensen, J., Kvannli, S.: Efficient generation of Parsons problems for digital programming exams in Inspera, Department of Computer Science. NTNU, Trondheim (2019)
30. García-Holgado, A., García-Peñalvo, F.J.: Architectural pattern to improve the definition and implementation of eLearning ecosystems. Sci. Comput. Program. **129**, 20–34 (2016)
31. Inspera Assessment. Assessment technology standards. http://www.inspera.com/standards
32. Dagger, D., et al.: Service-oriented e-learning platforms: from monolithic systems to flexible services. IEEE Internet Comput. **11**(3), 28–35 (2007)
33. García Peñalvo, F.J., et al.: Opening learning management systems to personal learning environments. J. Univers. Comput. Sci.: J. UCS **17**(9), 1222–1240 (2011)

Paper 6:

Cheating in e-exams and paper exams: the perceptions of engineering students and teachers in Norway

Aparna Chirumamilla, Guttorm Sindre, Anh Nguyen Duc

Assessment & Evaluation in Higher Education

Paper 7:

E-exams in Norwegian Higher Education: Vendors and managers views on requirements in a digital ecosystem perspective

Aparna Chirumamilla, Guttorm Sindre

Computers & Education

E-exams in Norwegian Higher Education: Vendors and managers views on requirements in a digital ecosystem perspective

Abstract

E-assessment has been supported in Learning Management Systems for decades. More recently, dedicated e-exam systems have emerged on the market, more specifically supporting the workflow and security needs surrounding high stakes exams. For instance, in Norway, LMS's Canvas and Blackboard are only used for ungraded assessment tasks, while e-exam systems like WISEflow and Inspera Assessment are used for graded ones. Since the systems are mass-market software, vendors must satisfy the needs of several customers, and needs that are specific to only one or a few customers will receive low priority, perhaps forcing teachers to adapt their assessments to what the tool supports, rather than having the tool adapt to the preferred pedagogy. So far, there has been considerable research on views of students and teachers on e-exam systems, much less on the views of vendors and managers. In this paper, we investigate what these stakeholder groups consider to be the key features of e-exam systems, and by what process they are determined. An exploratory case study was conducted, based on interviews with 12 participants belonging to three different groups: vendors, process manager and system managers in Norwegian universities. Our findings indicate much agreement among these groups about key features of e-exam systems, though observing that not all functionality requested by end-users will be prioritized. Also, there was much agreement that a movement towards standardization, open interfaces and digital ecosystems would allow smoother integration with other information systems in the higher education sector, and easier addition of plug-ins for specific functionality – but that there still is a way to go to reach the ambitions of a flexible ecosystem. Currently, vendors give more priority for adding functional features in e-exam systems rather than better interoperability, and integration with third-party tools remains a challenge.

Keywords: e-exam system; requirements engineering process; features; digital ecosystem; interoperability.

1. Introduction

The higher education sector (HE sector) is currently going through a massive digital transformation (Llamas-Nistal, Fernández-Iglesias, González-Tato, & Mikic-Fonte, 2013; Sandkuhl & Lehmann, 2017). One exciting trend is the digitisation of assessment. While e-assessment has been with us for decades (Buzzetto-More & Alade, 2006; Frosini, Lazzerini, & Marcelloni, 1998), we are now witnessing a much more consistent shift. Many countries have ambitions to abandon traditional pen and paper exams within five years (Butler-Henderson & Crawford, 2020; Fluck, 2019).

The general trend for e-assessment products is towards mass-market software, with competing vendors targeting the same needs. Many universities around the world use Blackboard or Canvas as their learning management system (LMS). For high-stakes exams, they may supplement the LMS with software to safeguard against cheating, such as the Respondus Lockdown Browser (Cluskey Jr, Ehlen, & Raiborn, 2011). Norwegian universities also have Blackboard or Canvas but use dedicated e-exam systems such as Inspera or WISEflow for graded tests. Mass-market software allows development cost to be divided on a large number of customers. On the other hand, there are also challenges with mass-market software:

- Functionality will often be a compromise between needs of various customers, often imperfectly understood by vendors (Naous, Giessmann, & Legner, 2020). Less common needs (e.g., of specific courses or innovative assessment practices) will tend to get low priority.
- For high-stakes e-assessment, the need to mitigate cheating makes security requirements particularly important (Dawson, 2016). Cheating can be defined as breaking the rules of an exam to gain unfair advantage (Cizek, 1999) and is a many-faceted issue. Partly, it is about ensuring the correct identity of the candidate and correct authorship of the delivered answer (Mellar,

Peytcheva-Forsyth, Kocdar, Karadeniz, & Yovkova, 2018), but also about ensuring that candidates do not use forbidden aids or illegitimate collaboration during the exam (Dick, Sheard, Bareiss, Carter, Joyce, Harding et al., 2002).

- Mass-market software products may be hard to integrate with other systems that the university already has. Integration requirements are often hard to capture in the acquisition stage (Lauesen, 2006). The modern approach to interoperability is to move away from proprietary systems and instead use open standards and governance frameworks, so that many different software products can collaborate smoothly in a digital ecosystem (Kerssens & Dijck, 2021) – as also proposed within e-learning (Uden, Wangsa, & Damiani, 2007) and e-assessment (Llorens, Molina, Compañ, & Satorre, 2014; Luo & Lin, 2013). In spite of improvements towards this vision in recent years, interoperability remains a major challenge in the e-learning domain (Chituc & Rittberger, 2019). Poor interoperability may cause a lot of double work, e.g., re-entering of data, meaning that the administrative simplifications and cost savings that one hoped to achieve from e-exams, may not materialise (or be smaller than expected).

Both limited functionality and trade-offs between functionality, cheating prevention and interoperability mean that teachers may end up having to adapt their assessment practices to what the exam tool allows, rather than having the tool adapt to the assessment practices that are ideal from a pedagogical point of view.

Our main research question is: *What are the key features for e-exam software, according to vendors, process managers, and system managers, and how are such features identified and agreed upon?*

In connection with this main question, we have four sub-questions:

1. What process is followed by vendors, process managers, and system managers to identify features and agree upon requirements?
2. What do vendors, process managers and system managers see as key features for functionality and security in e-exam systems?
3. What are the goals and challenges concerning integrations between the e-exam system and other exam supporting systems?
4. To what extent do the stakeholders envision a move towards a more open digital ecosystem for e-exams, and would this be assumed to impact security?

Notably, our main research question omits two of the most important user groups for e-exam systems, namely students and teachers. Instead, we focus on stakeholders involved in developing e-exam systems (i.e., vendors), and in acquiring and operating e-exam systems for the university sector (i.e., system managers involved in acquisition and operation of e-exam systems, and process managers involved in negotiating requirements). This selection of informants does not mean that the views of students and teachers are considered unimportant. However, the views of students and teachers on e-exam systems have been studied in several other publications (Fluck, 2019; Kuikka, Kitola, & Laakso, 2014; Wibowo, Grandhi, Chugh, & Sawir, 2016), while less has been published on the views of vendors, system and process managers. Moreover, with research questions focusing on rather technical issues like security, interoperability, and digital ecosystems, students and teachers except for the most technologically competent might be weak informants. Obviously, the requirements of students and teachers should be essential when a university is acquiring e-exam technology. However, the requirements of teachers would likely differ a lot from person to person, e.g., depending on the discipline taught and the assessment practice followed. Among students, there will also be much variation, e.g., based on personal preferences and special needs. The advantage of system managers (i.e., license administrators, project managers, team leaders, advisors, engineers) at a university is that they receive feedback from many students and employees, e.g., concerning dissatisfaction with the system or requests for new functionality, thus able to present a more aggregate idea of requirements. Similarly, process managers employed by the Ministry of Education and Research run joint procurement processes on behalf of all public universities and vendors (i.e., development managers, head of product development, product managers) who sell to several universities should also possess such aggregate views on needs. Hence, these three groups of people should be knowledgeable about general trends concerning requirements for e-exam software.

2. Literature review

There have been several case studies concerning usage of e-exam systems, e.g., (Fluck, Adebayo, & Abdulhamid, 2017; Fluck, Pullen, & Harper, 2009; Fluck, 2019; Kuikka et al., 2014; Wibowo et al., 2016), typically eliciting viewpoints of students and teachers to inform the further development and operation of such systems. They have reported various advantages for e-exam systems, such as improved exam logistics, support for auto-marking and question authoring – and for harvesting data that can be used in learning analytics (Fitzharris & Kent, 2020). Especially, Bring Your Own Device (BYOD) e-exams, which are cloud services letting students use their own laptops during the exam – have become increasingly popular, offering reduced costs and increased scalability for universities (Fluck & Hillier, 2017; Hillier & Fluck, 2013), and better accessibility and usability for students (Fitzharris & Kent, 2020) by using the equipment they are already familiar with.

Necessary security features for high stakes e-exams have been discussed by several authors. Huszti & Petho (2010) focussed on authentication, anonymisation of candidates, and confidentiality of questions. Mitigation of cheating is considered especially challenging for BYOD e-exams since the student controls the hardware (Dawson, 2016; Frankl, Schartner, & Zebedin, 2012). The two most prominent approaches to secure such exams – requiring booting from a memory stick or accessing the exam through a lock-down browser – both have their pros and cons. Security features for BYOD exams using Moodle LMS are provided through "Secure Exam Environment" in (Frankl et al., 2012) that mainly use wired LAN and boot from USB or DVD. Kaiiali et al. (2016) proposed a "Secure Exam Management System" for the security of BYOD exams run on Moodle LMS through WIFI. With remote exams, it becomes even more challenging to mitigate cheating through impersonation (another person taking the exam) or collaboration, answer sharing and plagiarism among students (D'Souza & Siegfeldt, 2017), which has been observed with many remote exams during the Covid lockdown period (Bilen & Matros, 2021). For such situations, authorship verification combined with remote proctoring technology utilizing biometry and candidate monitoring can be helpful (Okada, Noguera, Alexieva, Rozeva, Kocdar, Brouns et al., 2019). However, our study took place before the Covid lockdown, so the context for our interviews was the then-typical usage of e-exam systems in Norwegian universities, namely for on-campus exams, with face-to-face human invigilators checking the identity of students. Hence, remote proctoring technology is not much addressed in this paper.

As for design research, Adebayo & Abdulhamid (2014), Brink & Lautenbach (2011), Ferdiana & Hoseanto (2018), and (Fluck, Pálsson, Coleman, Hillier, Schneider, Frankl et al., 2017) all report on various design and implementation experiences with e-exam systems, with combinations of in-house development and already available software. When it comes to more general frameworks, Striewe (2019) based on a literature review, proposes components and design alternatives for e-assessment systems, for each component pointing out also the features that the component would be supposed to cover. Isaias, Miranda, and Pífano (2019) made a framework of eight evaluation criteria for e-assessment systems: variety of design options, scalability, security, access and usability, feedback features, personalisation, cost and interoperability. The framework was validated by a questionnaire gaining responses from academic staff across 37 countries. Among the criteria, the highest level of agreement was for variety of question types and feedback features, and for interoperability. The EU project TeSLA is a collaboration between a number of universities across Europe, implementing a system for secure online exams, as reported for instance in (Baró-Solé, Guerrero-Roldan, Prieto-Blázquez, Rozeva, Marinov, Kiennert et al., 2018; Mellar et al., 2018; Okada, Noguera, et al., 2019; Okada, Whitelock, Holmes, & Edwards, 2019). The project has focussed on security against cheating especially for remote online exams, combining biometry and authorship verification, and also on how technology can facilitate pedagogical improvement in assessment. TeSLA has also had a strong focus on accessibility for students with special needs, and interoperability with other e-learning tools (Ladonlahti, Laamanen, & Uotinen, 2020).

E-exam systems need to interoperate with other software products performing complementary functions in the organisation, such as the Student Information System, LMS, and single sign-on system for authentication. Such interoperability challenges are somewhat discussed by (Dagger, Connor, Lawless, Walsh, & Wade, 2007; Jakimoski, 2016). Also, e-exam systems may need to interoperate with other e-

exam systems, to allow universities to move and exchange information and collaborate on joint question bases for various disciplines (Sclater, Low, & Barr, 2002). Major obstacles to interoperability could be that systems use different interfaces and data formats. Standardization can reduce this problem. Specifically for e-learning tools, the IMS Learning Tools Interoperability (LTI) standard allows external tools to be launched within an application (Queirós, Leal, & Paiva, 2016; Severance, Hanss, & Hardin, 2010). The Question and Test Interoperability (QTI) allows tests and questions exported from the e-exam system of one university to be imported to the e-exam system of another university (Wills, Davis, Gilbert, Hare, Howard, Jeyes et al., 2009). However, Piotrowski (2011) considered the QTI specification too complex, ambiguous, and challenging to implement, and Sclater (2007) argued that interoperability testing must be included at acquisition to ensure that the vendors really deliver on this. In addition to adhering to standards, products with an open and well-documented Application Programming Interfaces (API's) will more easily allow for plug-ins so that universities may customise the e-exam system according to their needs (Chirumamilla & Sindre, 2019).

All in all, the publications reviewed above do present required features for e-exam systems in various ways, but none of them directly respond to our main research question of investigating these features as seen by vendors, system and process managers. Closely related to our study is one by Foss-Pedersen & Begnum (2017), targeting the same stakeholder groups that we have interviewed, using a questionnaire survey plus interviews. However, their focus was on the e-exam systems' support for universal access, not towards functional features in general, nor for security or interoperability.

3. Research Approach

3.1. Case context

Previously, each university or college in Norway might run separate procurement processes for e-learning products, though some might also have collaboration and joint procurement. The last big procurement done by a single university was NTNU's choice of Blackboard as its new Learning Management System (LMS) in 2017. Nowadays, the national organization Unit[1] (Directorate for ICT and Joint Services in Higher Education and Research, Merger of CERES, BIBSYS and parts of UNINETT)) works as process manager and runs joint procurement processes on behalf of all Norwegian higher education institutions, which could both increase bargaining power and save resources, as procurement processes are labour intensive. Hence, when most other Norwegian universities got Canvas in 2016, this was a result of such a joint process. Similarly, Unit currently manage Norwegian HE institutions' dialogue with e-exam system vendors and development of requirements. They also have responsibility for the development and maintenance of custom software, such as FS (Felles Studentsystem) which is a Student Information System (SIS) in use by almost all higher education institutions in Norway. The architecture diagram in Fig. 1 shows various systems involved, with links indicating information exchange. FS (second from left) contains authoritative information about students (e.g., personal information, enrolment, course registration, exams scheduled, grades received, etc.), courses, teachers, etc. StudentWeb (left) is a front-end to FS where students can register or withdraw from courses/exams, view and appeal grades, etc. Blackboard and Canvas are typical LMS's, which in Norway are used to handle communication within courses – except for exams and graded coursework, which will be delivered through the e-exam system. Two mass-market software products are used for e-exams in Norway, and some universities use Inspera Assessment (hereafter IA), others WISEflow (hereafter WF). Both are proprietary software products, run as cloud services using lock-down browsers (top and bottom) to mitigate cheating. Further to the right are some other systems involved, the document archival system, the single-sign-on authentication (used with several systems, but we only show links to the e-exam systems to avoid messing up the diagram), and the plagiarism checking tool, where Norwegian HE currently use Urkund.

[1] https://www.unit.no/

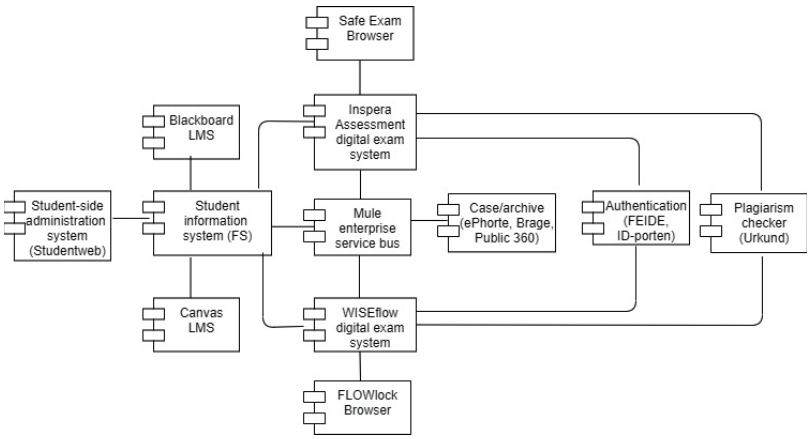

Fig. 1. Exam solutions interfaces [Adapted from (Melve & Smilden, 2015)]

3.2. Research design

The research questions for this study are exploratory, hence a qualitative approach would be most appropriate, as indicated by Robson (2002). A quantitative survey might have been used for this study if the purpose was to compare stakeholder views on features of e-exam tools in terms of ranked preferences. However, the research questions are not about the relative importance of the features in numerical terms, but rather what the key features are and why they are important. As argued by Yin (2002), such research questions are best suitable for case studies. The exploratory nature of the research questions points towards inductive reasoning (Braun & Clarke, 2006; Twining, Heller, Nussbaum, & Tsai, 2017), using the exploratory case study approach (Yin, 2017) to develop theory from the case.

In the literature, case studies have been mainly designed to be as either single case study or multiple-case study (Baxter & Jack, 2008; Gustafsson, 2017; Weishaupl, Yasasin, & Schryen, 2018). Although our study spans several universities and two different system vendors, it is most appropriate to define it as a single case study, rather than multiple cases since the two e-exam systems have been acquired in a joint process run by Unit, for use in various universities in Norway, with both vendors responding to the same set of requirements. The participants for this study were people involved in the larger procurement and development process coordinated by Unit. The responses from interviews did not provide much variation depending on whether a university was using Inspera Assessment and WISEflow – if there were more variation, a multiple case study would have been more appropriate (Yin, 2002), While a single case study does not achieve the same breadth as multiple cases, the single case gives more time for investigating that one case, thus potentially promoting deeper understanding of the subject, as suggested by Dyer Jr and Wilkins (1991). The two system vendors also have customers in other countries, but since only Norwegian HE institutions were included, the case is “e-exam tool support for higher education in Norway”.

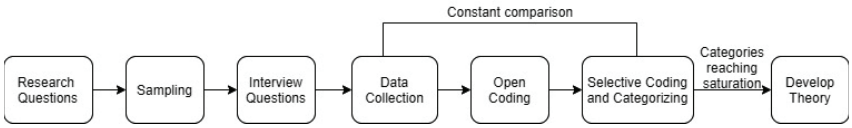

Fig. 2. Figure illustrating case study design

3.3. Data collection and analysis

Fig. 2 illustrates the process of data collection and analysis used in this study. Based on research questions, suitable participants were identified. We used a combination of key informant sampling (contacting persons known to be in central roles, with expertise on the topic) and snowball sampling (getting suggestions from initial participants about other potential participants), with the aim of covering both vendors IA and WF, and universities using each system. All in all, we interviewed n=12 participants from 7 different organizations. Participants (cf. Section 1 & Appendix B) were vendors, process managers, system managers at universities who were primarily involved in the execution of digital exams, and all consented to have their interview data researched and published in anonymized form.

Semi-structured interviews (Kallio, Pietila, Johnson, & Kangasniemi, 2016) were most appropriate for our purpose, keeping some structure for comparability between participants while at the same time allowing for participants to bring forth issues we might not have thought about. An interview guide (cf. Appendix C) was prepared and distributed to participants before interviews. All interviews were done by the first author, during April 2019 – Aug 2020, some face-to-face (4) and some (8) via Skype and Zoom video calls. Each interview lasted approx. 40 mins.

The interviews were transcribed line by line by the first author and coded in NVivo 12, using the constant comparative method, which has the advantage of making the analysis more explicitly theoretical (Urquhart, Lehmann, & Myers, 2010). It was first proposed by Glaser, Strauss and Strutzel (1968) in their grounded theory methodology, and further practically explained by others (Charmaz, 2006). We used it for data analysis in the same way as (Fluck, 2019). The purpose of our study is to mainly focus on answering the research questions rather than identifying emerging theory (Boeije, 2002).

We had a set of predefined categories from our research questions. Hence, our analysis was focused on finding the relation between the concepts that emerged from the analysis, and on grouping those concepts under our predefined categories. First, data analysis involved open coding of the responses using 'in vivo' in NVivo 12. This open coding abstracts concepts from the data. Second, axial coding was performed for grouping of the codes from open coding and further categorising the codes. Lastly, selective coding was conducted to interpret the relation between codes and categories from axial coding. The data collection, coding, and analysis were done together to enrich the existing category. During the coding process, the constant comparison of data and codes was made to compare responses and decide what data will be gathered next until the data get saturated. The coding was done by the first author but discussed with the second author along the way.

4. Results

This section presents the results emerging from the interviews, structured according to the research questions. Table 1 shows the most prominent concepts identified in data analysis.

Table 1. Most prominent concepts identified from data analysis

Categories	Grouped concepts (or codes)
Requirements engineering process	procurement process, requirements elicitation, requirements negotiation, requirements prioritisation, requirements specification
Key features	key functional features, key security features
Security	cheating prevention, cheating detection, cheating threats, security in BYOD exams vs exams in university-owned PCs, technical issues, vulnerabilities in tools, mitigations
Integration and interoperability	integration architecture, integration between e-exam system and LMS, integration between e-exam system and student information system, integration between e-exam system and lockdown browser, security challenges during integrations
Digital ecosystems	content sharing (e.g., sharing questions), monolithic vs digital ecosystem, open API requirements, third-party tools integration with e-exam system, impact of digital ecosystems on security

4.1. Requirements engineering process

There were findings (cf. Table 2) related to several aspects of the Requirements Engineering (RE) process (also called procurement process by participants), e.g., elicitation, analysis, negotiation, prioritisation, and specification of requirements.

Elicitation of requirements was facilitated by UNIT, the directorate overseeing joint procurement as mentioned in the previous section. They collect requirements from system managers in universities who again get input from end-users. UNIT align requirements with government strategy and evaluate e-exam systems based on offers from vendors. Pilots and feedback from universities were given high weight in the comparison of products, and some universities had local procurement processes before UNIT took the coordinating role. However, system managers still asserted that there was no structured RE process followed during procurements. Vendors and system managers used different tools for specifying the requirements, e.g. Confluence as their collaboration software program. Prodpad and Confluence were used by vendors to document customer input, and system managers are using OneNote and Visio for designing of the processes.

After procurement, requirements for future releases have been analysed and prioritized through seminars conducted by UNIT and vendors, with participation by universities. If universities have critical requirements that were not included in the requirements specification, this may lead to contract changes. Requirements differ from country to country, and even between universities in the same country. Products aim to satisfy the generic needs of several customers. One of the challenges mentioned by vendors is to balance needs of various customers. Hence, they will consider the prioritised customer list or perform votes and polls about future features.

Table 2: Findings on the RE process

RE activity	Stakeholder	Statements of Stakeholders
Elicitation	▪ Vendors	▪ "Unit come up with requirements from universities, based on those we provided offer, we also try to have a direct dialogue with end-users." (IA1)
	▪ System managers	▪ "We have had pilots on IA and WF, and in total, there were only four vendors [to choose from] at the start. So we [gathered] experiences of other schools, and then we tested it ourselves, but a lot has changed since [2015]" (NTNU2). ▪ "We don't have a structured RE process yet, that's something we would like to do. We collaborate with […some Norwegian universities] and Unit and write user stories. Then we [discuss] with vendor." (NTNU1)
	▪ Process managers	▪ "We ask universities about their requirements and align them with government strategy. We will have bi-weekly meetings with IA and WF together, to agree on workflow and integration for everyone so that we don't have to make a separate integration for every university." (Unit)
Analysis	▪ Vendors	▪ "We have webinars, user groups, annual seminar with customers for analysing, negotiating and prioritising which features to develop further in which order." (IA1) ▪ "We negotiate in multiple ways. We have to negotiate to change contracts because […. when] the contract is a year or two old [… customer] needs might have changed." (IA1) ▪ "Requirements prioritisation is always tricky. Because there might be functionality that's important to one segment that isn't important to another customer segment." (IA1) ▪ "We prioritise [...] based on votes and polls." (IA2) ▪ "We don't build stuff for only one customer, we build it for everyone" (WF)
	▪ System managers	▪ "We had the local procurement in the beginning before national procurement. There, we gave points on requirement specifications, and WF ended up with the best score." (HVL).
	▪ Process managers	▪ "There haven't been that many conflicts. When something developed not in line with our requirements, we have dialogue with the vendors. We discuss on checkpoints. [..]" (Unit)
Specification	▪ Vendors	▪ "We use many different tools, e.g., 'Prodpad' to get customer input and the final roadmap, different templates to define the requirements and acceptance criteria in confluence. For public tenders, we use 'UPO', which is a specialised tool for RE processes." (IA1)
	▪ System managers	▪ "Unit is more like a facilitator rather than specifying our requirements themselves" (NTNU1) ▪ "We use OneNote and Visio. When we collaborate on the documents with other universities, we have one institution responsible for that area, and it is the editor, but always others can contribute. Then we move requirements to vendor's wiki and get a specification." (NTNU1)

Regarding the similarities and differences in procurement process between countries, IA vendors and process managers mentioned that only Sweden has a similar setup as Norway with 'Sunet'. In contrast, UK has certain de facto standards concerning LMS systems identity federation (eduGAIN), but no central authority organizes procurement. Process managers especially mentioned that International organizations, Geant (Europe) and

NORDUNET (Nordic countries) attempted to conduct joint procurements. Vendors and process managers thought that other countries could benefit from considering a procurement process similar to the Norwegian one:

"There are benefits from standardizing workflows and integrations, for reaching many customers with the same solution." (IA1)
"I think it requires a culture of sharing and collaborating rather than competing in addition to a government requirement to do so [..] and depends on the Higher education sector attitudes, autonomy, and willingness to work together (even though in some cases it will result in development taking longer) rather than competing." (Unit)

4.2. Key features and qualities for e-exam systems

Functional features of e-exam systems face various user groups and address several process stages related to the exam. Table 3 shows functionality mentioned in interviews sorted by key user group (students, teachers/censors, administration) and process stage (before, during, after exam). Some of the cells are empty, as no such features were mentioned as important by the participants.

Table 3. Functional features stressed by participants, sorted by user group and process stage

User group \ Process stage	Before exam	During exam	After exam
Students		Answer questions	Receive grades
		Upload documents	Seek, receive explanations
		Deliver exam	Appeal grades
Teachers (and censors)	Authoring (of tests, questions)		Grading
	Upload documents		Explanations
			Analytics
Admin	Logistics	Monitoring	

Table 4. Findings on key functional features

Key features	Stakeholders	Statements of Stakeholders
Authoring	• System managers	• "In WF, you make a flow based on what kind of exam you're holding, e.g., FLOWlock, FLOWassign, whereas in IA, you make the exam and add features to exam based on what kind of exam you're holding." (UiT) • WF's authoring tool only supports FLOWmulti exam." (UiT)
Logistics system	• Vendors	• "Scheduling of exams, including things like the number of students and rough estimation of requirements for that room, e.g., power, special software, and grant access for students with disabilities." (IA2)
Question analytics	• Vendors	• "Ensure that exams that can be provided are fair, accurate, have relevant question types, help users providing good questions and give insight into the difficulty of questions and ability of students." (IA1)
Grading	• System managers	• "WF is more user-friendly than IA. it wasn't easy to see the whole grading process in one page in IA, but in WF it's easier to keep overlook of the whole." (KU)
Explanation of grades	• System managers	• "IA supports explanation for the test but not for the whole course since there is no synchronisation between FS and IA, e.g. if the project weighs 30% of the grades and final test weigh 70% of the grade." (NTNU2) • "Currently, there is an opportunity to add an explanation in IA, but it will not notify the student, so students have to log on to IA to check the explanation." (NTNU5) • "At the moment, censor may write a comment in WF, and share it with a students, soon they will see it when the grade is published. But it's not automatic yet." (HVL) • "At UiT, WF has been integrated with FS for grading explanations. So, an explanation written in WF can be exported to Studentweb now." (UiT)
Appeals and complaint grading	• Vendors	• "We have implemented appeal function in IA, which can manage requests for appeals, appeals grading, and invite graders into IA for appeal grading. It's based on integrating additional data from FS." (IA1)
	• System managers	• "Students request appeal from Studentweb then it goes to FS. When examiners get notified from FS, and when it is graded in WF, it goes back to FS. Students find the grade in Studentweb." (HVL)

E-exams need several different questions types, ranging from multiple choice and other auto-scored formats (where WF's question types are based on Learnosity[2]) to free-text essays and file uploads. For

[2] https://learnosity.com/

the writing of lengthy text, some participants favoured a limited scope of support from the e-exam system. A mainstream word processor would have better functionality than a more limited text editor inside the e-exam system, hence it was perceived more user friendly to write the text outside the e-exam system and just upload the document: *"Most of our teachers except a few in engineering and health subjects, don't use that [the in-built editor]"* (HVL) *"Most of our exams are made outside WF system and are uploaded into WF as a pdf, and students write answers using a word processor in FLOWlock."* (UiT)

Some participants were familiar with both IA and WF from procurement elaborations, often comparing the two systems in their answers. Generally, participants tended to focus less on well-established functionality, more on recent and maybe challenging features. Especially, many talked about grade explanations and appeals, previously done outside the e-exam systems but recently included in the feature scope of e-exam systems. Examples of statements concerning functional features can be found in Table 4.

Table 5. Findings on key non-functional features

Key features	Stakeholders	Statements of Interviewees
Scalability	• Vendors	• "Earlier IT has not involved with exams other than special needs accommodation, so our involvement has been different than before, e.g., instead of 100 people, we have 1,700 at one time." (IA1)
	• System managers	• "E-exam system can allow as many students, e.g. 1500 students to take a different type of exams at the same time on the same day." (KU)
Usability	• System managers	• "It's easier for students to download, use, update, follow system requirements, and find the information they need in WF, but in IA the steps are difficult when you're not known to the system. In contrast, the dashboard is better in IA." (KU)
Integration and Interoperability	• Vendors	• "From the content point of view, questions and data in IA can be exported and can be moved to other exam systems and the same for submissions that they can be archived and exported in standardised formats." (IA1)
	• System managers	• "IA is integrated with 'Brage' [publication system for academics and research] and 'ePhorte' [archiving system for exam data], but most of the public sector uses public 360." (NTNU1) • "Integration between FS and WF transfers student and assessors' data, exam times, assessor deadlines, exams relevant data from FS to WF, and WF to FS such as grades." (UiT)
Security	• Vendors	• "Currently IA auto-save exam data in cloud storage by Amazon web services." (NTNU1) and "[…] same for WF" (WF) • "Security depends on the type of encryption and its length." (IA1) • "IA exam data is hosted within a virtual private cloud of Amazon web services. So the data has limited access, firewalls protection, encryption, logging, and data lies in redundancy in three different physical locations." (WF)
	• System managers	• "Feide authenticates students and internal censors, and we use ID-porten for the external censors. But sometimes when we have foreign censors, we have to send them a link through WF to access that specific exam." (HVL) • "We can also use access tokens, but we try not to use it unless external doesn't have a Norwegian social security number. But sometimes, we have to give the access token to faculty, e.g., when systems or routines are so slow." (KU) • "WF supports in-built monitoring feature called FLOWmonitor for monitoring exams." (HVL)
Reliability	• Vendors	• "We manage and execute the exams securely from end-end in a stable manner. Because when something happens, that obstructs the exam then reset of the exam is very cumbersome, and expensive for customers." (WF) • "There is an allocated time and place for the exam to happen. Otherwise, the exam might not be fair to everyone or able to be completed." (IA1) • "When a submission is handed in, it is being reflected as test taker intended, and evaluator can see what the student intended for them to see." (IA1)

The key non-functional features required for e-exam systems were found to be scalability, usability, integrity and interoperability, security, and reliability (cf. Table 5). Scalability is important due to a huge load of exams in the peak period, some with large classes. Usability is essential as end-users are quite diverse, some with limited computer skills. We received more responses on integration and interoperability between systems, which we provide in detail in Section 4.3. As for integrations and

interoperability, participants felt integration of exam systems with student information system and archival systems were the most important.

Security was seen as essential for the validity of the exam results: *"To make sure that data is authentic and secure both in terms of securing the data of the ones taking the exams when somebody is breaking in to tamper with exams and test-takers breaking out [of the lock-down] during exams."* (IA1). Security and reliability are also crucial due to the cost of redoing an exam: *" it has to be very secure because it's challenging to do it over again if something went wrong with exams and grades".* (NTNU4). As can be seen in some of the quotes in Table 5, authentication of graders through single-sign-on was considered to have good security, while less secure ad hoc solutions like access links had to be used for foreign graders unable to use the national single-sign-on systems. Some participants considered WF's closed source FLOWlock browser more secure than the open source SEB browser used by IA: "*SEB being an open-source browser is however a problem we will have to deal with for as long as we still use the open source browser."* (NTNU2)

Table 6. Findings on detection and prevention of cheating during onsite invigilated exams

Findings	Stakeholders	Statements of Stakeholders
Prevention of cheating	• Vendors	• "The lockdown browsers close a lot of security holes, but there's no perfect solution for BYOD. Some organisations use computer labs to have a higher degree of security." (IA1)
	• System managers	• "If the student goes out of the FLOWlock, then invigilator has to help him with a password. So, we have not had any cheating cases that I know." (HVL) • "FLOWlock of WF works like SEB, but it is closed source. We experienced some problems with SEB that's why we changed to WF. We had more cases of cheating in SEB." (KU) • "If students are going to use MATLAB, then we will conduct exams in computer labs or VDIs." (NTNU2) • "We locked out our school computers with ad-blocker, proxies. We experienced fewer issues." (NTNU1) • "We have test computers for Mac OS, Windows, Linux. We tested if running exam from different virtual machines is possible, some of them worked, but we only test what is reported, more responsibility lies at Inspera and SEB." (NTNU2) • "It is always possible to hack something using internet, especially using third-party tools. So, vendors have had data protection agreements with third-party solutions." (NTNU1)
	• Process managers	• " Lockdown of the network can prevent students from sharing files when they use Excel in BYOD exams, but it is a challenging part of closed-book exams. it is better to change the type of exam to open-book exams to avoid cheating. (Unit)
Detection of cheating	• Vendors	• "We do case analysis, and flag if there is suspicion, e.g., copy-pasting or sudden increase in text. We log and store every response. There are not too many false positives, but we need to make sure that text is not copy-pasted from the question or the response." (IA1)
	• System managers	• "We have seen Macro keyboard and Mac book touch bar save something to the clipboard, but SEB wipes clipboard. SEB also detects if a large amount of text is dumped through macro, but if student use speech to text, it will not be able to tell the difference." (NTNU2) • "There is a monitoring feature in WF, but sometimes it doesn't indicate cheating, e.g., in WF, sometimes when students don't get the exam text up, they will copy-paste exam text into their text, and it gets the detection that students copy-pasted a lot of text." (KU) • "There's a built-in feature in IA that enables to get the list of active students when events triggered, e.g. when they go offline, screen share, copies a large amount of text. But when many events triggered, it's hard to differentiate what was the real issue." (NTNU2) • "A few students had manipulated certain files, and they got it to work in demo test but [..] not when they had their exam. Because we had a few students that have gotten error messages for SEB config file and config tool was in the recently viewed steps." (NTNU2)

Participants were reluctant to reveal concrete ways of cheating but indicated that the system would not be 100% secure: "*We don't want to share […] what we know about how you can cheat. We […] caught one student who [utilized] one of the known issues that we have had in WF. That issue was fixed. […] However, I'm sure that a hacker student would probably be able to do something with the coding of any program installed."* (UiT)

Table 6 shows a number of quotes related to the mitigation of cheating during exams. One finding is that mitigation of cheating has a trade-off with scalability, e.g., it is risky to allow usage of third-party

tools like MATLAB or Excel in BYOD exams, so these must rather be held in computer labs – but many institutions do not have sufficient capacity: "*University PCs might be secure, but we don't have labs, so we only use BYOD laptops."* (KU)

Some of the BYOD exam cheating practices mentioned during interviews include code injection using Macbook pro touch bar and Macro keyboard, speech to text, manipulating configuration file, modifying open source code, cheating via SEB configuration tool and third-party tools. Some of the mitigations mentioned include logging, penetration testing at vendors-site, using lockdown browsers, using ad-blocker and proxies for university computers, using virtual desktop infrastructure (VDI), in-built monitoring feature from the e-exam solution, and using passwords from invigilators. Participants mentioned that students also use cheating approaches which are totally outside the BYOD exam laptop, such as cheat notes, communicating through mobiles, wireless earpieces: "*We don't have many [discovered cheating] cases on invigilated onsite exams because it's difficult to find out cheating, e.g., with a note in their pocket and when reading it in the toilet, use their mobile phones when they are not allowed to use."* (HVL) and "*it happens outside of the computers like writing on the notes or hiding a phone in the bathroom. But that will happen if the exam is on the paper or if it is on their computer."* (KU).

Universities in Norway mainly depend on human invigilators for monitoring students during exams, typically retired elders who are hired part-time for the exam season. Biometric surveillance, though available in WF, is not used out of privacy concerns: "*We haven't used the facial recognition function of WF at HVL because we have the invigilators checking the ID of the student"* (HVL) and "*WF has face recognition functionality, but we don't have that because of regulation laws, and I think nobody in Norway uses that functionality."* (KU)

4.3. Digital ecosystems

Participants addressed different aspects of how e-exam systems might fit into a larger digital ecosystem, such as integrations between e-exam systems with other information systems that the university already has (cf. Fig.1), support of third-party tools by e-exam systems, well-documented Application Programming Interfaces (API) to enable easy integration, support of third-party plug-ins, and usage of standard data formats to enable sharing of content between systems and universities.

There was no overarching standard or framework used by vendors for integrations. Instead, different standards, practices and a collection of APIs have been used as a framework for integrations:

We follow a lot of best practices and standards. Data are often presented via different industry standards, e.g. IMS QTI or OneRoster. We use REST API's and provide Swagger documentation which is a widely used way to document APIs that sort of serves us as the framework. It is not an ISO standard as such, but it is sort of one of the industrial practices for documenting. But well-established standards often ease the integrations with new systems. (IA1)

We build our own framework and then use some technology standards depending on the customer systems that we integrate with. (WF)

Our developers use MuleSoft for the integrations, which transfers the information from exam systems, FS, and other systems to the right place. (Unit)

Vendors mentioned that they use REST API's for in-house integrations, whereas process managers mainly used Mule soft Enterprise Service Bus (ESB) for API integrations of different systems at a time.

Several challenges with integrations were reported during interviews (cf. Table 7). First, some of the integrations were not standard based yet, e.g., SEB lockdown browser was integrated at the functional level of IA rather than full integration between these two systems. Integration with the FS administrative information system was complicated both by low data quality and synchronization issues, many of these problems were on the side of FS due to lack of features and interfaces. System managers also thought that non-uniformity in choosing solutions by universities for integrations have led to slower implementations.

Table 7. Findings on challenges of integrations between e-exam systems with exam supporting technologies

Findings	Stakeholders	Statements from Stakeholders
Integration is not standard based	• Vendors	• "We control more of the ecosystem. Integration with SEB isn't standards-based, so in that sense, documentation is not as good as standards-based." (IA1)
Low data quality	• Vendors • System managers	• "There had been challenges with data quality." (IA1) • "The quality of the registration of data in FS is not uniform enough to do all the automation NTNU would like to do." (NTNU1) • "Duplicate users for both students and employees may happen." (NTNU2)
Non-uniformity in choosing solutions for integrations	• System managers	• "Biggest issue is that larger institutions that use IA have integrated their central user account with IA whereas NTNU integrated their central use account directly only to FS." (NTNU2) • "Since we have chosen different solutions [..] for developing solutions for the same purpose, and it takes a long time to implement changes." (NTNU2)
Poor synchronisation between systems	• System managers	• "Sometimes the integrations won't export the right sources from FS to WF due to synchronisation issues, e.g., synchronisation doesn't do the changes when grading commission is changed." (KU) • "Slow process of integration between FS and IA due to loading of the large database." (NTNU4)
Incompatibility of QTI formats	• Vendors	• "Integration of IA and Blackboard is not supported due to incompatibility in their QTI formats." (IA1)
Integration of exam system with LMS is not prioritised	• Vendors • System managers	• "We did not receive requests from our clients." (WF) • "Integration with Blackboard is not something we prioritise because we're not going to send any results from Blackboard to FS or from Blackboard to Inspera. But a lot of teachers have said that why should I have to make questions twice, so that's why we're looking at." (NTNU2) • "There is no integration between WF and Canvas. I don't see any need for that rather what I do see a need for better integration between Canvas and FS if that would exist then all necessary integration between the LMS and exam system would go through FS." (UiT)

According to participants, neither WF nor IA have been integrated with the LMS (Canvas or Blackboard) although many teachers have expressed wishes for such an integration (e.g., wanting to use the same pool of questions both for formative and summative assessment). One key obstacle for the integration is incompatibility between their Question and Test Interoperability (QTI) specifications. As an alternative solution, these e-exam systems support interoperability through IMS Learning Tools Interoperability™ (LTI), to launch e-exam application as a tool from LMS. But this received low priority at universities:

Currently, this has not been highly prioritised in Norway. So, we have not started working on integration. Norway has somewhat different regulations, thereby mandating that assignments that count towards final grade should be conducted in a specific assessment system and not in the LMS and wanting to avoid confusion for students of what system is to be used for assessments and assignments. (Unit)

In this paper, interview responses are presented according to categories (themes) and user group (i.e., stakeholders) in tables 2,4,5,6,7,11 since the responses are related to different categories associated to main themes mentioned in the captions of those tables. Whereas tables 8, 9, 10 represent only statements according to user group since all the statements are directly associated to the main theme.

As for access to third-party tools from the two e-exam systems, as indicated by statements of our interviewees in Table 8, IA has the option to configure the third-party tools in its SEB browser. This functionality is similar in WF's FLOWlock exam. However, universities instead run open-book exams or use their computer labs if the test requires usage of third-party tools (e.g., MATLAB, SOLIDWORKS, GeoGebra) during BYOD exams. Alternatively, both IA and WF support whitelisting of websites and web tools (e.g., EXCEL), using this practice, teachers can allow students to access external material during lockdown exams. Some universities have already tried whitelisting Excel, but this created a security vulnerability so that students could access files and desktop. At the same time, constraints and

increased cheating vulnerabilities around the use of this feature made system managers avoid it. According to participants, the need for using third-party tools depended much on the discipline. Those who had mainly essay exams (e.g. social sciences) or multiple-choice tests (e.g., written exams in medicine, to supplement clinical/oral examinations done outside the e-exam system) were largely satisfied with standard features of the e-exam systems. Whereas, STEM subjects were less satisfied, needing, e.g. design, programming and math tools to make exams more authentic. Regardless of discipline, there might also be students with special needs for which extra tools would be required. Some of the system managers found in-built Virtual Desktop Infrastructure (VDI) feature in e-exam systems would be the best solution for accessing third-party tools; however, they thought it would require extensive testing to assure the usability, stability and security of high-stakes exams with VDIs.

Table 8: Findings on third-party tools access during onsite invigilated e-exams

Stakeholders	Statements of Stakeholders
Vendors	• "Universities have a certain number of licenses to use third-party software. So, students have to be on a university computer. We do not sell the third-party license, but we do integrate with some of these." (IA2)
System managers	• "Third-party tools are added through the config file, we use mainly PDF, and we can also use whitelisting, e.g., some X URL, but behind that URL there is a static list of rules we are going to use, so we don't use that function." (NTNU2) • "WF doesn't support third-party integration as of today, but they support whitelisting. We used Excel through whitelisting, but students were able to access the other files and desktop that we cannot allow during our exams." (KU) • "Students with special needs must access third-party tools. Currently, they deliver exam on Word, and we will upload the file, that's better in IA than WF." (KU) • "When we need third-party tools in onsite exams, we conduct open book exams. So teachers have to choose what is more important, is it to test the students what they learnt in the course or is it that it's 100% secure." (HVL) • "We don't use FLOWlock when we use third-party tools. But there is a project currently going on to access VDI in FLOWlock, giving you access to third-party tools. but the question is, are VDIs stable enough to handle many students at the same time? […] what happens if it breaks down? How about user experience? Another solution would be to make specific third-party integrations with specific programs, e.g., MATLAB, GeoGebra." (UiT)
Process managers	• "Vendors mostly do integration of tools, a lot of people want to integrate Excel in WF, the problem is that when students use excel sheet, they can share it" (Unit)

Table 9: Findings on availability of open APIs for IA and WF

Stakeholders	Statements of Stakeholders
Vendors	• "Currently, our API's are designed to enable data flow with the admin side of IA mostly, but additional ways of extending and customising are possible for universities through IMS PCI standard." (IA1) • "There are primary advantages by opening APIs, so we're working on adding support for APIs to more parts of the application, but we have to ensure that they're being used correctly." (IA1) • "Our APIs are tightly controlled, and we don't let people query things against our database. To do integrations, developers have to work with us. You get normal standard problems with our APIs that you get with regular APIs security." (IA2).
System managers	• "There is a national project going on with Unit, and it is looking into it to find a good solution through API. Every institution that uses WF gets the API through Unit, but we haven't done much with it yet." (KU) • "Security always depends on the level of documentation of open API. Currently, Inspera's REST API's help with the administrative workflow." (NTNU1) • "Working on APIs would be interesting to me. We have access to do that if we saw the need to tweak the API in some cases to fit some of our needs. but that's not something we've discussed at UiT." (UiT) • "Inspera is using PCI standard to create exercises, and they're open to enable others to do it as well, but we haven't explored that, and it is something our teachers are working on." (NTNU1)
Process managers	• "In general, having open API's allows people with competence to create applications, but specific considerations have to be kept in mind when opening them up to ensure that they are being used correctly." (Unit) • "Currently, API's are not open enough to pull out real data because of not having access to the access token." (Unit)

Vendors and system managers thought open APIs would enable universities to integrate third-party tools or develop plug-ins (cf. Table 9), for instance, using the Portable Custom Interactions (PCI) standard. However, the current APIs of IA and WF were felt not be open enough, and system managers of IA at

NTNU thought currently available APIs only facilitated administrative workflows, and also thought it was not within their capacity to develop plug-ins.

Table 10: Findings on opinions on monolithic vs digital ecosystem

Stakeholders	Statements of Stakeholders
Vendors	• "It is an absurd idea to say one piece of software will do everything for the exam. The ecosystem is the only ideal choice for the universities because the idea of monolithic does not work" (IA2) • "From a high level, we have to focus on an ecosystem because there are so many different needs that we cannot possibly fulfil them all. The different systems are in use already a handful in specific tasks, so it makes sense to continue using those by integrating with the rest of the ecosystem." (IA1) • "When you have a combination of companies work together to achieve a goal [..] it might be slower to implement because each of these companies has their own agenda. However, with a monolithic solution, they have the same agenda. In theory, they have many processes digitised, so when customers get an incident, they can immediately send it to the vendor, and they can immediately look at it." (WF)
System managers	• "We have talked a lot with teachers about what system they would like, and what's become very apparent is that a lot of them want completely different systems. Blackboard has the functionality to perform exams as well, and we could use IA for delivering assignments, but in the end, IA is designed as an exam system and Blackboard as an LMS." (NTNU2) • "We get requests when teachers see a certain functionality in Canvas that does not exist in WF. They wish to use Canvas for a certain part of their exams, but that system again lacks all other necessary functionality to conduct an exam." (UiT) • "People always want more features, but the more features you put into a system, the more difficult it will be to use one system. It will be simple to use multiple systems because there are so many ways to do things already in Blackboard and IA. If you combine them into one system, people will struggle with finding the right option for them." (NTNU2) • "The most important thing is providing a good tool for our students and our employees, having well-standardised integrations and being able to plug-in from another system into e-exam system." (NTNU1) • "Having ecosystem would be a better solution. […]. Even it would benefit vendors. I would be surprised if they would not be open to working with other parties" (UiT)

As for using monolithic vs digital ecosystems (cf. Table 10), vendors mentioned an ecosystem is an ideal choice for e-exam toolsets, as a single tool cannot provide every needed functionality, but they feared more issues with availability when third party tools are integrated within the ecosystem compared to the monolithic system. System managers believed ecosystems would improve integration and benefit users and vendors. One of the reasons for not choosing monolithic system is that teachers want specific features from different systems, e.g., from LMS and e-exam solutions.

Table 11: Findings on factors that influence sharing questions among universities

Factor	Stakeholders	Statements of Stakeholders
Security	• Vendors	• "If we open up the ability for users to share through exam system, there is certainly a concern of users sharing the stuff they are not allowed to share, so a degree of access control should exist." (IA1)
	• System managers	• "I wouldn't recommend open sharing between the universities like opening it up to everyone who has authoring access to the system at any university. Because you would have to start trusting everyone else's security measures that would affect your security." (UiT)
Culture	• Vendors	• "It depends on the culture of the university as to whether sharing is important thing or not. In Norway, there is more share of information. Every year all the questions for the previous year's exams are available. If professors know each other […], they may talk about some of the old exam questions." (IA2)
	• System managers	• "Universities need to have a good collaboration outside the exam if they want to create the exams together. E.g., In medicine, they share questions across universities, and they had national tests where questions were imported into IA to conduct exams locally." (NTNU1)
Sharing between Cross-platforms	• System managers	• "We're involved in several national exams where universities use the same questions in different exam systems. They are made in a particular format by the national groups, but we have to do a lot of work after they've been imported to have them ready for the exam day. It will be a major development if it's possible to share questions between the different systems in a more dynamic and real-time way. (UiT) • "Having a shared database in cross platforms would be a major step for not only to cooperate and sharing questions, basically sharing the load." (UiT)

Currently, both IA and WF support the functionality of sharing questions within the university. But sharing outside university requires export/import questions through a QTI file. Participants thought that

access control would be a significant problem if e-exam systems support sharing across universities (cf. Table 11). Moreover, vendors mentioned that the culture of the university was the main factor that encourages sharing. A system manager said that if, e.g., medical faculties create questions together for national exams, that would require questions to be made for different exam systems, requiring extra work. So, a shared database would enable easier collaboration within and across universities.

5. Discussion

5.1. Interpretation of findings

1. ***What process is followed by vendors, process managers, and system managers to identify features and agree upon requirements?***

The interviews indicated that the identification and negotiation of features and requirements is an important process that needs a lot of attention. This is in accordance with literature in software engineering in general (Dick, Hull, & Jackson, 2017), and in digital ecosystems (Immonen, Ovaska, Kalaoja, & Pakkala, 2016), and e-learning ecosystems (Uden et al., 2007). Some e-exam systems requirements may differ between countries, institutions, disciplines, and assessment practices. Yet, the interviews revealed that the requirement process was somewhat ad hoc, and no standard requirements process seemed to be followed across various stakeholders. This is also in accordance with findings in the literature. For instance, Achimugu et al. (2014) observed that although several advanced methods for requirements prioritization have been proposed in the literature, these have had limited uptake in general software development practice. The stakeholders who addressed requirements prioritization seemed satisfied with their rather simple approach based on polling and web meetings.

A study from Norway by Foss-Pedersen & Begnum (2017) looked especially at universal access requirements for e-exam systems IA and WF, in a study involving some of the same organizations that we interviewed. Their findings indicate unclear division of responsibility between vendors and buyers for ensuring universal access and lack of quality assurance through user testing. Some of our participants did mention getting useful feedback from piloting and user testing at universities. This may indicate that the process has been somewhat improved after 2017, which is natural since more widespread use of the e-exam systems will have generated more user experience. To some extent, our findings indicate a clearer assignment of responsibilities between vendors, universities and system managers, thus approaching a clearer definition of roles as recommended in the literature (Immonen et al., 2016).

The arrangement to have one national organization – Unit – oversee the requirements and integration process, seemed to be satisfactory to the participants. No participant indicated a wish to go back to the previous situation when each institution would run separate procurement and integration projects. Unit maintains a requirements specification for e-exam systems for Norwegian institutions, negotiates with vendors on universities' behalf, and takes responsibility for developing integration software to make the e-exam systems fit into the common system infrastructure. Similar approach is used for Sweden and Netherlands higher education institutions (Boezerooy, Cordewener, & Liebrand, 2007). An interesting question is whether a similar collaboration could work even internationally. As indicated by vendors, some requirements will differ from country to country, such as those relating to the national IT infrastructure and those relating to laws, regulations, and student rights concerning exams and grade appeals. On the other hand, many features would be of interest regardless of country, such as support for specific exam types, question types, disciplines. E.g., a maths exam in country X would likely have more common needs with a maths exam in country Y, than with a history exam in country X. However, collaborating between several universities to establish requirements for e-exam systems as a basis for acquisition from commercial vendors is not the only viable path for international collaboration. Another option would be to have multiple universities collaborate on developing e-exam technology, as has been pursued, for instance, by the EU project TeSLA (Okada, Noguera, et al., 2019). However, this alternative was not suggested by any of the informants, likely because none of the Norwegian universities were involved in the TeSLA project.

2. ***What do vendors, process managers and system managers see as key features for functionality and security in e-exam systems?***

Explaining all the needed features of an e-exam system would take a lot of interviewee time. It is thus reasonable to assume that participants focussed on significant features that they found most worthwhile to mention, based on their experience. It was also noted that they spent more time discussing recent and perhaps challenging features, than more straightforward features. It is a well-known phenomenon in the field of requirements engineering that assumedly obvious features will tend to be omitted by interviewees, especially if they are experts on the technology in question (Firesmith, 2003).

Of the features that participants did talk about, there was comparatively more mention of authoring, explanation of grades, and appeals grading, less about the logistic system, question analytics and ordinary grading. Of course, ordinary grading (for all students) is a much more used functionality than appeals grading (for the smaller fraction of students who dispute their grades), so when interview participants talked more about appeals grading, this again is likely an effect of omitting the obvious. Moreover, appeals grading had probably gained extra attention from many participants because it was initially unsupported by the e-exam systems, thus addressed by a cumbersome manual workaround, and had just recently been included as a feature.

Our results on key features are in line with previous findings (Kuikka et al., 2014; Striewe, 2019). The features highlighted by our participants relate to all four major components proposed in the framework by Striewe (2019): front-end components (user interface), educational components (underlying logic of tests and question types, pedagogical modules), knowledge representation and storing components (e.g., question pools, tests, exam answers and results), connector components (related to integration and interoperability, as discussed in our section 4.3 on digital ecosystems). Admittedly, the development of features for knowledge representation is currently much less advanced than suggested by Striewe. Kuikka et al. noted that none of their compared systems satisfied all the needs of teachers. Likewise in our study, participants indicated that features with less backing would end up not being prioritized. Also, teachers would sometimes have to make trade-offs, such as only being allowed to have open book exams if providing usage of third-party tools. The expert survey by (Isaias et al., 2019), found authoring, interoperability and feedback features to be important, this corresponds to our findings, but participants in our study felt feedback as the basic feature. Different from the situation in Norway, Kuikka et al. suggested that it is an advantage to have the same software product work both as LMS and e-exam system. This may be true, but some of our system manager participants believed otherwise, thinking that too many features in the same system would confuse some teachers.

Among the non-functional features, usability, integrity and interoperability, and security were considered as essential for e-exam systems. Vendors also felt scalability and reliability were important for their customers as the BYOD exams were mainly run on the internet. System managers considered the availability of technical support from vendors during exams to be vital for them.

3. What are the goals and challenges concerning integrations between the e-exam system and other exam supporting systems?

Several integration challenges are presented in this study. There was functional level integration between a few systems, e.g., IA exam system and SEB lockdown browser. Although there was no issue explicitly mentioned with this particular integration, specification of this particular integration was mentioned as not as good as standards based. There had been challenges with data quality and synchronisation were reported due to universities in user group used different solutions for integrations. Hence, there should be a need for data management and governance before integrating e-exam systems with other systems.

System managers mentioned that teachers were concerned about content interoperability between LMS and e-exam systems. Both WF and IA e-exam solutions use IMS QTI specification for sharing and exchanging of the content. But question types do not follow a common standard, meaning a QTI exchange would be very complex and manual, and not fully supported across systems. The lack of common standard format for representation of questions was seen limiting the exchange of questions between LMS and e-exam solutions. This is in line with findings by several others (Boussakuk, Ghazi, Bouchboua, & Ouremchi, 2019; Tomberg, Savitski, Djundik, & Berzinsh, 2017). The lack of interoperability standards for LMSs to communicate with external applications due to their monolithic

architecture has received more attention in recent research (Alier, Guerrero, Gonzalez, Penalvo, & Severance, 2010; Chirumamilla & Sindre, 2019; Dagger et al., 2007).

Moreover, our results show that low prioritisation of integration between LMS and e-exam systems and limited demand for such integrations from HE institutions seemed to be the reason for not having upgrades in QTI specifications of LMSs towards better interoperability with other systems. Such interoperability shortcomings of LMS's were pointed out already by Sclater (2007), arguing that the standards and specifications had not reached a critical mass of adoption due to the insufficient demand by users in higher education institutions. More than a decade later, although the uptake of such systems has increased enormously, the progress concerning interoperability is not too impressive. Literature shows that only in the Netherlands, their primary education school system has considered integration between LMS and e-exam systems as a mandatory requirement in procurement of e-assessment tools (Kerssens & Dijck, 2021). While it would also be useful, e.g., to be able to reuse test questions across several platforms (Chirumamilla & Sindre, 2019), participants indicated that so far, this had not been considered important *enough* to make the priority list, as there had always been other issues needing more urgent attention. Both IA and WF vendors also support the IMS Learning Tools Interoperability™ (LTI) standard for seamless integration and secure data communication between several platforms. But the lack of personnel resources and time for testing seemed to be the reason for system managers not to pursue increased use of that functionality.

4. ***To what extent do the stakeholders envision a move towards a more open digital ecosystem for e-exams, and would this be assumed to impact security?***

Both IA and WF vendors are developing RESTful API's for flexible integrations of e-exam systems with other systems and allowing institutions to build plug-ins on their platforms. They followed effective API development and integration practices for e-exam systems adapting to institutions pedagogical requirements. However, e-exam systems and supporting systems use different APIs, so it is not easy to integrate multiple APIs of different systems. Moreover, API's are strictly controlled for security reasons, in literature this has been addressed as a strategy for limiting access to external users and protecting the system from malicious components (Striewe, 2019). So, integrations of e-exam systems with other systems were mainly performed at the process manager's site using a standardised third-party tool, Mule Soft ESB. MuleSoft serves as the framework for service organisation and orchestration, providing secure delivery of services (Menge, 2007). Still, integration of IA and WF with third-party tools seemed to be at the early stage. Both IA and WF support minimal external software during BYOD exams, requiring whitelisting or integration. Unit is running pilots, to use third-party software/tools in lockdown browser. So, currently, third-party tools can only be accessed at the cost of reduced security from the lock-down browser. Though the literature reports alternatives to lock-down browsers to allow third-party tools while maintaining good security, such as booting a dedicated operating system from memory stick (Fluck, Pálsson, et al., 2017; Frankl et al., 2012), this approach is not used with the e-exams used in Norway. One reason could be that the booting approach is more cumbersome, requiring additional actions at the start of every exam, plus preparations in copying a sufficient number of memory sticks. Participants instead suggested a VDI solution for allowing third-party tools without getting too high cheating vulnerabilities but admitted this would require extensive testing and implementation costs not yet undertaken.

There was clear collaboration reported between institutions, process managers and vendors in implementation, integrations, user testing. All participants preferred moving towards a digital ecosystem rather than a monolithic system, as this could provide more features and user convenience for teachers and students to organize learning and tests, better interoperability between systems, increased quality, reusability, and sharing of questions. This is in line with findings in (Kerssens & Dijck, 2021; Veiga, Campos, Braga, & David, 2016). Our results show that security, culture, sharing between cross-platforms seemed as factors that influence sharing questions among universities. On the other hand, system managers in our study suggested that more sharing within and between universities would be possible with the integration of question banks. This is in line with the previous findings (Chituc, Herrmann, Schiffner, & Rittberger, 2019; Laine, Sipilä, Anderson, & Sydänheimo, 2016). However, interoperability and IPR issues remain possible obstacles for the development of shared item banks (JISC, 2007).

5.2. Limitations of this study

5.2.1. Credibility (internal validity)

Several threats that are relevant to discuss.

Selection bias: The selection of whom to interview may have affected the results. Our approach to mitigate this threat was to interview a considerable number of persons, from several different organizations.

Participant bias: May the participants knowingly or unknowingly have given inaccurate information? As already mentioned in the discussion, some very obvious features of e-exam systems were hardly mentioned – omitting information that was taken for granted is a well-known phenomenon. There may also be other reasons to reply inaccurately, such as memory, embarrassment (if something went wrong with the system or project) or secrecy (e.g., vendor representatives not wanting to reveal business secrets). Again, interviewing several persons will reduce this threat. Moreover, we have explicitly reported cases where participants were reluctant to answer about cheating vulnerabilities and concrete ways to utilize them.

Researcher bias: Researchers may tend to interpret data in ways that confirm their preconceived ideas. Various measures taken to mitigate this threat include avoiding leading questions, not pushing participants in any particular direction, following a well-defined protocol for analysing the data.

Participant checking is a key measure in mitigating researcher bias. Transcriptions were sent to participants before analysis to verify whether transcriptions indicate what participants intended to say. After analysis, a copy of this article has been sent to participants for comment before journal submission, to let participants point out any cases where their statements may have been misinterpreted. Their suggestions have been accommodated in the article.

Triangulation of analyses can also contribute to mitigating researcher bias. The first author was primarily involved in data collection and transcription of data, and data have been further analysed together with a co-author. Participant checking and triangulation of analyses correspond to the best practice guideline for implementing and reporting of qualitative research (Twining et al., 2017).

5.2.2. Transferability (external validity or generalizability)

To what extent can findings from this study be generalized to other settings? One notable limitation is that the study is focusing on the situation in only one country, Norway. All the system managers were from Norwegian universities, and the purchasing organization a Norwegian directorate. True, the two vendors have customers in several countries, and one of the companies (WF) is Danish, so the vendors take on required features for their products will have had a somewhat more international perspective, also exemplified by specific statements from participants that requirements would be different from country to country. Still, the context of the case is specifically the situation in Norway, and a study including universities from other countries might have come up with different findings. The two companies involved in the study were both vendors of dedicated e-exam software, hence catering to universities who use different products for high stakes e-exams than what they use for e-learning in general. As mentioned in the introduction, many universities around the world may be using the same system (e.g., Canvas, Blackboard, Moodle) both for e-learning and for high stakes tests, which may lead to differences in expectations towards the products. So, further work is needed to take a more international approach and to get findings covering a broader spectrum of educational software products. Nevertheless, challenges such as security and interoperability are key to e-exams in many countries – as indicated by related work – so findings are believed to be of interest also outside the specific Norwegian context.

5.2.3. Dependability (reliability) and Confirmability (objectivity)

Often researchers find it challenging to synthesize results from the interpretation of qualitative data analysis (Twining et al., 2017). We followed guidelines by (Anney, 2014; Korstjens & Moser, 2018) to ensure dependability and confirmability of our findings.

Authors applied triangulation of analyses and participant checking to conduct audit trail of collected data, analysis documents. Authors were also in prolonged engagement with participants before submission to the journal to ensure whether the analysis was consistent with collected data rather than their own imaginations. Also, the analysis process has been constantly verified with the grounded theory approach to ensure whether the analysis is consistent with the approach. So, the research approach and instrument (i.e., interview questionnaire) can be used in a similar group of subjects and research settings.

6. Conclusion

This paper has reported on an interview study with 12 persons having central roles in the development, procurement, and operation of e-exam systems in Norwegian higher education, some being vendors, some system managers at various universities, and one being a process manager at the national organization Unit, coordinating the acquisitions and IT infrastructure development for all Norwegian public universities. Participants generally seemed satisfied with the requirements process, feeling that the coordinating role of Unit had improved the process compared to the previous situation where each university was running separate acquisitions and integration projects – though it was observed that parts of the requirements process were still somewhat ad hoc. There was much agreement between participants about both functional and non-functional features of e-exam systems, system managers much focussed on security and interoperability. As for security, while there might be some vulnerabilities for cheating, especially with the BYOD type of e-exam, stakeholders generally felt that systems were satisfactory and believed that more cheating was taking place outside the digital exam infrastructure – e.g., old fashioned cheat notes, concealed mobile phones. As for interoperability, stakeholders generally agreed that long-term ambition should be a move towards a digital ecosystem where open standards and APIs would allow for smooth integrations between various tools. However, at the current stage, many integrations were challenging – both considering the usage of third-party tools during the exam and exchanging information between e-exam systems and other tools, such as Learning Management Systems. Since our case study was restricted to Norway, some findings may be specific to the national setting. However, both e-exam system vendors have customers in several countries, in Europe and beyond. Hence, the challenges surrounding requirements specification, security and interoperability of e-exam systems are likely of much broader interest. Interesting avenues for further work could be to perform similar studies in other countries, or across several countries, and also to compare the views of the stakeholder groups interviewed here with those of students and teachers. It will also be exciting to see how the e-exam systems of these two vendors – and other competitors – develop over the next couple of years. While the move towards open digital ecosystems is a fine ideal that many agree about, there are also obvious business interests that would go in favour of maintaining a situation dominated by proprietary software.

What advice should then be given to universities pursuing better e-exam systems in the future? Is it a good idea to continue relying on commercial vendors of learning management systems and e-assessment systems – as is currently done in Norway – or would it be better for universities to develop their own systems, as has been done, for instance, in the TeSLA project? If continuing with commercial vendors, our study indicates that it is a good idea that several universities collaborate in the requirements and acquisition process, to have more influence and better bargaining power versus the vendors. Likewise, with a development approach, a collaboration between many universities seems more likely to give successful results than each university making its own system. In any case, it is important to have a long-term view, focusing on the interoperability between the e-exam systems and other IT systems in the higher education sector, and to keep in mind that the ultimate goal is not the technology itself, but pedagogical improvement of assessment practices.

Appendix

Appendix A: Participant Information Letter

Invitation

I would like to talk with you about your experiences with e-exams.

Purpose of this study

I hope to compare experiences of e-exam systems between several individuals, institutions and countries to publish in an academic journal.

What does participation involve for you?

The participant will be engaged in the interview. The interview will last approximately 45-60 mins.

If you are agreeable, I would like to record our conversation. I will ask you to review the draft article to ensure your comments have been accurately reported.

Participation is voluntary

Participation in this study is voluntary. All information about you will then be made anonymous.

Your personal privacy – how we will store and use your personal data

We will process your personal data confidentially following data protection legislation (the General Data Protection Regulation and Personal Data Act). The data will not be handed to other researchers or professors at NTNU or outside of the institution.

What will happen to your personal data at the end of this study?

The project is scheduled to end 2020. The data will be stored in a university computer and will be kept until the end of the project after that data will be deleted.

Your rights

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What if I have questions about this study?

- Please contact at aparna.vegendla@ntnu.no or guttorm.sindre@ntnu.no

This study has been approved by the Norwegian Centre for Research Data AS.

This information letter is for you to keep. Your consent will be indicated by your agreement to speak with me at a mutually agreed appointment.

Appendix B: Demographic information of interview participants

Person & date	Position	Institution and site code	E-exam technology
IA1, 15Aug19	Development manager	Inspera AS, Norway, **IA**	Inspera Assessment, SEB, Blackboard, FS
IA2, 11Sep19	Product manager	Inspera AS, Norway, **IA**	Inspera Assessment, SEB
WF, 17Sep19	Head of product development	UNIWISE, Aarhus, Denmark, **WF**	WISEflow, FLOWlock, Canvas, FS
Unit, 29Aug19	Senior advisor	The Norwegian Directorate for ICT and joint services in higher education and research, Norway, **Unit**	Inspera Assessment, WISEflow
NTNU1, 13Aug19	Project manager, IT Digital Assessment	Norwegian University of Science and Technology, Norway, **NTNU**	Inspera Assessment, SEB, Blackboard, FS

NTNU2, 7July19	Team leader, IT Digital Assessment	Norwegian University of Science and Technology, Norway, **NTNU**	Inspera Assessment, SEB, Blackboard, FS
NTNU3, 14June19	Engineer, Digital Security Section	Norwegian University of Science and Technology, Norway, **NTNU**	Inspera Assessment, SEB
NTNU4, 14June19	Engineer, IT department	Norwegian University of Science and Technology, Norway, **NTNU**	Inspera Assessment, SEB, Blackboard, FS
NTNU5, 8Apr19	Senior consultant, IT Digital Assessment	Norwegian University of Science and Technology, Norway, **NTNU**	Inspera Assessment, SEB
HVL, 3July20	Senior advisor, IT Digital Assessment	Western Norway University of Applied Sciences, Norway, **HVL**	WISEflow, FLOWlock, Inspera Assessment, SEB, Canvas, FS
KU, 2July20	Senior advisor, IT Digital Assessment	Kristiania University College Norway, **KU**	WISEflow, FLOWlock, Inspera Assessment, SEB, Canvas, FS
UiT, 21Aug20	Advisor	University of Tromsø - The Arctic University of Norway, **UiT**	UiT, 21Aug20

Appendix C: Research questions and corresponding interview questions

Research Question	Interview Questions
What process is followed by vendors, process managers, and system managers to identify features and agree upon requirements?	▪ How do you conduct the RE process?
What do vendors, process managers and system managers see as key features for functionality and security in e-exam systems?	▪ What are the key functional requirements for digital exam system? ▪ What are the key security requirements for digital exam system?
What are the goals and challenges concerning integrations between the e-exam system and other exam supporting systems?	▪ How do you conduct integrations? ▪ What are the challenges of integrations between your organisation e-exam system with exam supporting systems, e.g., LMS, student information system, and lockdown browser?
To what extent do the stakeholders envision a move towards a more open digital ecosystem for e-exams, and would this be assumed to impact security?	Questions specific to vendors: ▪ Would you like to provide monolithic or a system as a part an ecosystem? ▪ How your e-exam system supports customisation? ▪ How do you support different third-party plug-ins in your e-exam system? ▪ What are the challenges of integrations between your e-exam system and third-party tools? Questions specific to process managers and system managers: ▪ Which system would you like to experience between monolithic or an ecosystem? ▪ Will the requirements to have an open API and allow for third-party and university-made plug-ins be important? ▪ Will the requirements to enable the sharing of questions across institutions to be important? Will there be any security challenges if we do so? ▪ Are there any security threats of paper exams that have been removed or reduced by digital exams? Common questions to vendors, process managers and system managers: ▪ What are the most important cheating/security risks to e-exam technologies during on-campus digital exams, especially using BYOD devices?

References

Achimugu, P., Selamat, A., Ibrahim, R., & Mahrin, M. N. r. (2014). A systematic literature review of software requirements prioritization research. *Information and Software Technology, 56*(6), 568-585.

Adebayo, O., & Abdulhamid, S. M. (2014). E-exams system for Nigerian universities with emphasis on security and result integrity. *International Journal of the Computer, the Internet and Management (IJCIM), 18*(2), 1-12.

Alier, M. F., Guerrero, M. J. C., Gonzalez, M. A. C., Penalvo, F. J. G., & Severance, C. (2010). Interoperability for LMS: the missing piece to become the common place for e-learning innovation. *International Journal of Knowledge and Learning, 6*(2-3), 130-141.

Anney, V. N. (2014). Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. *5*(2), 272-281.

Baró-Solé, X., Guerrero-Roldan, A. E., Prieto-Blázquez, J., Rozeva, A., Marinov, O., Kiennert, C., Rocher, P. O., & Garcia-Alfaro, J. (2018). Integration of an adaptive trust-based e-assessment system into virtual learning environments—The TeSLA project experience. *Internet Technology Letters, 1*(4), e56.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report, 13*(4), 544-559.

Bilen, E., & Matros, A. (2021). Online cheating amid COVID-19. *Journal of Economic Behavior & Organization, 182*, 196-211.

Boeije, H. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity, 36*(4), 391-409.

Boezerooy, P., Cordewener, B., & Liebrand, W. (2007). Collaboration on ICT in Dutch Higher Education: The SURF Approach. *Educause Review, 42*(3).

Boussakuk, M., Ghazi, M. E., Bouchboua, A., & Ouremchi, R. (2019). Online assessment system based on IMS-QTI specification. In *2019 7th Mediterranean Congress of Telecommunications (CMT)* (pp. 1-4).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Brink, R., & Lautenbach, G. (2011). Electronic assessment in higher education. *Educational Studies, 37*(5), 503-512.

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education, 159*, 104024.

Buzzetto-More, N. A., & Alade, A. J. (2006). Best practices in e-assessment. *Journal of Information Technology Education: Research, 5*(1), 251-269.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*: sage.

Chirumamilla, A., & Sindre, G. (2019). E-Assessment in Programming Courses: Towards a Digital Ecosystem Supporting Diverse Needs? In I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, J. Krogstie & M. Mäntymäki (Eds.), *Digital Transformation for a Sustainable Society in the 21st Century* (pp. 585-596). Cham: Springer International Publishing.

Chituc, C.-M., Herrmann, M., Schiffner, D., & Rittberger, M. (2019). Towards the Design and Deployment of an Item Bank: An Analysis of the Requirements Elicited. In (pp. 155-162). Cham: Springer International Publishing.

Chituc, C.-M., & Rittberger, M. (2019). Understanding the Importance of Interoperability Standards in the Classroom of the Future. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society* (Vol. 1, pp. 6801-6806).

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*: Routledge.

Cluskey Jr, G., Ehlen, C. R., & Raiborn, M. H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics, 4*(1).

D'Souza, K. A., & Siegfeldt, D. V. (2017). A conceptual framework for detecting cheating in online and take-home exams. *Decision Sciences Journal of Innovative Education, 15*(4), 370-391.

Dagger, D., Connor, A. O., Lawless, S., Walsh, E., & Wade, V. P. (2007). Service-Oriented E-Learning Platforms: From Monolithic Systems to Flexible Services. *IEEE Internet Computing, 11*(3), 28-35.

Dawson, P. (2016). Five ways to hack and cheat with bring-your-own-device electronic examinations. *British Journal of Educational Technology, 47*(4), 592-600.

Dick, J., Hull, E., & Jackson, K. (2017). *Requirements engineering*: Springer.

Dick, M., Sheard, J., Bareiss, C., Carter, J., Joyce, D., Harding, T., & Laxer, C. (2002). Addressing student cheating: definitions and solutions. *ACM SIGCSE Bulletin, 35*(2), 172-184.

Dyer Jr, W. G., & Wilkins, A. L. (1991). Better stories, not better constructs, to generate better theory: A rejoinder to Eisenhardt. *Academy of Management Review, 16*(3), 613-619.

Ferdiana, R., & Hoseanto, O. (2018). The implementation of computer based test on BYOD and cloud computing environment. *International Journal of Advanced Computer Science and Applications, 9*(8), 121-124.

Firesmith, D. (2003). Specifying good requirements. *Journal of Object Technology, 2*(4), 77-87.

Fitzharris, R., & Kent, S. (2020). Adoption of Bring-Your-Own-Device Examinations and Data Analytics. In D. Ifenthaler & D. Gibson (Eds.), *Adoption of Data Analytics in Higher Education Learning and Teaching* (pp. 327-348). Cham: Springer International Publishing.

Fluck, A., Adebayo, O. S., & Abdulhamid, S. i. M. (2017). Secure e-examination systems compared: Case studies from two countries. *Journal of Information Technology Education: Innovations in Practice, 16*(1), 107-125.

Fluck, A., & Hillier, M. (2017). eExams: Strength in Diversity. In A. Tatnall & M. Webb (Eds.), *IFIP Advances in Information and Communication Technology* (pp. 409-417). Cham: Springer International Publishing.

Fluck, A., Pálsson, H., Coleman, M., Hillier, M., Schneider, D., Frankl, G., & Uolia, K. (2017). eExam symposium: design decisions and implementation experience. In *11th IFIP TC 3 World Conference on Computers in Education*.

Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology, 25*(4).

Fluck, A. E. (2019). An international review of eExam technologies and impact. *Computers & Education, 132*, 1-15.

Foss-Pedersen, R. J., & Begnum, M. E. N. N. (2017). Universell utforming og digital eksamen i UH-sektoren: 5 anbefalte tiltakspunkter. In *Norsk konferanse for organisasjoners bruk at IT* (Vol. 25).

Frankl, G., Schartner, P., & Zebedin, G. (2012). Secure online exams using students' devices. In *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-7).

Frosini, G., Lazzerini, B., & Marcelloni, F. (1998). Performing automatic exams. *Computers & Education, 31*(3), 281-300.

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The discovery of grounded theory; strategies for qualitative research. *Nursing research, 17*(4), 364.

Gustafsson, J. (2017). Single case studies vs. multiple case studies: A comparative study. In.

Hillier, M., & Fluck, A. (2013). Arguing again for e-exams in high stakes examinations. *Electric dreams. proceedings ascilite*, 385-396.

Huszti, A., & Petho, A. (2010). A secure electronic exam system. *Publicationes Mathematicae Debrecen, 77*(3-4), 299-312.

Immonen, A., Ovaska, E., Kalaoja, J., & Pakkala, D. (2016). A service requirements engineering method for a digital service ecosystem. *Service Oriented Computing and Applications, 10*(2), 151-172.

Isaias, P., Miranda, P., & Pífano, S. (2019). Framework for the analysis and comparison of e-assessment systems. In *ASCILITE 2017-Conference Proceedings-34th International Conference of Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 276-283): Australasian Society for Computers in Learning in Tertiary Education (ASCILITE).

Jakimoski, K. (2016). Challenges of interoperability and integration in education information systems. *International Journal of Database and Theory and Application, 9*(2), 33-46.

JISC. (2007). Effective Practice with e-Assessment: An overview of technologies, policies and practice in further and higher education. In: Joint Information Systems Committee.

Kaiiali, M., Ozkaya, A., Altun, H., Haddad, H., & Alier, M. (2016). Designing a Secure Exam Management System (SEMS) for M-Learning Environments. *IEEE Transactions on Learning Technologies, 9*(3), 258-271.

Kallio, H., Pietila, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing, 72*(12), 2954-2965.

Kerssens, N., & Dijck, J. v. (2021). The platformization of primary education in The Netherlands. *Learning, Media and Technology*, 1-14.

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice, 24*(1), 120-124.

Kuikka, M., Kitola, M., & Laakso, M.-J. (2014). Challenges when introducing electronic exam. *2014, 22*.

Ladonlahti, T., Laamanen, M., & Uotinen, S. (2020). Ensuring diverse user experiences and accessibility while developing the TeSLA e-assessment system. In *Engineering data-driven adaptive trust-based e-assessment systems* (pp. 213-238): Springer.

Laine, K., Sipilä, E., Anderson, M., & Sydänheimo, L. (2016). Electronic exam in electronics studies. In *SEFI Annual Conference 2016*.

Llamas-Nistal, M., Fernández-Iglesias, M. J., González-Tato, J., & Mikic-Fonte, F. A. (2013). Blended e-assessment: Migrating classical exams to the digital world. *Computers & Education, 62*, 72-87.

Llorens, F., Molina, R., Compañ, P., & Satorre, R. (2014). Technological Ecosystem for Open Education. In *IDT/IIMSS/STET* (pp. 706-715).

Luo, M.-Y., & Lin, S.-W. (2013). From monolithic systems to a federated e-learning cloud system. In *2013 IEEE International Conference on Cloud Engineering (IC2E)* (pp. 156-165): IEEE.

Mellar, H., Peytcheva-Forsyth, R., Kocdar, S., Karadeniz, A., & Yovkova, B. (2018). Addressing cheating in e-assessment using student authentication and authorship checking systems: teachers' perspectives. *International Journal for Educational Integrity, 14*(1), 1-21.

Melve, I., & Smilden, B. (2015). ICT Architecture for Digital Assessment. In *UNINETT* (pp. 102).

Menge, F. (2007). Enterprise service bus. In *Free and open source software conference* (Vol. 2, pp. 1-6).

Naous, D., Giessmann, A., & Legner, C. (2020). Incorporating the voice of the customer into mass-market software product management. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing* (pp. 1397-1404).

Okada, A., Noguera, I., Alexieva, L., Rozeva, A., Kocdar, S., Brouns, F., Ladonlahti, T., Whitelock, D., & Guerrero-Roldán, A. E. (2019). Pedagogical approaches for e-assessment with authentication and authorship verification in Higher Education. *British Journal of Educational Technology, 50*(6), 3264-3282.

Okada, A., Whitelock, D., Holmes, W., & Edwards, C. (2019). e-Authentication for online assessment: A mixed-method study. *British Journal of Educational Technology, 50*(2), 861-875.

Piotrowski, M. (2011). QTI: A Failed E-Learning Standard? In L. Fotis, G. Steve & P. Elaine (Eds.), *Handbook of Research on E-Learning Standards and Interoperability: Frameworks and Issues* (pp. 59-82). Hershey, PA, USA: IGI Global.

Queirós, R., Leal, J. P., & Paiva, J. C. (2016). Integrating Rich Learning Applications in LMS. In (pp. 381-386). Singapore: Springer Singapore.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*: Wiley-Blackwell.

Sandkuhl, K., & Lehmann, H. (2017). Digital transformation in higher education–The role of enterprise architectures and portals. *Digital Enterprise Computing (DEC 2017)*.

Sclater, N. (2007). The Demise of eAssessment Interoperability? *WWWrong, 317*.

Sclater, N., Low, B., & Barr, N. (2002). Interoperability with CAA: does it work in practice?

Severance, C., Hanss, T., & Hardin, J. (2010). Ims learning tools interoperability: Enabling a mash-up approach to teaching and learning tools. *Technology, Instruction, Cognition and Learning, 7*(3-4), 245-262.

Striewe, M. (2019). Components and Design Alternatives in E-Assessment Systems. In (pp. 220-228). Cham: Springer International Publishing.

Tomberg, V., Savitski, P., Djundik, P., & Berzinsh, V. (2017). Design and Development of IMS QTI Compliant Lightweight Assessment Delivery System. In (pp. 159-170). Cham: Springer International Publishing.

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education, 106*, A1-A9.

Uden, L., Wangsa, I. T., & Damiani, E. (2007). The future of E-learning: E-learning ecosystem. In *Digital EcoSystems and Technologies Conference, 2007. DEST '07. Inaugural IEEE-IES* (pp. 113-117).

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory'back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal, 20*(4), 357-381.

Veiga, W., Campos, F., Braga, R., & David, J. M. (2016). A Software Ecosystem approach to e-Learning domain. In *Proceedings of the XII Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era - Volume 1* (pp. 574–581). Florianopolis, Santa Catarina, Brazil: Brazilian Computer Society.

Weishaupl, E., Yasasin, E., & Schryen, G. (2018). Information security investments: An exploratory multiple case study on decision-making, evaluation and learning. *Computers & Security, 77*, 807-823.

Wibowo, S., Grandhi, S., Chugh, R., & Sawir, E. (2016). A pilot study of an electronic exam system at an Australian University. *Journal of Educational Technology Systems, 45*(1), 5-33.

Wills, G. B., Davis, H. C., Gilbert, L., Hare, J., Howard, Y., Jeyes, S., Millard, D., & Sherratt, R. (2009). Delivery of QTIv2 question types. *Assessment & Evaluation in Higher Education, 34*(3), 353-366.

Yin, R. (2002). *Case Study Research: Design and Methods* (Vol. 5): SAGE Publications, Inc.

Yin, R. K. (2017). *Case study research and applications: Design and methods* (6th ed.): Sage publications.

ISBN 978-82-326-5579-3 (printed ver.)
ISBN 978-82-326-6341-5 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology