Utrecht University

Master's Thesis

Submitted in partial fulfilment of the Master's Programme

Business Informatics

Increasing MOOC Completion Rates Through Adaptive Learning: A Case Study

Cas van Cooten

c.h.vancooten@students.uu.nl

supervised by

Dr. Matthieu Brinkhuis (Utrecht University)

Dr. Fabiano Dalpiaz (Utrecht University)

mr. drs. Tim van Essen (Deloitte Netherlands)

& Niels Pompe (Deloitte Netherlands)

Academic year 2016/2017

17th July 2017

Abstract

Massive Open Online Courses (MOOCs) are a highly popular, relatively new phenomenon in the world of online education. However, due to their massive scale they encounter many problems such as low completion rates and a lack of personalised learning at scale. Several reasons for low completion rates are discussed in literature, among which are course difficulty and course workload. In addition, some authors have suggested adaptive learning as a solution for personalised learning. However, little research has been performed towards the combination of MOOCs and adaptive learning.

Therefore, this research project aimed to investigate the effect of adaptive learning on learner satisfaction, learner engagement, and ultimately completion rates in the context of MOOCs. It was hypothesised that adaptive learning would increase all three variables. To investigate this, an online adaptive learning system was designed using an approach based on the principles of design science. The designed system was then applied in practice in a case study. The designed system represented a challenge-based MOOC on cyber security and contained three experimental conditions: a randomised condition to calibrate the adaptive system, a linear condition, and an adaptive condition. The system collected quantitative usage data as well as qualitative data by means of two surveys.

The system was deployed in practice at five educational institutions. A total of 156 users registered on the system, of which 131 users participated actively. Adaptive learning was found to significantly reduce learner dropout and completion time when compared to the linear condition. However, adaptive learning also significantly reduced learner satisfaction. No significant effect of adaptive learning on engagement or completion rates was found. Additionally, no effect of satisfaction or engagement on completion rates was found, though satisfaction and completion were found to be highly correlated. It is concluded that implementing adaptive learning in MOOCs is a viable option for MOOC developers, but that it depends on the mission and context of a specific MOOC whether this is desirable. Concrete recommendations to support decision-making are provided.

Contents

Preface

Before you lies my Master's thesis, a product of what I can proudly call the most elaborate project of my life thus far. For eight months I have been able to dedicate my time to combine two fields that greatly interest me; cyber security and education. I am very proud of what has been achieved during this project, and I sincerely hope I managed to enthuse some young souls about cyber security along the way.

Of course, I wouldn't have been able to complete a project of this scale all on my own. Therefore, I would like to thank everyone who assisted me during this project in any way. Specifically I would like to thank Matthieu Brinkhuis for his continuous intensive and involved guidance. I would also like to thank Fabiano Dalpiaz for his feedback and reviews throughout this project. Finally, I would like to thank Tim van Essen and Niels Pompe for their valuable insights and guidance regarding the practical side of this project.

In addition, I would like to express my gratitude to everyone that facilitated the execution of my case study at various educational institutions: Anneloes Leijenhorst, Paul Kool & Wim van der Weiden (Alberdingk Thijm College), Hakan Akkas (Metis Montessori Lyceum), Hanneke Kool & Vincent Raphael (ROC Mondriaan), Reza Esmaili (Amsterdam University of Applied Sciences), and of course Fabiano Dalpiaz (Utrecht University).

I hope you enjoy reading my Master's thesis and it is of some value to you personally or professionally.

Cas van Cooten

1 Introduction

Massive Open Online Courses (MOOCs) are a relatively new phenomenon in the world of online learning. Popularised in 2012, a MOOC is defined as "an online course with the option of free and open registration, a publicly-shared curriculum, and open-ended outcomes", which "build on the engagement of learners who self-organise their participation according to learning goals, prior knowledge and skills, and common interests" [1, p. 10]. Two types of MOOC are often distinguished: Connectivist 'cMOOCs', which have a focus on connectivist theory and take a social approach to learning, and the later emerged transmissive 'xMOOCs', which take a more traditional approach to learning [2]–[4]. These two MOOC types are very different in pedagogy; cMOOCs focus on learning by creation and collaborating with peers, while xMOOCs focus on learning by studying material and practising. In this thesis, the distinction between cMOOCs and xMOOCs will only be made where relevant. The term MOOC will be used as a more general term covering both branches of MOOCs.

MOOCs are "massive" and "open", and are thus freely available to large audiences at once. Therefore, they provide many advantages over traditional classroom education. Due to the "massive" aspect, the reach of MOOCs extends far beyond traditional courses or e-learnings. Furthermore, the large user base and integrated social networking tools also allow MOOCs to facilitate knowledge sharing and creation in a unique way [1]. The "open" aspect of MOOCs delivers value by making universities embrace openness within their core business models [5] and opening up large bodies of knowledge for anyone to access. Due to the "online" aspect of courses, barriers to entry are low. All that is required to participate in a MOOC is an Internet connection, an interest in the MOOC's subject and in some cases a small enrolment or certification fee. These low barriers of entry make it easy for students to enrol and learn about a new subject.

1.1 Problem Statement

Despite the various advantages and possibilities, MOOCs have various downsides and some experts in the field of higher education remain sceptical about their potential for various reasons, including low instructional quality [6], [7]. One major limitation to the adoption of MOOCs is the low completion rates. The completion rate for most courses is below 10%, with a median average of 6.5% [8].

On the other hand, Onah *et al.* [9] nuance these numbers by stating that they are derived from baseline registration numbers which include learners who never intend to engage with the course, or engage in their own way without completing assessments. In addition, Kizilcec *et al.* [10] define four prototypical learner trajectories: 'Completing', 'Auditing', 'Disengaging' and 'Sampling'. Of these, 'Auditing' (following the course, but not completing assessments) and 'Sampling' (only interacting with some material) would be considered "non-completing" in most metrics, even though the learners are still engaging with the study material in their own preferred way. Still, the 'Disengaging' group was consistently larger than the 'Completing' group, the latter only accounting for 5% to 27% of participants.

Reasons for disengagement are various. Factors like a lack of incentive, course difficulty, course workload, work conflicts, rigid or inflexible course structures, and a lack of digital or learning skills have all been listed as reasons for disengagement [1], [9]–[12]. Additionally, since the technology of MOOCs is relatively new, learners may simply want to try participating in a MOOC without any intention of actually engaging with or completing the course, further contributing to dropout rates [13]. A list of important reasons for MOOC dropout identified from literature can be found in Table 2.1 in the next chapter.

The low completion rates that MOOCs are currently facing pose a threat to the long-term viability of MOOCs. Currently, MOOC providers are adopting a "freemium to premium" strategy: Initial services are free, but once a user base is established additional products or services are offered at a cost [14]. To realise these "premium" MOOCs, providers are experimenting with various business models in order to develop one that is suitable for MOOCs [15]. These business models include, but are not limited to, paid certification, recruitment, applicant screening, tuition fees, and sponsorships [5]. Business models like these greatly depend on a large user base in order to realise value. However, since the completion rate of MOOCs is so low, the profitability of MOOCs is just a fraction of what it could be.

Due to the massiveness of MOOCs, the population of learners is very heterogeneous. Failure to adapt to this heterogeneity is also a reason for learner dropout [13]. The problem is that most MOOCs are generally "flat" and presentational in nature [16]. Onah *et al.* [9] point out that adaptively structuring MOOCs would give learners more freedom to reach their learning objectives. In his blog, Savi [17] also outlines the need for adaptive learning in massive education. According to him, adaptive technology can be applied to tailor massive education to an individual learner's needs. In an open letter in *Science*, Savi *et al.* [18] draw the analogy with a GPS navigation device: rather than steering all learners in the same direction, each learner should get their own, customised route. In the context of MOOCs, implementing adaptive learning could help improve learner performance at scale and subsequently increase perceived education quality and completion rates.

In sum, MOOCs are facing many issues which keep them from reaching their full potential. There is a high dropout rate, which threatens the future viability of MOOCs. Furthermore, there is a need for learner-tailored material on a massive scale in order to serve the various needs of a heterogeneous population of learners. In this thesis, an attempt will be made to counter these two main issues by implementing adaptive learning in MOOCs in order to raise completion rates.

1.2 Reading Guide

This thesis encompasses the entirety of a research project on massive open online education and adaptive learning. It is acknowledged that not all parts of this project and thesis are interesting to everyone. Therefore, a brief reading guide is provided below to guide you in navigating this thesis.

If you are interested in the product of this research, an adaptive online education system on cyber security, it is suggested to review Chapter 4 and the accompanying Design Specification document[1]. If you are interested in the scientific process and background of this research, chapters 2 and 3 are recommended. The practical process and analysis of results are discussed in chapters 5 and 6, respectively. Finally, if you are only interested in the findings and practical implications of this thesis, please refer to Chapter 7.

[1] If, for any reason, the design specification was not provided with your copy of this thesis, it can be downloaded separately here: `https://bit.ly/MScDesignDoc`

2 Literature Review

In this section, existing scientific studies from various fields of expertise are reviewed and compared in order to provide a complete and actual overview of existing, relevant scientific literature. The main focus points are MOOCs and Adaptive Learning, and how they are applied in practice.

2.1 MOOCs

As discussed in 1, MOOCs are a relatively new phenomenon in the field of learning. In this thesis, a MOOC is defined as "an online course with the option of free and open registration, a publicly-shared curriculum, and open-ended outcomes", which "build on the engagement of learners who self-organise their participation according to learning goals, prior knowledge and skills, and common interests" [1, p. 10].

In literature, MOOCs are most extensively discussed in the fields of educational and social sciences, but studies from all fields of research will be included in this section. Various types of MOOCs are defined in literature, these are discussed in more detail below. Subsequently, the potential of MOOCs will be discussed, followed by their application in practice and the aforementioned problems of completion and engagement.

2.1.1 MOOC Evolution and Types

The concept MOOC was introduced in 2008, when Stephen Downes and George Siemens published their online course "Connectivism and Connected Knowledge" using publicly available tools [19]. This course was later considered the first cMOOC. In 2012, the three major MOOC platforms edX, Coursera, and Udacity launched as university spin-offs [14]. Since these platforms took a different, more traditional approach to education, they were considered a different branch of MOOC, which was later named xMOOC. Because of the the sudden support from many prestigious universities, a (media) hype started growing around the subject with some even calling MOOC the "educational buzzword of 2012" [5]. In the years after, MOOCs have opened up several avenues of research, but initial results have failed to change education and learning [20].

As mentioned before, there are two branches of MOOC which each have their own distinctive features and underlying pedagogical principles. cMOOCs are based on a social, connectivist approach to learning, which facilitates knowledge sharing and creation through a high level of learner interaction [21]. A cMOOC "may be compared to seminars in which participants evaluate and structure new contents, create texts, and write comments that are then made available to other participants"[22, p. 2]. On the other hand, xMOOCs fall into the "cognitive-behaviourist" pedagogical category [3]. They take a more traditional approach to learning, where learners view video lectures and test their knowledge afterwards. In contrast to cMOOCs, information in xMOOCs is transmitted largely individually. However, xMOOCs may also contain optional social elements like peer grading or a discussion board about the MOOC's subject.

2.1.2 The Potential of MOOCs

According to Yuan *et al.* [14], MOOCs have all the characteristics of a disruptive innovation in the higher education market. However, large educational institutions see MOOCs as an incremental rather than a disruptive innovation, and use them to improve their existing educational practices. For institutions that were already involved in e-learning, for example, "MOOCs represent more of an incremental step along a pre-existing trajectory than a major innovation" [15, p. 66]. Still, Sharples *et al.* [23] believe that MOOCs have the potential to provoke major shifts in educational practice.

For example, MOOCs can bring education from top institutions to a diverse, worldwide learning audience for free. This includes third-world countries, where MOOCs have the potential to greatly increase the level of education. Furthermore, Wulf *et al.* [22] outline additional benefits that a MOOC could offer, like learner co-creation, low marginal (scaling) costs, positive network effects and the potential for individualisation of teaching services.

Apart from the educational potential, the magnitude and diversity of data that MOOCs generate create an enormous analytical potential [24]. Kay *et al.* [25] mention several other advantages of MOOC data: Not only do MOOCs provide much larger data sets, the data sets are also more heterogeneous due to the larger body of learners. These data sets provide value for educational data mining and learning analytics. Furthermore, historic data about successful and unsuccessful learners can be used to guide new learners or improve an entire online course.

2.1.3 MOOCs in Practice

Due to the high potential of MOOCs, many prestigious higher education institutions like Harvard and MIT have created and published MOOCs in order to provide open access to education, increase residential education and create a platform for educational research using data from their MOOCs [26]. Due to the immense popularity around the concept of MOOCs generated by these top institutions offering them, many other educational institutions followed.

An example of this is Utrecht University, which published its first MOOC on the Coursera platform in 2016 [27]. Three more MOOCs were planned to be published in 2016, in line with the university's strategic research themes. In doing so, Utrecht University aims to contribute to open education, increase its educational and scientific reputation and allow students to experience the content of various courses and programmes that the university offers on-campus and in their accredited, revenue-generating “Small Private Online Courses” (SPOCs).

From the learner’s perspective, MOOCs are also gaining popularity. For example, the aforementioned 68 courses published by Harvard and MIT accounted for a cumulative course enrolment growth of 2,200 learners daily, a total of 1.71 million learners over a two-year time period [26]. Daniel *et al.* [15] claimed a total MOOC enrolment of over five million learners worldwide in 2015, most of which have a prior higher education degree.

Hew *et al.* [11] point out the four main reasons why learners sign up for MOOCs: A desire to learn, curiosity about MOOCs, personal challenge and a desire to collect completion certificates. Due to the freely available high-quality educational content that MOOCs offer, they fit very well with the concept of “lifelong learning”, which seems to be confirmed by the highly educated learners and aforementioned factors for participation.

Of course, introducing courses at such a large scale introduces problems from an educational perspective. In describing his experiences with an early MOOC, Martin [28] states that he noticed weaker students struggle with the difficulty of the course material, while a few strong students were bored due to a lack of challenge. Additionally, Margaryan *et al.* [7] analysed a random sample of 76 MOOCs and found that they generally score poorly on most instructional design principles, suggesting a lower educational quality than traditional education. These articles underscore one of the key problems with the core concept of MOOCs: it is hard, if not impossible, to provide the support or teaching quality that a tutor can provide in their classroom on a massive scale. To combat this problem, MOOCs may employ several didactic mechanisms in order to better facilitate the large enrolment numbers. Among these mechanisms are peer support, peer grading, gamification and learning analytics [22].

2.1.4 Completion Rates and Engagement

As discussed in section 1.1, the low completion rate of MOOCs is a serious problem which may even pose a threat to its long-term business model. This high dropout rate for MOOCs has been described as a “funnel of participation”, which seems to be typical for many MOOCs [4]. A problem with literature on MOOC completion rates is that most studies focus on a small number of (early) MOOCs for their analysis [8]. Additionally, most studies focus on learner motivations for dropout rather than analyse the impact of course characteristics on enrolment and dropout.

In an attempt to counter this problem and aggregate distributed data on MOOC completion rates, Jordan [29] created a webpage where completion data on 217 MOOCs is collected and analysed. This analysis shows some interesting results. For example, peer graded courses show lower completion rates than automatically graded courses, and courses with lower enrolment rates show higher overall percentages of learners completing the course.

MOOC completion rates are a binary representation of learner performance: A learner either fulfils the requirements set by the course facilitator in order complete the course, or they do not and are subsequently “non-completing” or “drop out”. However, these statistics can be nuanced by looking at learner *engagement*. Engagement measures the amount of student activity with regards to an educational subject. It is not a new term in the field of education: in higher education literature, it has been shown that a high student engagement is a significant positive factor in a student's educational performance on various levels [30], [31]. According to Reich [20], MOOC research must look at learner engagement in order to advance the science of learning at large.

Within MOOCs, engagement provides a measure to identify and explain various user behaviour patterns. A learner may for example just watch video material, or a learner may only complete tests without interacting with course material or peers. In these cases, the learner is engaging with course material even though they would be considered "non-completing" otherwise.

The aforementioned learner trajectories proposed by Kizilcec *et al.* [10] discern typical patterns of engagement within MOOCs based on clustering. These classifications of behaviour shed more light on a learner's motivations and behaviour than the binary representation of dropout or completion rates. Because of this, using engagement as a measure allows for more nuanced research into learner engagement patterns and their predictors. For example, Alraimi *et al.* [32] found that perceived reputation and perceived course openness were very strong positive predictors for learner engagement in MOOCs.

Ramesh *et al.* [33] look at learner engagement in MOOCs by applying probabilistic soft logic. They model engagement as an interaction of behavioural, social and linguistic cues and subsequently train the model using data from a Coursera MOOC. Using the resulting algorithm, they can classify a learner's type and level of engagement based on their interactions with the course.

Engagement itself can also be used as a predictor for user behaviour or performance. For example, Sinha *et al.* [34] use a network-based approach to predict learner dropout based on engagement patterns, and He *et al.* [35] use predictive modelling to predict dropout by analysing their amount and type of engagement. Following these pieces of literature, it can be inferred that learner engagement is a predictor for both their performance and potential dropout within MOOCs.

2.2 Adaptive Learning

In section 1.1, adaptive learning is outlined as a potential solution for the dropout issue that MOOCs are facing. In this section, the notion of adaptive learning will be elaborated on and its need in MOOCs will be further highlighted. In this paper, the definition of adaptive learning proposed by Newman *et al.* [36, p. 3] will be used:

> *"Solutions that take a sophisticated, data-driven, and in some cases non-linear approach to instruction and remediation, adjusting to each learner's interactions and demonstrated performance level and subsequently anticipating what types of content and resources meet the learner's needs at a specific point in time."*

An important aspect of this definition is the notion of adaptive learning systems being data-driven. This differentiates adaptive learning systems from other feedback systems that for example return feedback to a learner when they answer a question wrong.

In this section, studies about the history and potential of adaptive learning will be discussed first. Subsequently, some practical studies involving adaptive learning will be highlighted, followed by studies that attempted to implement it in MOOCs. At the end of this section, a concrete list of reasons for dropout and their hypothesised adaptive measures is provided.

2.2.1 History

Adaptive learning is not a new concept, it has been around since the early days of the computer and has existed in many different forms: Adaptive learning has its roots in 'intelligent tutoring systems', which date back to the 1950s. In 1999, Brusilovsky [37] provided an overview of adaptive learning technologies in the context of web-based education. Even then, the goal of adaptive learning was to increase the quality of web-based education, and to allow for personalised learning at scale.

In 2013, educational advisory and market research firm Tyton Partners published their first report on adaptive learning [38], making a case for faster adoption of adaptive learning in education. They argue that the potential of adaptive learning technology was being overshadowed by the "MOOC mania" happening at that time. In recent years, the increasing ability to collect and process learner data through (online) platforms caused a steep increase in popularity for adaptive learning.

2.2.2 The Potential of Adaptive Learning

Today, the development and adoption of adaptive learning systems is mostly driven by an increasing need for scalable, personalised learning. Because of this, adaptive learning is recognised as one of the most promising technologies in the field of education. In 2016, market research firm Gartner identified adaptive learning as the number one strategic technology in education for the second year in a row [39]. Over the last few years, Newman *et al.* [36] have perceived a broader adoption of adaptive learning, a wider range of possible applications for it and an increased focus on benefits for the education provider, called "adaptive teaching".

Adaptive learning promises to offer many advantages in (online) education, especially when applied in education at scale. By tailoring educational content per learner based on their performance in real-time, each student is learning at their own pace and at the right level of challenge. Challenge has been shown to be an important factor to create fun in digital learning environments [40]. Fostering fun and learner satisfaction by adaptively providing the right level of challenge is therefore hypothesised to increase learner engagement with the material.

2.2.3 Practical Applications of Adaptive Learning

Following the developments described above, more and more educational institutions are adopting some form of adaptive learning technology. According to Gartner [39], institutions that have implemented adaptive learning technologies are reporting positive results, both in terms of learning results and student satisfaction. In academic literature, various case studies on the development and implementation of adaptive learning are described, some of which are discussed in this section.

For example, Klinkenberg *et al.* [41] implemented adaptive learning in an on-line practice system for children's maths abilities. By implementing a system that continuously adapts to the learner and subsequently offers maths questions that offer the right level of challenge, they increased engagement for both highly skilled and less skilled children.

Torrente *et al.* [42] created a gamified system that helps medical students familiarise with a laboratory exercise that is difficult to facilitate in real life. By creating a game that simulates the real-life environment and tracks and adapts to the user's performance, they increased student familiarity with the exercise and efficiency in their real-life lab time. In addition, they achieved a high level of system acceptance with the students.

Martín *et al.* [43] designed a system that is capable of automatically generating adaptive mobile learning interfaces. By offering an interface that adaptively recommends the optimal learning content, learners were more motivated to engage with the material and discuss it with peers, and better able to organise their learning time. The system, especially the dynamic adaptation to content and learning styles, was perceived as useful by its users.

Finally, Szafir *et al.* [44] created a system that adaptively recommends the best video content for learners to review, based on a learner's observed attention. They found that doing so optimises the time spent on reviewing content and therefore improves learning outcomes. Additionally, they found that a full review of all video content is a non-optimal strategy when learning. These observations underline the possibilities of adaptive learning and its potential when providing on-line education.

2.2.4 Adaptive Learning in MOOCs

The case studies described in the previous section clearly show the potential positive effects offered by adaptive learning systems for increasing engagement. It makes sense then to combine adaptive learning and MOOCs in order to increase MOOC engagement and quality or reduce learner dropout. In literature, some authors attempted to combine adaptive learning with MOOCs.

Fidalgo-Blanco *et al.* [13], for example, designed a methodological approach and technological framework, combining aspects of both X and C-type MOOCs and integrating them with knowledge management, learning analytics, and adaptive learning capabilities. By doing so, they reduced MOOC dropout from 90% to 70%. Their adaptive system increased learner interaction and engagement through dynamic grouping, early dropout detection, and dynamic content provision.

Kay *et al.* [25] call for another adaptive solution: automatically identifying learners at risk of failing or dropping out. He *et al.* [35] took a first step towards doing this with positive results, allowing for potential personalised interventions for learners at risk of dropping out. Adaptive learning could help further improve the mechanisms devised by them by providing more detailed insight in a learner's characteristics and why they are disengaging.

Sonwalkar [45] of Synaptic Global Learning developed cloud and pedagogy frameworks to support personalised learning in MOOC environments. His study shows that it is realistically possible to realise a tailored learning system on a massive scale.

Even though steps have been taken towards combining MOOCs and adaptive learning, a clear gap in literature remains. This gap is demonstrated by many authors calling for the implementation of adaptive learning to personalise learning within MOOCs. Some examples have already been given in section 1.1 [9], [13], [17], [18].

In addition, Daniel *et al.* [15] state that "There is a need to develop sophisticated adaptive learning mechanisms that will require the establishment of MOOC working partnerships between educators, instructional designers, and programmers" [p. 69]. Clark [16] of Cogbooks, a platform for creating adaptive courseware, states that the need for adaptive learning in MOOCs stems from two factors. The first is the "massive" element of MOOCs, which causes MOOCs to have a diverse audience with varying backgrounds. To account for this, adaptive learning systems can facilitate a more diverse pedagogy according to the needs of this heterogeneous learner population. The second factor is the "open" characteristic. Due to this, everyone is allowed to enrol in courses which causes a big diversity in skill levels. Adaptive learning can remedy this issue by tailoring content according to a learner's individual skill level.

Table 2.1 summarises the reasons for MOOC dropout mentioned before and the hypothesised (adaptive learning) measures that could be taken to reduce their effects.

Dropout Reason	**Source(s)**	**Hypothesised Measures**
Course difficulty	[9], [11]	Track user skill level and dynamically scale course difficulty and/or pace.
Course inflexibility	[1], [9], [12]	Release course content all at once [46] to make course self-paced.
Course workload	[10]	Adaptively pace course in accordance with user needs.
Lack of digital skills	[9], [12]	Guide user through MOOC functionality and procedures. Provide help when needed.
Lack of incentive	[11], [12]	Create direct incentive by providing certification, or create indirect incentive by challenging the user. The latter can be achieved by adaptively scaling the MOOC's difficulty level in accordance to user needs.
Lack of interactivity	[12]	Increase MOOC interactivity by providing adaptive content.
Lack of learning skills	[9], [12]	Guide user through assignment process and learning goals. Provide help when needed.
Lack of time	[12]	Shorten MOOC length by providing only content that delivers the most value.
No help available	[11]	Provide the user with dynamic assisting elements.
Other priorities (work/personal)	[10], [11]	Increase MOOC incentive (see above).

Table 2.1: Dropout reasons and their hypothesised (adaptive learning) measures

3 Research Method

As discussed in the previous chapter, adaptive learning has been proposed as a possible tool to increase learning quality and remedy the problem of dropout in massive learning. For MOOCs the integration of adaptive learning could be especially fruitful, since many opportunities for improving pedagogy, learning outcomes, and completion rates exist. Since completion rates seem to be the most pressing issue for the future of MOOCs, it will be the main focus of this research.

It was also discussed that learner engagement is an important measure in (online) education, which allows for the explanation and prediction of a user's learning behaviour. It is therefore interesting to look at learner engagement in relation to MOOC completion rates, and the effect of adaptive learning on a learner's engagement. Finally, there is some evidence that a learner's satisfaction level can also be influenced by adaptive learning, and in turn influences MOOC completion rates. Because of this, it is also integrated as a variable in the hypotheses.

Though there are likely many other indirect factors that affect completion rates and are influenced by adaptive learning, this thesis will limit itself to engagement and satisfaction. Therefore, combining the factors above, the main research question of this thesis is as follows:

> *"How can a MOOC's completion rates be increased, directly, or indirectly by increasing learner satisfaction or learner engagement, through the implementation of an adaptive learning system?"*

3.1 Research Questions

To help answer the main research question, several sub-questions have been formulated. These sub-questions are as follows:

SQ1. What is engagement?

SQ2. What is an effective method for the implementation of adaptive learning?

SQ3. What is an effective implementation of adaptive learning for the domain of cyber security?

SQ4. To what extent does this adaptive learning implementation increase completion rates?

SQ5. To what extent is the increase in completion rates affected by an increase in learner engagement?

SQ6. To what extent is the increase in completion rates affected by an increase in learner satisfaction?

3.2 Hypotheses

Based on results and deductions from the reviewed literature, some hypotheses have been formulated with regards to the interrelationship between adaptive learning, engagement, satisfaction, and completion rates. In this section these hypotheses will be enumerated and elaborated. Hypotheses will be referred to using their abbreviated form (*e.g.* Hypothesis 1 is referred to as H1).

H1: Adaptive learning has a positive effect on completion rates
Null H1: Adaptive learning does not have a positive effect on completion rates
This hypothesis assumes that an adaptive learning system has a positive effect on a MOOC's completion rates. This effect could be explained by a number of currently unidentified, indirect factors. The positive effect could for example be achieved by providing personalised help messages about the system when the user needs them: A lack of available help and a lack of digital and learning skills were all shown to be motivators for MOOC dropout [9], [11], [12]. Therefore, an adaptive system like this would reduce dropout caused by frustration or inability to handle the system. Furthermore, Fidalgo-Blanco *et al.* [13] found that implementing adaptive learning capabilities in a MOOC (among other capabilities) significantly raised its completion rates.

H2: Adaptive learning has a positive effect on learner engagement
Null H2: Adaptive learning does not have a positive effect on learner engagement
This hypothesis is based on the assumption that learner engagement can be increased through adaptive learning. By providing only learning material (like videos or challenges) that is relevant to the learner at that point in time, it is assumed that the learner is more likely to engage with this material than when the learner is confronted with all possible material at once. This would be in line with the findings of Szafir *et al.* [44], who found that adaptively reviewing only selected study material is a better approach than reviewing all material.

H3: Adaptive learning has a positive effect on learner satisfaction
Null H3: Adaptive learning does not have a positive effect on learner satisfaction
It is hypothesised that an adaptive learning system can increase user satisfaction by providing the right level of challenge for a specific learner. Since one of the main motivators for MOOC enrolment is personal challenge [11], this is assumed to increase a learner's satisfaction with the system. Furthermore, adaptive learning can omit unnecessary learning material which lowers the time a user needs to learn new material. This should further contribute to a high learner satisfaction. Fidalgo-Blanco *et al.* [13] found that implementing adaptive learning (as well as other capabilities) increased users' satisfaction about their learning.

H4: A high learner satisfaction has a positive effect on completion rates
Null H4: A high learner satisfaction does not have a positive effect on completion rates

This hypothesis is based on the assumption that learners who are highly satisfied with a MOOC are more inclined to complete it. If this assumption is true, a high user satisfaction rate would directly contribute to a high completion rate. Supporting this, Alraimi *et al.* [32] found that satisfaction was the third strongest predictor for a learner's intention to continue following a MOOC, after perceived reputation and perceived openness.

H5: A high learner engagement has a positive effect on completion rates
Null H5: A high learner engagement does not have a positive effect on completion rates
It has been shown in section 2.1.4 that engagement is a measure of a user's learning behaviour, as well as a predictor of potential dropout. It is hypothesised that increasing engagement (*i.e.* increasing the amount of interaction with learning material, as well as positively changing learning behaviour) also increases MOOC completion rates. This follows from the assumption that users who are highly engaged with a MOOC are also committed to finishing it. This hypothesis also follows from the assumption that learners drop out once engagement falls below a certain threshold. If these assumptions hold, it is also true that raising a learner's level of engagement reduces their chance of dropping out of the course.

3.3 Conceptual Framework

The conceptual model of this research, which comprises all hypotheses, is depicted in figure 3.1.

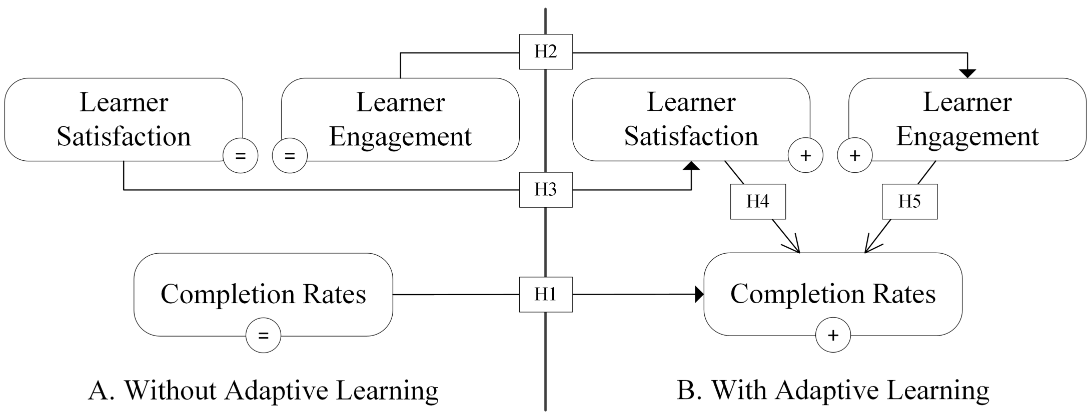


Figure 3.1: Conceptual Model

Figure 3.1 shows the expected effects of the introduction of Adaptive Learning on the other variables from the hypotheses. As can be seen, it is expected that all variables experience a positive effect (*e.g.* learner satisfaction is increased because of the adaptive learning system). In addition, the increase in learner satisfaction and engagement is expected to further increase completion rates. However, it should be noted that the direction of these relationships is merely hypothesised. Since the variables will be measured separately, these directions cannot be explicitly proven.

As can be seen in the conceptual model, there is a high expected level of interrelation between the various hypotheses and variables. By testing all hypotheses individually, the indirect effects of the implementation of adaptive learning on MOOC completion rates can be determined. However, the increase in completion rates may be explained by other indirect factors which are not tested within this research model, as it is not within the scope of this thesis to provide an exhaustive list of indirect factors. Therefore, it is not possible to distinguish the effect on completion rates that is caused by adaptive learning directly. To account for this, other possible explanations, as observed during the case study, will be provided. Once the relations between these variables are known, the sub-questions and main research question can be answered. To assist in the measurement of variables, measures are defined in the next section.

3.4 Measures

To assist in measurement and make the variables used in the conceptual model explicit, measures have been defined for every variable. These measures are shown in table 3.1 and are used to guide data collection.

Variable	**Measure(s)**
Adaptive learning	Whether or not the 'Adaptive Learning' condition has been implemented.
Learner engagement	"How you learn": Type of interaction with videos, assignments and help items. Due to the complexity of this construct, the exact measures are determined at a later stage. (self-reported, or by means of a quantitative construct)
Completion	Whether or not a learner has completed the MOOC (dichotomous), what percentage of the system they completed (percentage), or whether they dropped out (dichotomous).
Learner satisfaction	A learner's perceived satisfaction with the system (1-7 rating, self-reported)

Table 3.1: Research variables and their measures

4 Research Approach

In the first phase of this project, an adaptive learning system was designed and developed for use in the case study. The design of the adaptive learning artefact followed the guidelines of design science in information systems research proposed by Hevner *et al.* [47] where appropriate. These guidelines are listed in Table 4.1. The relevance of the system to be developed (guideline 2) is considered established by the literature study and problem statement provided in this thesis. The remaining guidelines have been applied throughout the design process.

To structure the development of the adaptive learning system, an adapted version of the waterfall software development model [48] has been applied. This model was chosen because of its clear phasing, which help with planning a project in advance. The "maintenance" phase of the traditional waterfall model has been left out, since it is not relevant in the context of this project. Furthermore, the phases have been renamed to 'steps' to prevent confusion with the phases of the project described in this document. The development process is visualised in Figure 4.1. The activities and outcomes for every step are described in the following subsections.

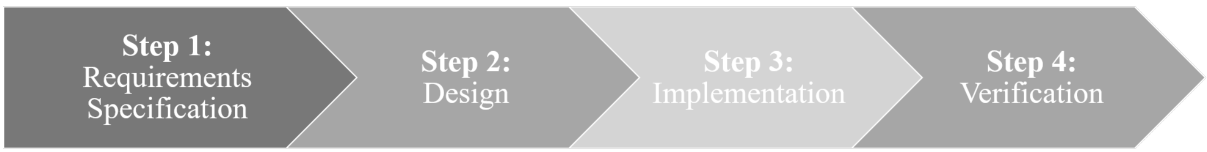


Figure 4.1: System Design and Development Process

To limit the amount of content in this chapter, the specifics of the design process have been included in a design specification document which is provided separately with this thesis[1]. This chapter discusses the main design decisions and outcomes of every step.

4.1 Requirements Specification

In this section, the requirements for the adaptive learning system are specified. The goal of this requirement specification was to provide a clear overview of the necessary features of the system and to allow for upfront prioritisation of features. The requirements are (partially) based on use cases, which are discussed in the next section.

Furthermore, the requirements were also based on the following criteria. These criteria were considered the bare minimum features that the system needed to have in order to

[1] If, for any reason, the design specification was not provided with your copy of this thesis, it can be downloaded separately here: `https://bit.ly/MScDesignDoc`

Guideline 1.	**Design as an artifact**. Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2.	**Problem Relevance**. The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3.	**Design Evaluation**. The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4.	**Research Contributions**. Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5.	**Research Rigor**. Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6.	**Design as a Search Process**. The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7.	**Communication of Research**. Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Table 4.1: Design Science Research Guidelines. Reprinted from Hevner *et al.* [47]

adhere to the provided definition of adaptive learning and satisfy the requirements for the stated hypotheses:

- Provide the right level of challenge to the learner [Hypothesis 3]
- Omit learning material that is not relevant to the user [Hypothesis 2]
- Provide help regarding learning material [Hypothesis 1]
- Provide help regarding the system itself [Hypothesis 1]

The resulting list of requirements is discussed below, the complete list included in the accompanying design specification. The deliverable of step one was a light-weight requirements specification document, which contains all prioritised requirements. This specification document was used as an input for the next steps.

4.1.1 Use Cases

In this section brief use cases are provided which were applied to support the externalisation of requirements and establish a clear view of the system's purpose. First, the system is described from the user's perspective. Afterwards, the administrator/developer's perspective is provided. To maintain readability, no formal use case notation was utilised.

User Perspective

The user is contacted through their institution with an invitation to participate in an experiment about an online hacking course. Out of interest, they follow the web link in the invitation (`www.leer-hacken.nl`). The landing page explains the system, but not the actual goal of the experiment. The user completes a simple username/password registration, accepts a brief informed consent and fills in their demographic survey in a short pre-survey. Afterwards, they are presented with a series of open-ended challenges about hacking. These challenges are preceded by informative sections, but they can be skipped if they are too hard for the user. The user is unaware of the adaptive or control condition they are in. After completing a fixed number of challenges, the user is presented with a post-survey with questions about their perception of the system. Finally, the user is presented with their score. After this, they can close the test.

Alternatively, the user may decide to drop out during the hacking challenges. To that end, a button which allows the user to drop out is shown to the user at all times. Should the user click this button, they are presented with a confirmation which, if confirmed, brings them directly to the post-survey.

Administrator Perspective

The administrator's main concern is the maintainability, stability, and integrity of the system and its data. The administrator is able to sign in through a sign-in screen. After authentication, they are able to access a control panel where the flow of the test, questions and data can be managed. From here, the administrator can make any necessary alterations to the system, review gathered data, backup the system, or download the database for analysis.

4.1.2 Requirements

Based on these use cases a list of system requirements was formed to support the design and development steps. The requirements are divided into three main categories: 'Functional requirements', 'Educational and content requirements', and 'Technical requirements'. The distinction between functional and quality requirements is also made: Functional requirements specify some form of system functionality or behaviour, where quality requirements relate to system performance.

In addition to being ordered based on content, the requirements were also prioritised in accordance with the MoSCoW-method [49]. By applying this low-cost prioritisation method, a clear view of requirement importance was formed. The aim of doing so was to create a clear distinction between essential and non-essential requirements for later in the design and development process.

An overview of requirement counts in each category is provided in Table 4.2 below. As mentioned, the complete list of requirements is included in the design specification. As becomes apparent from this overview, there is a vast majority of functional requirements, the requirements are spread over all type categories, and most requirements are classified as high priority ('must have').

Requirement Type	Priority	Functional	Quality
Feature	Must Have	5	0
	Should Have	2	0
	Could Have	2	0
	Won't Have	2	0
Educational/Content	Must Have	3	3
	Should Have	3	0
	Could Have	4	0
	Won't Have	1	0
Technical	Must Have	5	1
	Should Have	1	1
	Could Have	2	0
	Won't Have	2	0
	Total	**32**	**5**

Table 4.2: Requirement Overview

4.2 System Design

To ensure that the experiment went smoothly, measured the right variables, and the conditions were set up correctly, the system had to be designed and realised in the proper manner. In the 'system design' step, the design of the system was established and visualised from different perspectives, based on the output of the preceding step. Furthermore, the rationale behind various design decisions (which are not covered by the requirements listed in the previous chapter) was documented to provide insight into the design process.

First, the experimental design was elaborated on. Subsequently, the experimental conditions and their implications on the technical implementation of the system were established. This includes the selection and configuration of used adaptive learning algorithms. Finally, the survey elements of the system were detailed.

The system designed in this project is comparable to a prototype, since it is developed solely for this case study. Therefore, the design phase did not cover every aspect of the system in detail. Most low-level design aspects were left for the implementation step.

4.2.1 Experimental Design

The main research question involves three dependent variables that were to be measured in the experiment. These variables are 'Completion Rates', 'Engagement' and 'Satisfaction'. These variables were measured and determined based on data gathered from the system itself (measured) and a survey (self-reported). Figure 4.2 shows which variables were measured in what way. The measurements of learner engagement and completion rates are further elaborated in Chapter 6.

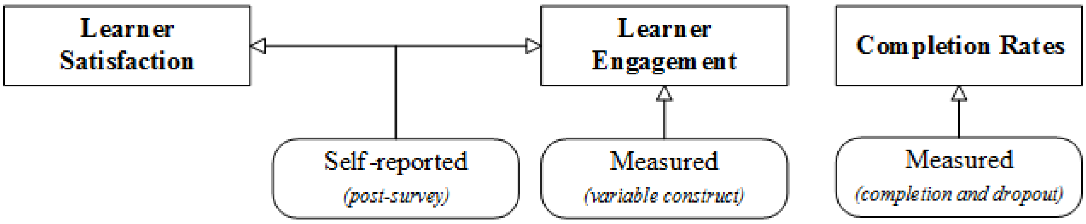


Figure 4.2: Variable Measurement Schema

As mentioned, the system consists of one linear (control) condition, one adaptive (experimental) condition, and a randomised (calibration) condition. The results from these systems were compared to establish the impact of adaptive learning on the dependent variables mentioned above. To accurately compare the scenarios, many factors must remain the same over the three conditions. An overview of the variables that were expected to be influenced by the experiment is provided in Table 4.3. The research variables that are included in the hypotheses are denoted by a question mark.

	Linear	**Random**	**Adaptive**
Difficulty (Challenge)	=	=	+
Speed	=	=	+
Perceived Satisfaction	=	-	+?
Perceived Engagement	=	=	+?
Completion Rates	=	-	+?

Table 4.3: Influence of Experimental Condition on Variables

As can be seen, there are two key factors which are expected to be influenced by the adaptive learning system: The difficulty (increased, depending on a user's performance) and time (decreased due to the adaptive routing through challenges). Furthermore, it was hypothesised (as denoted by the question marks) that the alteration of these factors would increase the main dependent variables: Perceived satisfaction, perceived engagement, and completion rates. These assumptions follow from the summary of dropout reasons in Section 2.2.4, where course difficulty, inflexibility, workload, and a lack of time are all listed as reasons for dropout.

4.2.2 Experimental Conditions

To structure the MOOC and ensure coherence in the order of challenges, the MOOC was divided into several modules. Each module consists of theory and challenges about one distinct subject. The adaptive learning element was implemented at the module level, so it determined the learner's skill level on a per-module basis. Applying this structure ensured that the flow of the MOOC remains logical in the adaptive condition, and that theory and explanation elements did not have to be shown multiple times. However,

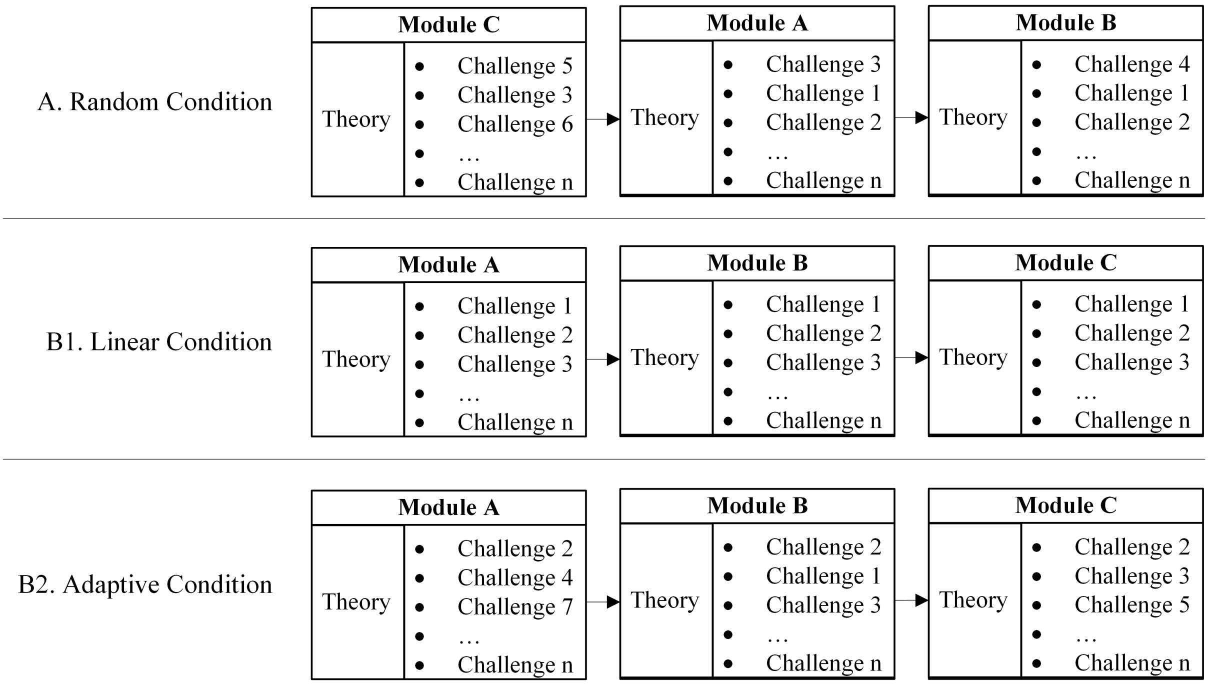


Figure 4.3: Schematic Representation of System Design

ordering the system in modules constrains the adaptive learning system, requiring that more challenges per module are available to reach the same level of adaptivity.

The system contained three experimental conditions, divided over two experimental phases. To facilitate an experimental analysis of the effects of adaptive learning, the adaptive and linear (control) conditions were implemented side-by-side in the same phase. However, the adaptive learning system required difficulty estimates of the challenges, which can be derived from completion data on the challenges. To obtain this data and subsequently calculate the challenge difficulties a third condition was implemented *prior* to the other two conditions: this was the random condition. The random condition was optimised for the collection of data before the actual experiment started, as users (and subsequently observations on the various challenges) were equally spread over the modules and challenges. A schematic representation of the system's design is shown in Figure 4.3. Further specification of the three conditions is provided in the following sections.

Random Condition

In order to serve challenges at the right difficulty level to the users in the adaptive condition, all items in the pool of available challenges needed to be calibrated. In this context, calibration entails the calculation of item difficulties from completion data. The process of calibrating the question difficulties is described in Section 4.3.6. To calibrate the challenges, binary completion data (the amount of times a challenge has been completed correctly and incorrectly) was required. To gather this data, a separate measurement phase had to be conducted prior to the execution of the adaptive and linear phase of the experiment.

The random condition was equal to the linear condition in length, but instead ordered both modules and assignments randomly. The randomised order of the assignments ensured an even spread of participants over challenges. This in turn ensured that all challenges had a number of observations that is as high as possible, thus maximising the accuracy of difficulty estimates. Furthermore, the random condition added to the overall amount of collected data and enabled comparison between participants in the randomised and linear conditions.

This condition was served to one institution, which was projected to contain a maximum of one-third of total participants based on participation estimates. This ensured that the acquired measurements are accurate and that participants are uniformly divided over all conditions, while keeping most participants in the 'main' conditions.

Linear Condition

As control condition, the linear condition was designed to be representative of the Hacklab MOOC as it is designed (and, at large, the way classical tests are structured). In this condition, learners followed a linear path through modules and challenges. Just like the random and adaptive conditions, each module consisted of a theoretical section and a set amount of challenges. However, in the linear condition the challenges were ordered logically from easy to difficult. This order was based on expert judgement; a combination of the existing order of challenges and experience of the author. Once a user completed all the challenges in a module (correctly or incorrectly) they were brought to the next module. The user also had the option to skip questions that they considered too difficult (which were then marked as wrong), or drop out of the system altogether (which brought them to the system's post-survey as per the requirements listed above).

Adaptive Condition

The adaptive learning condition of the system was similar to the other conditions, with the exception that the challenges within the modules were served in an adaptive manner. This was realised by dynamically providing challenges using Computerised Adaptive Testing (CAT) technology, and using Item Response Theory (IRT) for estimating challenge difficulties and ability levels. For specifics on the implementation of this condition, refer to Section 4.3.7.

As described by Weiss [50], CAT works by assessing the learner's skill level and continually updating this ability estimate. The ability estimate is based on the user's performance on questions while accounting for their difficulty levels. The more questions a user answers, the more accurate their ability estimation becomes. New questions are chosen in such a way that, given the difficulty of the question (parameter b) and the user's estimated skill level (θ), the estimated probability of the user solving the challenge (p) becomes approximately 0.5 (50%), as questions around this probability provide the most information about a user's skill level.

Motivation and Success Rate In order to increase user satisfaction, engagement, and completion rates it is important that the user stays motivated. A success rate of 50% may seem optimal from a CAT-perspective as it provides the most information about a learner's skill level, but a user may perceive this as a low rate of success since they are used to much higher success rates in classical tests. According to Linacre [51], this can be a "traumatic experience" for learners. This effect can be reduced by raising the success rate of users to 60%, 70%, or even 80%. However, doing so impacts the amount of questions needed for accurate results. Eggen *et al.* [52] found that raising this number to 80% or beyond has an even worse effect on system efficiency than random selection of questions. However, raising the predicted success rate to up to 70% has an acceptable impact; a 10% to 20% increase of the amount of questions required compared to a success rate of 50%.

Stopping Rules and Duration There are various stopping rules available for CAT tests, which have an impact on the duration of these tests. The tests can stop after a set number of items, a set amount of time, or once a set estimation accuracy has been reached. Due to the relatively low amount of challenges and high expected variability in challenge duration, stopping based on estimation accuracy or time would have likely introduced biases in gathered results. In addition, stopping based on accuracy or time would have made it more difficult to compare linear and adaptive scenarios. Therefore, the stopping rule for the adaptive learning system was decided to be a fixed number of questions.

In order to set a realistic and comparable amount of challenges for the adaptive condition, the increase in learning efficiency caused by the adaptive learning system needed to be considered. To make a fair comparison between the conditions, this increase in efficiency needed to be estimated and translated to a fixed number of assignments for the adaptive condition.

Vispoel *et al.* [53] found that, with optimal item selection of test items from the same item bank, learners in adaptive systems needed 50% to 93% fewer items to achieve the same level of reliability and validity when compared to fixed-item tests in an auditive testing environment. In addition, they mention 10 further case studies that all found that CAT tests can reduce test length by *at least* 50%. However, these percentages are based on the optimal selection of items. Due to the various constraints of the developed adaptive system (*e.g.* the relatively low amount of challenges and modular structure of the system), the selection of items cannot be considered optimal. Therefore, this percentage was reduced to 40%. The amount of challenges in each module was decreased to a logical number based on this percentage (refer to Section 4.3 for specifics). For example, the 'Web Hacking' module was decreased from 12 to 7 challenges (a 42% decrease).

4.2.3 Survey Design

As stated in the requirements, the experimental system needed to capture the user's demographics and experiences with the system. To realise this, two surveys were implemented in the system. Both surveys have different goals: The first survey ("pre-survey") served to capture a learner's demographics and expectations about the system, whereas the second survey ("post-survey") was to be filled in after a learner had either completed all challenges or dropped out of the system. The latter captured the user's experience with the system, including their self-reported satisfaction and engagement scores.

Pre-survey

The pre-survey was implemented to capture the user's demographics and some of their expectations about the Hacklab MOOC, so that different types of learners could be distinguished and compared during analysis. The pre-survey was shown to the user after registration as a simple form where all fields were mandatory. The main traits that had to be distinguished through this survey were 'Age and 'Institution', as these properties were used to distinguish population groups in the analysis phase. In addition, the traits 'Gender', 'Experience', and 'Interest' were added to provide extra information about the learner population. The questions asked in the pre-survey are listed in Table A.1 in Appendix A.

The questions were designed to be as basic and unambiguous as possible. The input types ranged from (validated) free input to dropdown boxes, in order to guide the user with choosing their answer. One notable input range was provided with the question for experience, "How much experience do you have with Cyber Security?". Here the user had the choice of the options 'None', 'A Little', 'Some', 'Quite Some', and 'A Lot'. These response options are non-standard, but were chosen to reflect a five-point Likert scale specific to experience. The response options were ordered logically to ensure the user understood the implication of their choice.

Post-survey

The post-survey served to capture the user's experience with the system. It was implemented in such a way that the independent variables 'Satisfaction' and 'Engagement' were measured through this survey. Therefore, it was of high importance that users complete this survey after either completing the system or dropping out. To support this goal, the survey was kept as short as possible and users were immediately redirected to this survey after dropping out. During experiment execution, users who did not complete the survey were be identified and contacted by e-mail with a reminder and instructions to finish the post-survey.

The survey attempted to capture the elements listed below. The elements printed in bold are the two independent research variables of this study. The factors to be measured for each variable were derived from the reasons for MOOC dropout established in the literature review in Chapter 2.

- **Learner Satisfaction**:
- Course content [9], [11]

- Course difficulty [9], [11]
- Course duration and workload [10], [12]
- Quality of available help and guidance (theory) [9], [11], [12]
- **Learner Engagement**:
- Course content [9], [11], [12]
- Interaction with available help (theory) [11], [12]
- Motivation to complete challenges [11], [12]
- Motivation to finish course [11], [12]

The post-survey was shown to the user after completing the system or dropping out. The user was then asked to fill in whether or not they agreed with various statements that relate to the above subjects. To maintain usability and simplicity, answers for all questions were recorded on a five-point Likert scale (Strongly Disagree, Disagree, Neither Agree Nor Disagree, Agree, Strongly Agree). The statements provided in the post-survey are listed in Table A.2 in Appendix A.

The questions on the trait 'content' were based on the work of Peltier *et al.* [54]. However, there were no validated questions available in literature on the remaining traits. Therefore, these questions were carefully formulated in a manual fashion. During the validation phase, these questions were checked for ambiguity with trial users.

4.3 System Implementation

In this step the system was developed, implemented and deployed. The system was implemented in an online environment and based on an open source platform for adaptive learning. This was the most time-consuming step of the system development process, as the system had to be implemented in accordance with the design specified earlier.

The output of this step was a working version of the experimental prototype of a MOOC on hacking (for more information on the case study subject please refer to Chapter 5). Only key steps of the development process have been documented included in this thesis. For a more elaborated description of the development process and output please refer to the accompanying design specification.

4.3.1 Modules

The challenges were divided over several modules based on their subject, see Section 4.2.2. Three modules were chosen to be implemented in the experimental prototype: 'Web Hacking', 'Network Hacking' and 'Encryption'. By choosing these three modules a broad range of challenges could be provided, while limiting the length and development time of the system.

These modules were mainly chosen for their suitability within the context of this project. For these modules, the author's experience was the highest and most challenges were readily available within an online environment. Furthermore, it was expected that these challenges would provide a broad range of observations: Based on observations of the traditional (offline) Hacklabs organised by Deloitte, 'Web Hacking' was likely to be perceived as the most interesting and "exciting" module by learners with no prior experience, whereas 'Encryption' might be considered more difficult or less interesting. 'Network Hacking' was expected to be somewhere in between the two other modules. This broad spectrum allowed for observations about the effect of learning materials to be made during the experiment.

Module 1: Web Hacking The web hacking module involved the user finding and exploiting vulnerabilities in online environments. The module consisted of 13 challenges in three different online environments. The challenges ranged from very easy to hard difficulty. The web hacking module had a wide variety of challenges and required skills, ranging from looking through (hidden) webpages to guessing passwords or exploiting vulnerable websites. To account for this variety, a comprehensive documentation section was provided to ensure that learners are pointed in the right direction and have the means necessary to solve the provided challenges. For more details on the documentation please refer to Section 4.3.2.

Module 2: Network Hacking and Forensics The network hacking and forensics module revolved around analysing and exploiting local network vulnerabilities. Since it is hard to simulate a real network on demand within the limitations of this project, the decision was made to work with local "network capture" (.pcap) files which can be opened in the program Wireshark[2]. The user could then analyse traffic in this file (also called network forensics) to answer various questions. Documentation on the installation and use of Wireshark was provided. The module consisted of seven challenges which utilised five different network capture files. The difficulty of challenges ranged from moderate to hard. In addition, most users also had to learn the basics of Wireshark. Therefore, this module was expected to be somewhat more difficult than the web hacking module.

Module 3: Encryption The third and final module of this prototype, encryption, was about cryptography and deciphering various encoded and encrypted types of text. The encryption section of the documentation page provided the user with various examples of cryptographic ciphers, so that even inexperienced users could recognise them and solve the challenges. The module consisted of six challenges, each with their own file and unique encryption type. The challenges ranged from easy to very hard. Overall, the module was expected to be perceived as slightly easier than the network hacking module.

[2]Wireshark is a free network traffic and protocol analysis program available at `www.wireshark.org`

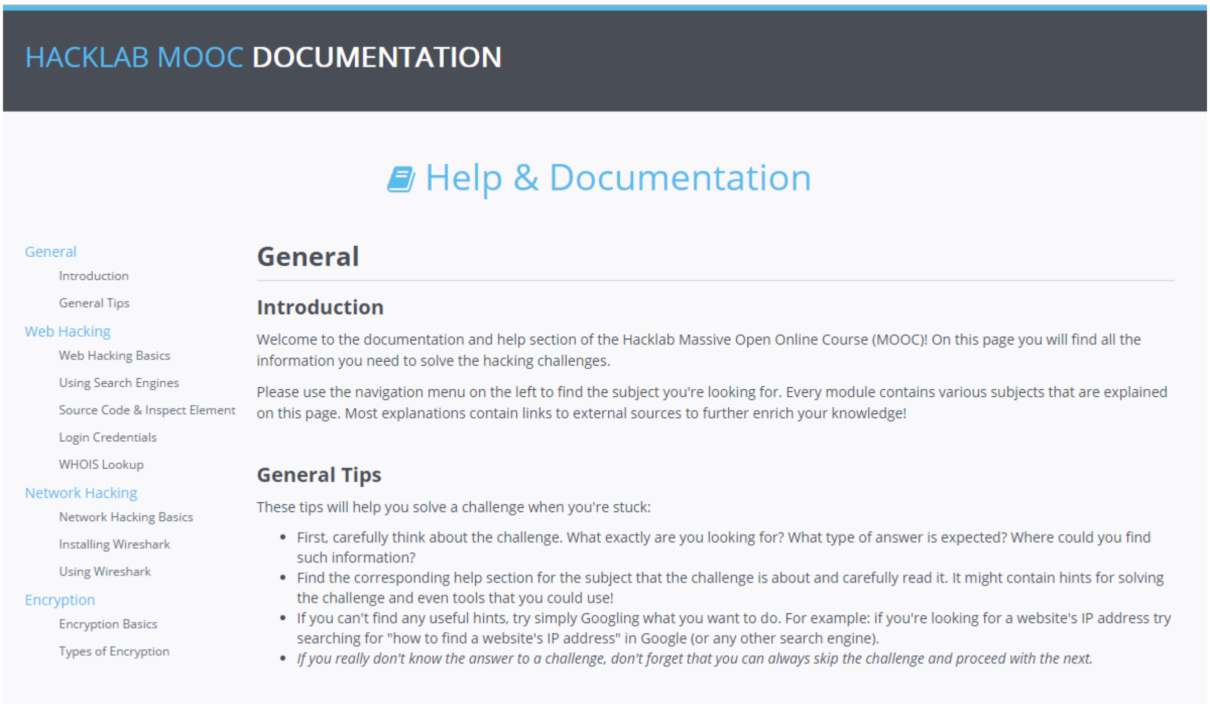


Figure 4.4: Documentation Page Landing Section and Navigation

4.3.2 Theory and Documentation

To complement the hacking challenges and provide beginning learners with the necessary knowledge to complete said challenges, a theory section was developed. This theory section served to make all learning material available in an online form. It contained the content that was explained orally in the hacking workshops, as well as basic step-by-step explanations for each challenge type and additional content to provide additional context and depth to the learning material. The documentation pages were written from scratch in order to optimally connect with the hacking challenges provided in the MOOC. However, to complement the provided theory, many links to external help sources were also provided throughout the documentation.

The content of the documentation was available at `https://www.leer-hacken.nl/help/`. Learners were provided with a button that links to the relevant section of this page. Usage of this button was monitored to provide insight in user behaviour. The navigation and landing page of the documentation section are shown in Figure 4.4.

4.3.3 Technical Implementation

The choice was made early in the implementation process to base the system on an existing platform for (adaptive) learning. The open-source platform for adaptive learning Concerto [55] was chosen as it fit the requirements of supporting CAT and IRT. Furthermore, Concerto showed a high degree of customisability which allowed for the implementation of custom features and tweaks. Concerto is based on the statistical programming language R [56] and the R package CatR [57], [58] to enable CAT functionality.

The adaptive learning system was realised through a cloud-based environment. Ubuntu 14.04 and Concerto platform version 5.0 beta 2.186 were installed on a small-tier, easily scalable virtual machine in Amazon's Elastic Compute Cloud (EC2) [59], using a pre-built Amazon Machine Image (AMI).

In order to serve the assignments to the user, an Apache web server was deployed on the same virtual machine. Most of the assignments feature downloadable files, but some web hacking challenges feature vulnerable web pages. The "Web Exploitation - First Steps" assignments run from a Flask (a web-based framework based on Python) application, which was installed in a separate sub-directory.

To support the Concerto system and securely store data, a MySQL database was also running on the same virtual machine. This was done to minimise infrastructure costs, but it did make the virtual machine a single point of failure. This risk was minimised by creating regular full-image backups and local database dumps.

In addition, the virtual machine was monitored throughout the project's execution phases to ensure that it could withstand the load and did not show any signs of malfunction. To support monitoring, a custom dashboard was built through Amazon's "CloudWatch" feature, and an automatic warning message was enabled for periods of critical CPU usage (>80% for a period of five minutes). The monitoring dashboard is shown in Figure 4.5. More information on resource monitoring is provided in Section 4.4.

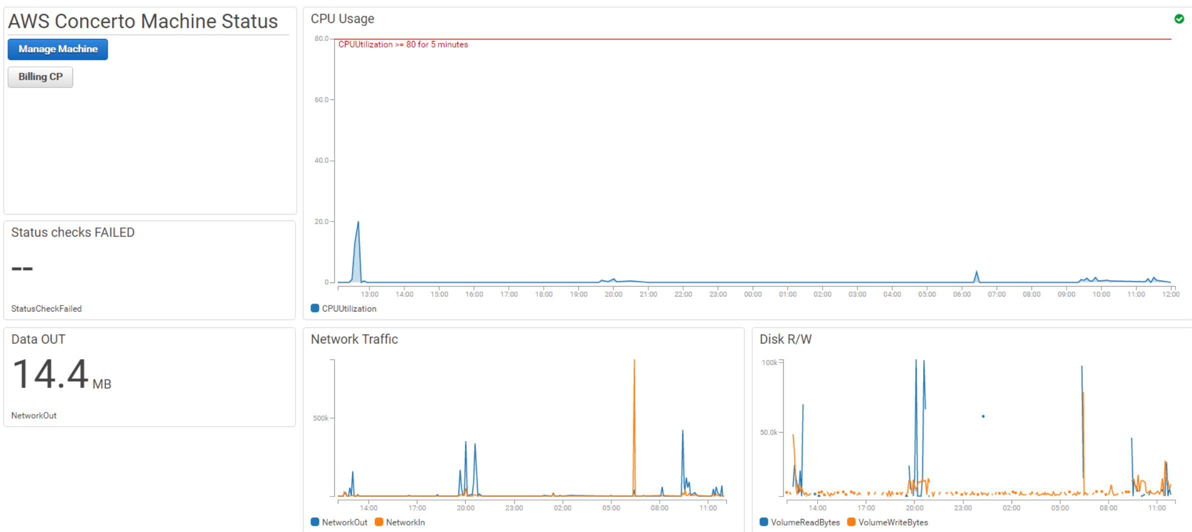


Figure 4.5: Amazon CloudWatch Monitoring Dashboard

To increase the accessibility of the system, the domain name `www.leer-hacken.nl` (learn-to-hack) was registered. Furthermore, a SSL certificate was installed on the server using Certbot[3] to increase security and ensure system integrity.

[3]Certbot by the Electronic Frontier Foundation is available at `https://certbot.eff.org`

4.3.4 Concerto Platform Modifications

As mentioned, the adaptive learning system was built on the adaptive learning platform Concerto. However, this platform did not support all required functionality. In order to realise the system in such a way that all requirements are met, the platform had to be adapted by altering or adding source code. Major modifications that were made to the platform are listed below. Extensive descriptions, including source code snippets, of the modifications are provided in the accompanying design specification.

- Implementation of an instant feedback feature
- Implementation a 'dropout' feature
- Implementation of a 'skip question' button
- Implementation of a help / documentation button
- Implementation case insensitive input fields
- Implementation of custom session tracking and saving
- Mirroring interfaces functionality of linear and adaptive nodes
- Fixing support for open questions in adaptive modules
- Various visual and logical tweaks

4.3.5 Test Implementation

As mentioned in Section 4.2.2, the three experimental conditions were divided over two experimental phases: one to calibrate the difficulty of the challenges and one to conduct the main experiment by comparing linear and adaptive conditions. For each phase, a Concerto test was developed using the flowchart editor introduced in Concerto version 5.

The randomised condition was developed first and used as a starting point for the implementation of the features described in the previous section. Once the first phase was considered final, a copy was made and used as a basis to create the second phase of the experiment. In this section, the implementations of the two phases is described. An extensive description and flowcharts of the implementation of these phases in Concerto is provided in the design specification.

Common Elements

Both variants of the system incorporated common elements, which were implemented in all phases of the system. Among these common elements were the 'introduction and informed consent' and the 'registration and login' modules. In addition, the surveys were also implemented identically over all phases. The pre-survey was implemented as a "form" module, which allowed for the combination of various question types. The post-survey was implemented in the form of a "questionnaire" module, which allowed for

Likert scale-type questions. In addition, the post-survey also included an optional form element for open feedback on the system.

The remaining elements differed per experimental phase. These elements are discussed in the following sections.

Phase One: Randomised Condition

The randomised module put each user in a random module. After the introduction of the system and registration screens, it selected a random, incompleted module and started serving questions from this module in random order. Once the user completed a module, they were sent to another randomly picked module until all modules had been completed (or the user dropped out). The system then proceeded with the post-survey and scoring screen, similar to the other conditions.

Phase Two: Linear and Adaptive Conditions

The second phase was in part similar to the first, with the main difference being the implementation of two parallel conditions. Up until the first module, this phase was identical to its predecessor. However, once the user reached the first module, they are randomly put in either the adaptive or linear condition. Since both conditions look exactly alike, the user was not aware of this division.

If the user was put in the linear condition, they followed all three modules in order. In contrast to the first phase, the questions within these modules were statically ordered based on expert judgement. Each module started with the same introduction as before, and the user still had the option to drop out at any time. If they finished all modules, their scores were saved and they were brought to the post-survey and subsequently the score screen.

If the user was put in the adaptive condition, they were brought to a different part of the system. The main difference was that they were not presented with "linear test" modules, but with "CAT" modules. These modules were configured to look and feel exactly the same, but perform an adaptive test rather than a linear one. Data from the first phase was used to calibrate the question difficulties for the adaptive tests.

In the adaptive condition, the user was presented with the same challenges as the other conditions, but they were adaptively ordered. Another key difference was that the Adaptive modules were approximately 40% shorter due to the higher measuring efficiency, as discussed in Section 4.2.2. If the user finished or dropped out, their score was saved and they were brought to the post-survey and score screen. The score screen for the adaptive condition was modified to account for the shorter duration.

4.3.6 System Calibration

As described in Section 4.2.2, a difficulty estimate had to be established for every question before the adaptive learning system could be applied in practice. To facilitate this, the random condition of the experiment was introduced. This condition was executed prior to the other two conditions and was designed in such a way that answers provided by users were spread over all questions. Doing so minimised the amount of required condition participants for accurate estimates.

The R package 'eRm' [60]–[62] was used to calculate the difficulty estimates. First, a binary response matrix was generated based on provided answers through an R script. As per the requirements of the 'eRm' package, some user entries and questions had to be removed from this matrix to prevent ill-formedness (the implications of this are explained below). Following that, the 'RM' function from the 'eRm' package was used to calculate the difficulty estimates based on a Rasch Model of the provided answer matrix. This returned a list of difficulty estimates that could be implemented in the Concerto Platform.

The 'eRm' package uses conditional maximum likelihood (CML) estimation to estimate the item parameters (difficulties) of the items (challenges)[60]. As described in Mair *et al.* [60]: "The main idea behind the CML estimation is that the persons raw score $r_v = \sum_{i=1}^{k} x_{vi}$ is a sufficient statistic. Thus, by conditioning the likelihood onto $r' = (r_1, ..., r_n)$, the person parameters θ, which in this context are nuisance parameters, vanish from the likelihood equation, thus, leading to consistently estimated item parameters $\hat{\beta}$." In this context, k is the number of items, and x_{vi} is a single element in the response matrix. In other words, CML works by estimating item difficulties based on the raw score of users in the matrix of items (X). Due to the complexity of the difficulty estimation formula, it is not further discussed in this thesis.

Initial difficulty estimates were based on 29 valid user entries, with a total of 284 item responses over 25 questions. To increase system accuracy, the difficulty estimates were recalculated twice when the second phase of the system was already operational. The final version of the difficulty estimates was based on 80 valid user entries, containing 885 valid responses over 25 questions.

Based on this data, the Rasch difficulty estimate for challenge ID 1 could not be calculated due to ill-formedness in the answer matrix, this challenge was therefore left out of the system altogether. Furthermore, the difficulty estimate for challenge ID 24 could also not be calculated since no user answered that challenge correctly. Since it was known from experience to be the hardest question, the difficulty was set at 2.5, the highest value. Finally, the answered were normalised in such a way that the mean difficulty of every module became 0. This was done to ensure that Concerto interpreted the question difficulties correctly over the various modules.

The final Rasch difficulty estimates are compared to initial expert estimates in Figure 4.6. It should be noted that the expert estimates were only transposed to fit on the same scale, the two estimates might therefore not be directly comparable.

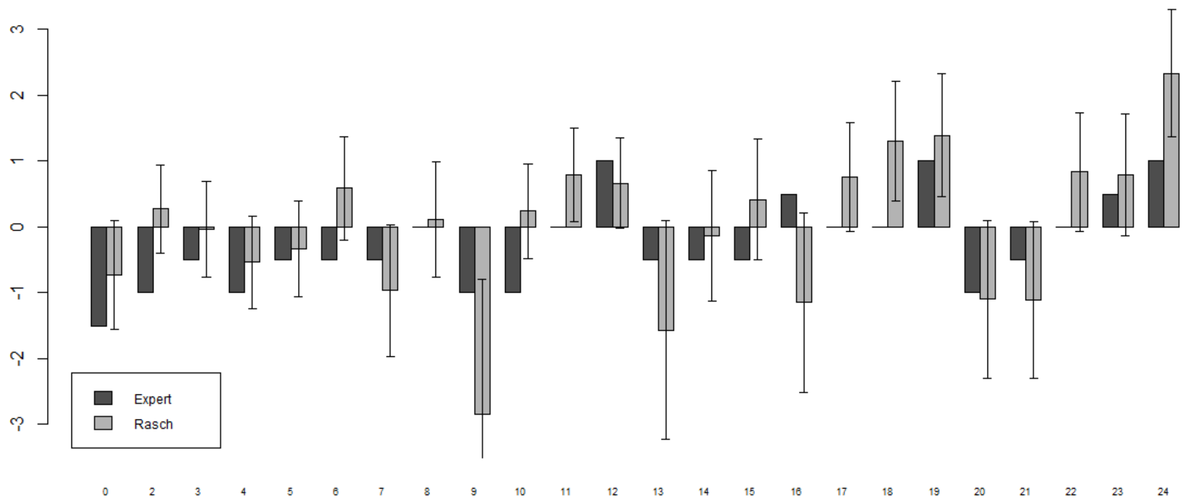


Figure 4.6: Expert and Rasch Difficulty Estimates Per Question

4.3.7 System Configuration

As mentioned, Concerto uses the R package 'CatR' [57], [58] to serve adaptive tests. CatR has various configuration options which determine the type of adaptive test provided within Concerto. In this section, the chosen configuration options and the underlying algorithms are briefly discussed.

IRT model CatR uses the four-parameter logistic model (4PL) of item response theory [57]. This model not only takes the item difficulty (described in the preceding section) into account, it also has three additional parameters for every item: a discrimination parameter (the slope of the item-characteristic curve), a pseudo-guessing parameter (the lower asymptote of the item curve), and the inattention parameter (upper asymptote). However, in this study the one-parameter logistic model (1PL) was implemented due to the relatively low added value of the 4PL model and to simplify the implementation of the adaptive system. To that end, Concerto was configured as follows: all pseudo-guessing parameters were set to 0, all inattention parameters were set to 1 and all discrimination parameters were set to 1.7, as recommended by Linacre [63].

Ability Estimation To estimate the ability trait of participants (θ), the Bayes modal (BM) estimator was used. Under the 1PL model, this estimator is equal to the Maximum likelihood (ML) estimator. The BM estimator was chosen over the ML estimator as it is the default estimator implemented in Concerto. As described by Magis *et al.* [57], BM works by maximising the posterior distribution of the ability level $g(\theta)$, which is a combination of the prior distribution of the ability level $f(\theta)$ and the likelihood function $L(\theta)$: $g(\theta) = f(\theta)L(\theta)$. Therefore, the BM estimate is the ability value $\hat{\theta}_{BM}$ that maximises the posterior distribution $g(\theta)$ or its logarithm:

$$\log g(\theta) = \log f(\theta) + \log L(\theta) \tag{4.1}$$

In the context of this project, the prior distribution of skill levels $f(\theta)$ was assumed to be the normal distribution. What remains is the likelihood function $L(\theta)$, which is equal to that of the ML estimator:

$$L(\theta) = \prod_{j=1}^{J} P_j(\theta)^{X_j} Q_j(\theta)^{1-X_j} \tag{4.2}$$

where $Q_j(\theta) = 1 - P_j(\theta)$ is the probability of an incorrect answer and J is the test length. For specifics regarding these equations please refer to the publication by Magis *et al.* [57].

Item Selection The next configuration parameter that determines the behaviour of the adaptive system is the item selection rule. This rule determines which item from the item bank is selected next and presented to the user. In the context of this project, Urry's rule for item selection [64] was applied, implemented in CatR as "bOpt". This rule simply selects the next item in such a way that the difference between the item difficulty (b) and user skill level (θ) is minimised. There are various other rules available in the CatR package, for example to maximise the information gained from the selected item, but for the purposes of this study Urry's rule was sufficient. It should be noted that for the first item, the user's skill level was set to zero (the prior mean ability level).

Stopping Rule The adaptive learning system keeps serving questions and updating the ability estimate until a certain stopping rule is satisfied. If this happens, the system stops serving questions, calculates a final ability estimate for the user, and proceeds to the next module. The system can stop when a certain level of estimation confidence is reached, after a certain amount of time, or when a certain amount of items is served. As discussed in Section 4.2.2, the stopping rule for this study was set at a fixed number of items. This was the only viable option in the context of this study, as stopping based on time or the achieved level of confidence would make the adaptive system incomparable to the linear system. The amount of items was determined per module, by subtracting approximately 40% from the amount of questions in that module in the linear condition. For specifics on this process refer to Section 4.2.2.

4.4 System Validation

After the first phase of the system was implemented, a validation process was started to ensure that all aspects of the system were functioning correctly. This included the system's front-end (mainly usability, accessibility and stability) and back-end (performance and data output). Matters that were considered important within the context of this project and required validation included:

- The system can handle at least the expected load
- Data gathered by the system is usable and complete

- The system's algorithms and other mechanisms function as expected
- The content provided is representative of the Hacklab MOOC
- The content provided is understandable for the target audience
- The system has a high usability and stability

Based on these points, the validation of the system was divided into four types: functional verification, user testing, load testing and data verification. To ensure that the system validation incorporates all (relevant) quality aspects of the system, the ISO 25010:2011 standard for software quality [65] was mapped to these validation types. This mapping is shown in Table 4.4. In the following sections the validation process and results are described.

ISO Quality Characteristic	**Validation Test(s)**
Functional Suitability	Functional Verification
Performance Efficiency	Load Testing; Functional Verification
Compatibility	Data Verification
Usability	User Testing
Reliability	Load Testing
Security	Functional Verification
Maintainability	(Functional Verification)
Portability	(Data Verification)

Table 4.4: ISO 25010:2011 validation mapping

4.4.1 Functional Verification

To ensure that the system contained all required functionality, a functional verification was performed using the requirements listed in Chapter 4.1. As the requirements specification contains many requirements regarding functionality, performance, implementation, and security, it is important that at least all high-priority requirements had successfully been implemented. To that end, all requirements were analysed to check the system for functional (and qualitative) completeness. An overview of the amount of (non-)implemented requirements is provided in Table 4.5.

As becomes apparent from this analysis, almost all high-priority (must have and should have) requirements have been implemented. The only exception is the technical requirement "The system should be able to handle a load of 10 simultaneous participants". This requirement has been marked as 'Not Implemented' since the exact load that the server can handle was unknown prior to launch (for details refer to Section 4.4.3).

Half of the low-priority (could have) requirements have been implemented. Requirements in this category that have not been implemented have generally been decided against because they required a high implementation effort but offered relatively little added value in return. Examples of this are the second variant of the adaptive condition, implementation of help videos, and Dutch language implementation. Furthermore, these

Requirement Type	Priority	Implemented	Not Implemented
Feature	Must Have	5	0
	Should Have	2	0
	Could Have	2	0
	Won't Have	0	2
Educational/Content	Must Have	6	0
	Should Have	3	0
	Could Have	1	**3**
	Won't Have	0	1
Technical	Must Have	6	0
	Should Have	1	**1**
	Could Have	1	**1**
	Won't Have	0	2
	Total	**27**	**10**

Table 4.5: Amount of (non-)implemented requirements

requirements contained one functional alternative that has been decided against (multi-stage testing rather than CAT).

As expected, none of the 'won't have' requirements have been implemented. The functionality described in these requirements could be adopted in the full version of the Hacklab MOOC.

4.4.2 User Testing

To ensure that assumptions made about learner behaviour during system development were correct, a simple user test was conducted by means of pre-launch trial runs. Several colleagues and friends of the author were invited to give the system a try. Given their experience with cyber security and information technology in general, these learners can be considered expert users when compared to the system's target audience.

Based on the performed trial runs (of approximately 6 users), several observations were made. The most important observation was that several users returned variations of the correct answer (for example a full web link or the correct answer with an extra space). One of the limitations of open questions in Concerto is that only one answer can be marked correct. Therefore, to avoid false negative answers, it is essential that the user inputs the right answer format. To achieve this, every question already included an 'answer format' and submission confirmation dialog ("Are you sure you would like to submit this answer? Please double check that it matches the expected answer format!"). However, to provide extra emphasis, the layout of the answer format was changed to attract more attention.

Another observation was that only one participant filled in the post-survey. This may be due to the fact that these users were asked to simply "give the system a try", but it could also be an indication that users were under the impression that they were finished before completing the post-survey. To avoid this issue extra care was taken during the experiment phase to ensure that users only stop after the post-survey.

Overall, users reported that the system appeared stable and was functioning well. In addition, some users made use of the session resume feature and reported no issues.

4.4.3 Load Testing

To ensure that the system could handle the load of several concurrent, active participants, a simple load test was performed on the server. This load test consisted of two aspects: a simulated load test and resource monitoring during the user tests.

The first test was a simulated load test of the Apache server running on the Ubuntu virtual machine. This test was conducted by running the 'ApacheBench' tool from the virtual machine. The server was able to easily handle several hundred requests (both HTTP and HTTPS) in rapid succession. However, the tool stopped due to bad requests after about 600 requests, most likely due to Amazon's infrastructure and load handling measures. During the tests, resource usage was carefully monitored and remained within normal operating levels.

The second aspect of the load test consisted of participant and resource observation during normal system use by the trial participants. Monitoring these resource usage patterns is important since the Concerto platform, R server and file hosting server running in the background put a different strain on the server than just the Apache server. During regular use with at most two concurrent active users, the machine did not experience any exceptionally heavy loads. CPU usage never exceeded 25%, memory usage remained stable at approximately a fifth of capacity and the disk was mostly idle. Based on these observations, the system was estimated to be able to easily handle at least five, highly active users at the same time. The machine's memory was established to be the main concurrency bottleneck.

One observation that was made during these tests was that hosting the downloadable challenge files would likely make the machine exceed Amazon's free limit for outgoing network data (1GB/mo). However, the costs for additional GBs are quite low and the usability advantage of simply downloading the file at high speeds with one click far exceeded the benefits of hosting these files externally. To monitor outgoing data network, an additional entry was made on the monitoring dashboard.

4.4.4 Data Verification

To ensure that the data provided by the learning system is accurate and usable, a basic analysis of the data output was performed. This analysis consisted of two aspects: modelling the database and a simple offline import and analysis test using RStudio.

Database Model

First, the database tables and their relationships were modelled in a simple relational database schema. This schema can be seen in Figure 4.7. The primary key (unique identifier) of table entries are marked bold and underlined, and references to other tables are denoted by arrows and underlined entries.

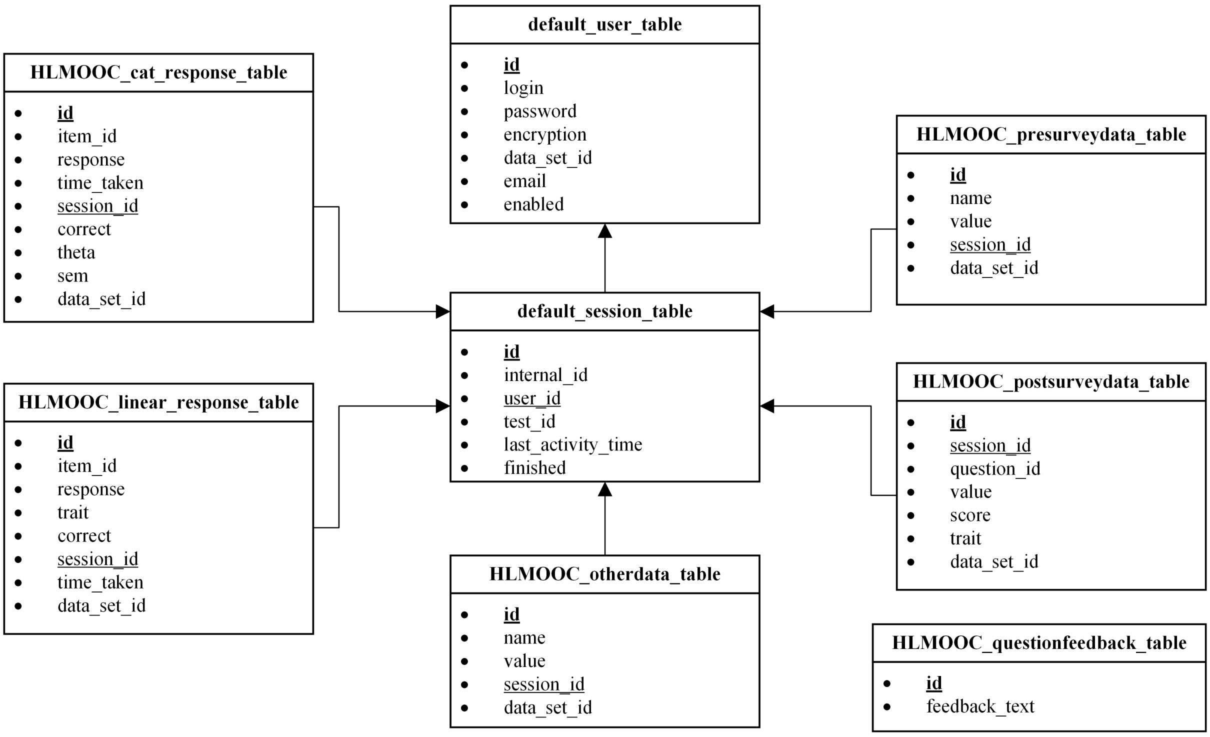


Figure 4.7: Relational Database Schema

It is worth noting that the question feedback table is not connected to the other tables since it only functions as secure storage for the feedback texts provided to the user if they get a question wrong.

Processing Data

To check whether or not the output data was usable, six relevant tables which contained initial data (users, sessions, linear test answers, pre- and post-surveys and other) were downloaded locally as CSV files and imported using a basic R script. This simple script was designed to return all session IDs for a single user ID in the old session structure. Since every user now only gets one session ID this code is now deprecated. However, it still allows for the verification of data formats and the linkage between tables. The script functioned as expected, returning the correct session IDs for the specified user and loading all database dump files into memory. This test showed that the database format and output type were usable and ready for analysis.

4.4.5 Monitoring

The validation tests described above provided a solid indication that the system was stable enough for release, but validation is by no means a one-time process. To ensure that the system remained in stable operation throughout the execution phase, the system was continually monitored using the Amazon CloudWatch dashboard described in Section 4.3.3. Key metrics that were monitored in real-time included 'CPU usage', 'network traffic in/out' and 'disk read/write'. Unfortunately, the Amazon CloudWatch platform does not provide insight into metrics like memory usage and disk space usage due to operating system limitations. Therefore, these metrics had to be manually monitored through the virtual machine itself.

During the first phase of the experiment, the system's memory was filled at one point when approximately 15 students connected at the same time. This memory overflow caused the system to crash, but a reboot of the virtual machine fixed the issue without any lasting effects. To prevent this problem from occurring again in later phases, the system was scaled up to the 'medium' package of amazon's elastic compute cloud. This added an extra virtual core and increased memory from 1 to 4 gigabytes, greatly increasing the capacity of the system. This allowed the system to handle approximately 20 concurrent users in the second stage of the experiment, never exceeding 20% CPU usage and 2.5GB of memory usage.

Besides key resource monitoring, regular checks were made to ensure that all data was stored correctly and users were progressing through the system in the desired manner. It was found that all data was stored correctly and data integrity was maintained throughout the process.

4.4.6 System Validity

In addition to the system validation steps that were taken before and during the execution of the case study, the system was also statistically validated after the experiment was conducted to validate that the adaptive learning system functioned properly and did not introduce any threats to validity. In this section, this validity analysis is discussed.

There are various ways to formally validate the fit of a calculated Rasch model as described by Mair *et al.* [61]. However, the provided data set was too small to run Andersen's LR-test [66], and the remaining, nonparametric tests could not be run on the Rasch model as the answer matrix contained NA entries (*i.e.* entries where the user did not provide the answer to a challenge).

To evaluate the model fit of the adaptive learning system without a formal test, a simulation test was run. For every user in the system, their estimated user skill level (θ), and the set of questions that they answered, a hundred performance simulations were run using the 'randomCAT' function from the 'CatR' package [57], [58]. The simulated scores were subsequently compared with the user's actual score. This comparison is shown in Figure 4.8.

Score was indicated on a probability scale of 0 to 1, where 1 means that the user answered all questions within their set correctly and vice versa. Users with an actual or simulated score of 0 or 1 have been removed from the set to prevent a simulation bias. The figure shows the logit function of probability p, $log(p/1-p)$, which indicates the log-odds of the actual and simulated probabilities of answering correctly. This transformation makes the graph more suitable for correlation and regression.

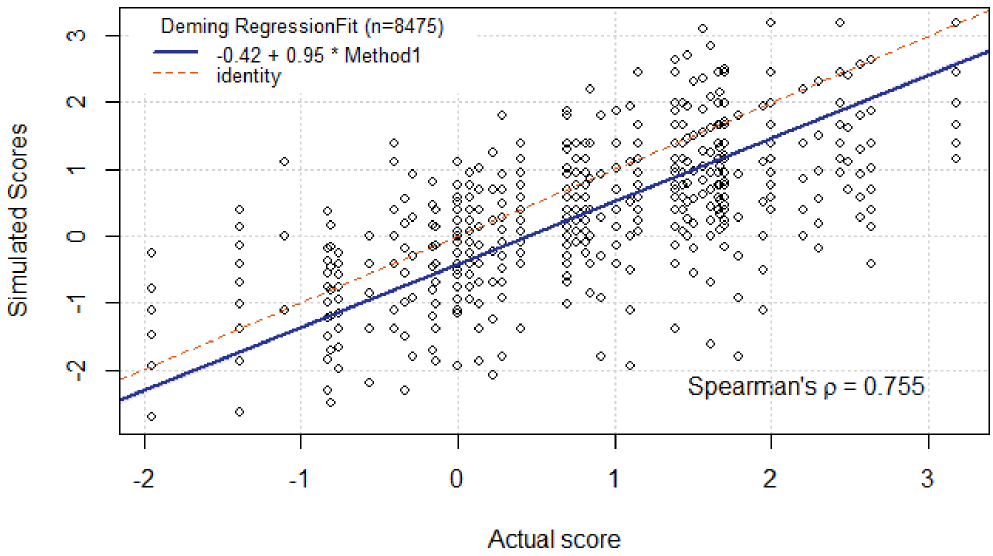


Figure 4.8: Actual User Performance Versus Simulated User Performance

As can be seen in this figure, the simulated values are highly correlated with the actual values, which indicates a good Rasch model fit and, by extension, show that the adaptive learning system is valid. A Deming regression was performed as it is very suitable for estimating systematic bias in a symmetric context [67]. The regression line shows a slight negative bias in simulated scores compared to the identity line, indicating that users scored slightly higher than would be expected. This is most likely caused by a slight negative offset in difficulty estimations caused by the experimental setup.

To further validate the performance of the adaptive learning system, the calculated standard error of user skill levels was compared between the adaptive and linear conditions. Even though users in the adaptive condition were provided with 40% less challenges than users in the linear condition (15 rather than 25), the calculated standard errors were not statistically different between the linear condition (M=0.47, SD=0.15) and the adaptive condition (M=0.48, SD=0.11); t(90)=-0.4761, p=0.6351. This result suggests that the adaptive learning system functions as expected and retains measurement accuracy, while reducing the amount of served challenges.

5 Experiment

To conduct the experiment and gather data for analysis, the designed adaptive learning system was applied in practice in the context of a case study. It should be noted that the system design and case study phases were executed in a nonlinear fashion, as steps like the technical implementation of the second version of the system and system monitoring were also executed while the system was operational.

In this chapter, the execution and results of the case study are discussed.

5.1 Case Study Description

After the system was designed, implemented and verified, it was applied in practice as a case study. The subject of the case study was a prototype of the Hacklab Massive Open Online Course (Hacklab MOOC). This MOOC is currently being developed by Deloitte Netherlands with the author being involved in its development and implementation. The designed system was based on this MOOC and integrated theory and practical challenges on cyber security in one system.

5.1.1 Origin and Need

The Hacklab MOOC is based on Deloitte's existing "Hacklab High School" formula [68], where high school students are invited for a day of theoretical and practical sessions on various subjects around cyber security. Subjects covered by Hacklab High School involve Google hacking, social engineering, privacy, web hacking, network hacking and ethical hacking. In these sessions, students are provided with theory and hands-on exercises in which they are given the opportunity to apply this theory in practice within a safe context. The focus on practical application of knowledge and tools allows learners to experience the life of an ethical hacker.

Hacklab High School sessions have been organised for over 600 students with a variety of backgrounds to date. These sessions were very well-received and fulfil their goal of enthusing young students about cyber security and generating awareness. However, Deloitte encountered the problem that the Hacklab High School concept was not scalable, as sessions have to be individually organised for groups with a maximum of 60 students, which requires a lot of effort. Therefore, scaling up this concept is not feasible.

5.1.2 Hacklab MOOC

The solution to the scalability problem encountered by the Hacklab High School concept is a new Hacklab proposition: The Hacklab MOOC. As established in this thesis, MOOCs are openly and widely available to anyone interested in following them. Therefore, the Hacklab MOOC could potentially be served to thousands of young learners worldwide, greatly increasing the amount of knowledge and awareness generated by the concept. The full proposition of the Hacklab MOOC is described in the following sections.

Mission

The aim of the Hacklab MOOC is twofold. The first goal is to promote cyber security as a field of work and study by enthusing young, computer-savvy learners about following a study, and by extension getting a job, in the field of cyber security. The second goal is to improve cyber security awareness and basic understanding for a broad range of future (IT-) professionals and to improve the overall visibility of the cyber security field by providing this course at an early stage.

Overall, the mission of the MOOC is to increase awareness, knowledge, and interest in the field of cyber security for the new generation of IT professionals.

Content

The Hacklab MOOC is primarily challenge-based with a strong focus on gamification. Participants need to be as resourceful and creative as real hackers. The participant is hired by the director of a fictive energy company called *Cybervoltage*. The company is facing a serious cyber-attack by an initially unknown actor, and its up to the participant to help the company. Throughout six modules on various difficulty levels, participants will learn about the basics of cyber security (awareness, phishing, cryptography, privacy, and hacking). In order to (im)prove their skills, participants also have to solve a number of challenges as part of these modules.

Audience

The target group for the HackLab MOOC consists primarily of students within the age group of 12 to 20 years, including pre-college to bachelor and current bachelor students. As mentioned, the course aims to prepare and enthuse these students for further education. There is no prior (hacking) knowledge required to follow this MOOC. The MOOC could also be interesting for current master students following an IT or cyber security programme.

5.2 Data Gathering

The developed system was presented to participants as an online prototype of the Hacklab MOOC described above. Through various channels, participants were asked to participate in testing the system and reporting their experiences. Participants were made aware through an informed consent form that data from the system was used for a scientific study. However, they were not told the goal nor the means of this scientific experiment.

5.2.1 Educational Institutions

Prior to the execution of the experiment, teachers responsible for computer science (or cyber security) at several educational institutions within the target demographic of high school to bachelor-level students were contacted to gauge interest in participation. Five institutions were found willing to let their students participate in this experiment, including two high schools, one secondary vocational institution, and two (applied) universities. Together, these institutions had a potential reach of some 350 to 400 students. An overview of the participating institutions is provided in Table 5.1.

Name	**Location**	**Type**	**Participants**	**N**
Alberdingk Thijm College	Hilversum	High school	Two computer science classes	40-60
Amsterdam University of Applied Sciences	Amsterdam	University of applied sciences	Students from a cyber security minor	60
Metis Montessori Lyceum	Amsterdam	High school	One computer science class	30
ROC Mondriaan	The Hague	Secondary vocational	Third- and fourth year application development students	100
Utrecht University	Utrecht	University	Students from the bachelor course ‘information security’	120

Table 5.1: Overview of participating institutions. N indicates the expected maximum amount of participants.

The first phase of the Hacklab MOOC prototype, which entailed the implementation of the random condition, was introduced at the Alberdingk Thijm College, a high school in Hilversum. This educational institution was chosen for the first phase as the projected amount of participants (40 to 60) was deemed a suitable amount for the calibration of question difficulty. Furthermore, running this phase with a small group of participants allowed the author to fix any errors with the system without them having a big impact on experiment execution.

The second phase of the experiment, containing the linear and adaptive conditions, was delivered at the remaining four institutions. This way a high number of participants

5.2.2 Delivery

To introduce and explain the system and increase participant engagement, several workshops on hacking were held by the author. To further complement the amount of participants gained from these workshops, additional email invites were sent to students at several institutions.

Workshops

A total of five hacking workshops were organised by the author in the context of this thesis. Two were provided to the Alberdingk Thijm College for the first phase of the experiment. For the second phase, three additional workshops were organised: two for the ROC Mondriaan in The Hague and one for the Metis Montessori Lyceum in Amsterdam.

The content of the workshops were tailored to the level of the participating students and the available time, but the overall structure of the workshop was similar. The workshops were divided in three theoretical sections followed by an optional practical session. The theoretical part of the workshops took approximately half an hour.

First, the theory and definition of hacking was provided and some examples were discussed. The difference between 'good' (white hat) and 'evil' (black hat) was explained and the importance of white hat hackers in society was explained.

Subsequently, the need for the Hacklab MOOC was briefly explained, and the students were asked for their help. The interesting aspects and additional incentives (certificates and prizes) were highlighted. Following this call to action, the system and its elements were explained to avoid confusion.

Finally, some practical hacking tips related to web and network hacking were provided if time allowed. These tips were designed to help the user along with their first steps regarding web hacking, network hacking and the required program Wireshark.

The workshops at ROC Mondriaan included a practical session where learners started working in the system individually. The author was present in the classroom to answer any questions they might have on the functionality of the system for approximately two hours. Help on the contents of the system was not provided to avoid a knowledge bias. Participation in the system was not mandatory, but since the students had to be present either way many students participated in the system for the duration of the session.

Email Invites and Reminders

To further increase the amount of participants involved through the workshops, email invites were sent out to students of two institutions: Amsterdam University of Applied Sciences and Utrecht University. Approximately 180 students received participation invites from their teacher. To improve response rates from these email invites, the author also provided a physical reminder to students of the Amsterdam University of Applied Sciences during one of their lectures.

In addition to these invites, email reminders were sent out to participants who registered in the system but did not complete it, including students who were introduced to the system through workshops. This was done to persuade these students to finish the system and its post-survey. A total of 150 reminder emails were sent on three separate occasions. One reminder was sent in the first phase and two in the second phase of the experiment.

5.3 Observations

Throughout the execution phase of the case study several observations were made, both during the delivery of the workshops and by looking at the incoming data. Some noteworthy observations are discussed in this section.

Participation

The number of active participants in the system exceeded initial expectations. Of the approximately 400 students that were invited to participate in the system, 156 registered an account in the system. 131 users answered at least one question and are considered 'active participants'. It should be noted that the vast majority of participants were introduced to the system through a workshop: Only 16 out of 131 active participants joined from an email invite. A schematic representation of the flow of participants is depicted in Figure 5.1.

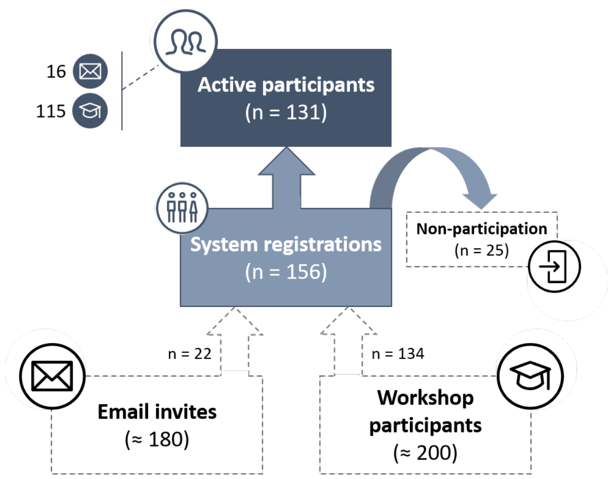


Figure 5.1: Participation Model

Even though response rates exceeded the expected level, many participants stopped participating the system after answering only a few questions or skipping many questions. It is assumed that these learners were simply 'sampling' the content of the system, which is normal behaviour for a MOOC as discussed in sections 1.1 and 2.1.4. However, the high dropoff rate did cause issues for the post-survey response rate. This issue is further discussed in Section 5.4.

Engagement

During the workshops it became apparent that the subjects of cyber security and hacking peaked the interest of students. Many students were highly engaged during the presentation, and a high level of interactivity was maintained throughout all sessions. One teacher responsible for one of the groups even called student engagement during their session "remarkable". It was also noted that a possible reason for the increased engagement was that the workshop provided "content outside of the students' regular curriculum".

The main challenge was to translate student interest and engagement during the sessions to their participation in the system. As discussed in the preceding section, the majority participants came from the workshop sessions which indicates that the workshops were at least to some extent succesful in inspiring learners to join. However, many learners seemed to join out of curiosity which may have caused the aforementioned high dropoff rate. This effect was hard to combat, as it is a common phenomenon with MOOCs and a high level of engagement and perseverance was required to complete many or all questions within the system.

To increase and sustain the overall level of participation and engagement in the system, two measures were in place. First, there was the possibility to earn a signed Hacklab MOOC prototype certificate. To be eligible for a certificate, the user had to achieve a score of at least 40% correct, with at least one correct answer in all three modules. If a learner achieved this score they were presented with instructions on how to request a certificate upon completing the system. A total of 13 learners requested and received a Hacklab MOOC certificate.

Second, a prize was awarded to the best-scoring learner of every educational institution, with five prizes handed out in total. As the possibility of winning prizes was communicated prior to learners participating in the system, it provided an extra incentive to perform well for all participants, especially those who were already performing well. The prize entailed a Deloitte goody bag with a hand-signed winner certificate and several small prizes such as a lockpicking kit and a build-your-own-robot kit. The prizes were presented to the winners in several award ceremonies.

5.4 Issues and Limitations

Obviously the developed system was not without limitations, and several issues occurred during its delivery. In this section, these issues, their impact on the case study, and their remediation are discussed. Limitations with this study that are not specific to the implementation and execution of the case study are discussed in Section 7.3.

One key limitation of the Concerto platform was that it was only able to accept one correct answer per question. Since an open input field was shown to the user, this meant that the user had to input *exactly* the right answer. For some questions, for example those who required an URL or copy-pasted flag, this caused issues where the URL was slightly different or blank space was copy-pasted with the answer, respectively. In turn, this caused correct answers to be marked wrong by the adaptive learning system. This issue became apparent in the user testing phase, as discussed in Section 4.4.2.

To minimise the effects of this issue, an emphasised 'answer format' field was implemented for every question. This field showed the expected, correct answer format for every question next to the answer input field. Users were instructed to always double check this answer format before submitting, which they were also prompted to do by a dialog box which appeared upon every submission. Despite these measures, many users still input correct answers that were marked as wrong by the system. Even though a user's score and estimated ability levels could be re-calculated, this did cause an issue for learners in the adaptive condition where they were routed to an easier question where they should have received a harder question. Luckily, most users who experienced this issue were in the linear and random conditions, minimising the impact of this issue on the effectiveness of the system.

The technical implementation of Concerto platform also caused two issues during the execution of the case study. First, the system crashed once during the first workshop, when most participants tried connecting at the same time. This crash was quickly fixed by rebooting the virtual machine, no lasting damage was caused and all data was preserved. Further crashes were avoided by upgrading the virtual machine to allow for more traffic in later stages. Concerto also caused an issue where some user sessions were corrupted in some way, causing them to be unable to log in again after leaving the system. This issue was unsuccessfully diagnosed by the author and reported to the developers of Concerto. Participants who encountered this issue were instructed to report it to the author and create a new account. These entries were then manually fixed in the database by merging their old and new scores.

One limitation caused by human factors was that only a small amount of learners pressed the 'drop out' button when they were finished to quit the system and proceed to the post-survey as instructed. This issue caused the response rate to the post-survey to be lower than expected. Several measures were taken to reduce this effect, including placing special emphasis on the dropout feature in the workshops and email reminders. When this issue was discussed during one of the award ceremonies, it became apparent that many learners had apparently forgotten about the system until after the participation deadline.

6 Results

In this section, the results of the adaptive learning implementation and case study will be discussed. First, user demographics and system validity will be established. Afterwards, user satisfaction, engagement, completion, and dropout will be discussed. These metrics will be analysed to accept or reject the stated hypotheses and answer most (sub-)research questions.

6.1 Response and Demographics

6.1.1 Response

As described in Section 5.3, a total of 156 users registered an account in the adaptive learning system. These users were introduced to the system through a workshop ($n = 134$) or email invite ($n = 22$). Out of these registered users, 131 answered at least one question and are considered 'active users' (this distinction was made for further analyses). The division of users over conditions is shown in Table 6.1. As can be seen in this table, the amount of active users is quite balanced between the three conditions, with slightly less than a third of participants in the random condition.

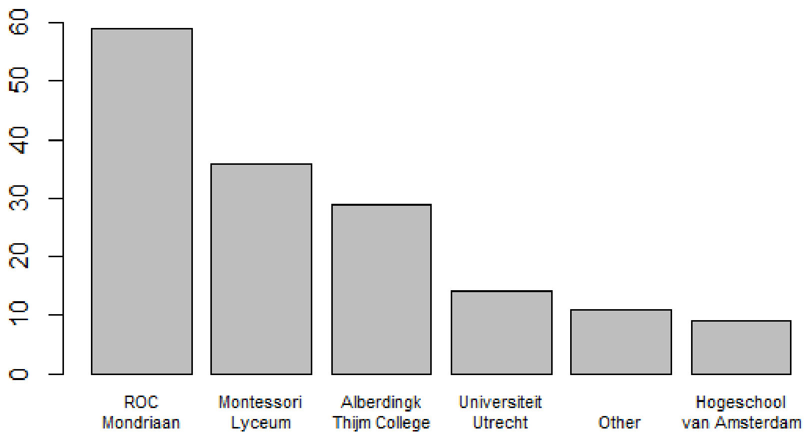
n	**Random**	**Linear**	**Adaptive**	**NA**
Registrations	53	47	55	1
Active Users	41	46	44	-

Table 6.1: Division of Users Over Conditions

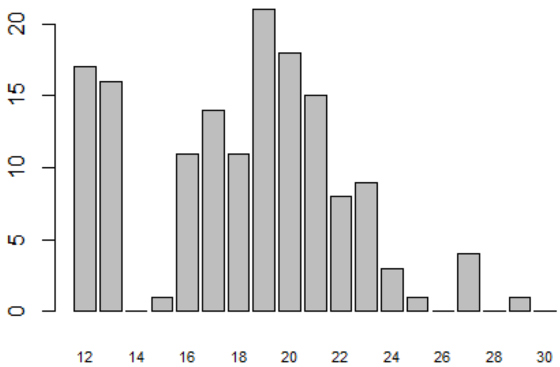
6.1.2 Demographics

The system's pre-survey was designed to gain insight into the demographics of participants (refer to Section 4.2.3). Of the 156 participants that filled in the pre-survey, 141 (90.4%) were male and 11 (7.1%) were female. The remaining 4 (2.6%) participants indicated being a different gender by tampering with the provided input field. A majority of 93 (59.6%) participants indicated upfront that they were most interested in the 'Web Hacking' module. The 'Network Hacking and Forensics' and 'Encryption' modules followed with 35 (22.4%) and 28 (17.9%) votes, respectively.

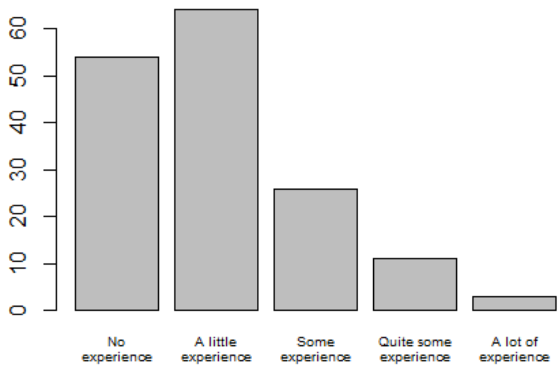
Most participating users were students at one of the five participating educational institutions. However, the amount of participants was not equally divided over these institutions. The amount of participants from each institution is shown in Figure 6.1a. The 'Other' category was comprised of colleagues and friends of the author who agreed to test



(a)



(b)



(c)

Figure 6.1: Participant Distributions for (a) Institution, (b) Age, and (c) Perceived Experience with Cyber Security

the system in the random condition. As noted before, the majority of the participants originated from an institution where workshops were given by the author.

Due to the varying backgrounds of participants, a broad demography was expected, as well as desired, for the execution of the case study. The demographic factors that were expected to influence participant performance within the system the most were 'age' and 'experience with cyber security'. The distributions for these factors are shown in figures 6.1b and 6.1c. As becomes apparent from these figures, most participants are quite young (high school to early higher education age) and inexperienced with cyber security (as per their own judgement). This demographic distribution fits quite well with the target group for the 'real' Hacklab MOOC.

6.1.3 Post-survey Validity

In order to generalise the results of the post-survey, it needed to be established that the sample of users who completed it was representative of the entire learner population, and no bias existed. This is important since, by design, the post-survey could only be completed by people who either finished all the challenges (27 out of 156 participants) or explicitly pressed the 'drop out' button (10 out of 156). Users who stopped the test in any other way are therefore not included in post-survey results.

To verify whether or not the subset of post-survey completers ($n = 37$) was representative of the entire population of learners ($n = 156$), the demographics of both groups were compared. A summary of the demographics for both groups is shown in Table 6.2.

	All Users (n=156)	**Post-Survey Completers (n=37)**
Age	M=18.32	M=18.86
Gender	91% Male 7% Female	94% Male 6% Female
Institution	37% ROC-M 18% ATC 45% Other	47% ROC-M 6% ATC 47% Other
English Skill Level	25% Basic 34% Intermediate 41% Advanced	32% Basic 29% Intermediate 38% Advanced
Cyber Security Experience	75% Inexperienced 25% Other	78% Inexperienced 22% Other
Most Interesting Module	60% Web hacking 22% Net hacking 18% Encryption	64% Web hacking 19% Net hacking 17% Encryption

Table 6.2: Demographics of All Completers Compared to Post-Survey Completers

As becomes apparent from this comparison, both groups are nearly equal in every demographic aspect apart from the institution; students from Alberdingk Thijm College are relatively underrepresented. This can be explained by the fact that they participated in the first phase of the experiment (the random condition), in which there was less pressure on (and reminders for) completing the post-survey. Furthermore, no one from the "other" category (not shown in the table) filled in the post-survey, as they were only invited as trial participants.

From this information it was concluded that the results from the post-survey can be generalised, at least for learners within the linear and adaptive conditions. It is, however, important to keep in mind that the post-survey does not represent users that did not complete the system without dropping out explicitly.

6.2 Satisfaction and Engagement

As described in Chapter 3, user satisfaction and user engagement are hypothesised to play a key role in a user's decision on whether or not to complete a MOOC. The way of measuring these variables was established in Section 4.2.1: Both satisfaction and engagement were self-reported by the user through the post-survey. The statements in the post-survey used to measure satisfaction and engagement are shown in Table A.2. In addition, engagement was also measured by means of a construct, elaborated below.

6.2.1 Engagement Construct

To operationalise a convenient way of measuring engagement based on various observed variables, the decision was made to create a construct for engagement. Variables that were hypothesised to indicate user engagement were time taken (in minutes), completion (in percent), questions skipped (in percent), and help used (as boolean). As a first attempt to create a construct for engagement, a simple linear regression was calculated to predict user's self-reported engagement based on these variables. A weakly significant regression equation was found ($F(4,32) = 2.784, p = 0.04$), with an R^2 of 0.258. However, the intercept and amount of skipped questions were the only significant coefficients in this regression equation which rendered it unusable in this context.

Hence, the choice was made to create an engagement construct based on expert judgement instead. Due to it's weak validity, this construct was meant to provide an indication of a user's interaction with the system during the exploratory analysis phase, it was explicitly *not* implemented to measure the independent variable engagement. The construct is comprised of the elements listed above. The formula for the construct is shown in Equation 6.1.

$$EngagementC = \frac{10 + C - S + (H * 10) - \frac{30 - min(T,30)}{3}}{1.2} \tag{6.1}$$

Where C is completion in percent, S is the amount of skipped questions in percent, H indicates whether or not the user activated the help feature (1 if yes, 0 if no), and T denotes the amount of minutes they spent in the system. $EngagementC$ was the internal name given to the engagement construct, in contrast to $Engagement$ which denoted the user's self-reported engagement score.

In short, this formula takes a user's completion as a base (ranging from 0-100), adds 10 bonus points if the user activated the help feature, deducts up to 10 points if the user spent less than 30 minutes in the system, and finally normalises the score to a 0-100 scale. This was considered a logical combination of factors based on the observed variables and frequency distributions. Overall, this provided a good indication of the distribution of user engagement. However, only the user's self-reported engagement measure was used for further analysis to ensure validity.

6.2.2 The Relation Between Satisfaction and Engagement

During the exploration phase of this study, the suspicion arose that a relation existed between satisfaction and engagement. Further analysing this relationship is interesting, as it helps with interpreting and understanding the data set, and could also impact other findings regarding satisfaction and engagement. Therefore, the decision was made to further explore this relationship.

To assess the relation between a learner's satisfaction and engagement, a Pearson's product-moment correlation coefficient was computed for these two variables, using the self-reported scores for user engagement. A strong positive correlation between satisfaction and engagement was found ($r = 0.84, n = 36, p = 0.00$), indicating that users with a high satisfaction score were also highly engaged. A scatterplot of this correlation is shown in Figure 6.2. From this figure it becomes apparent that test completers (shown in green) generally score higher on both satisfaction and engagement than dropouts (shown in red).

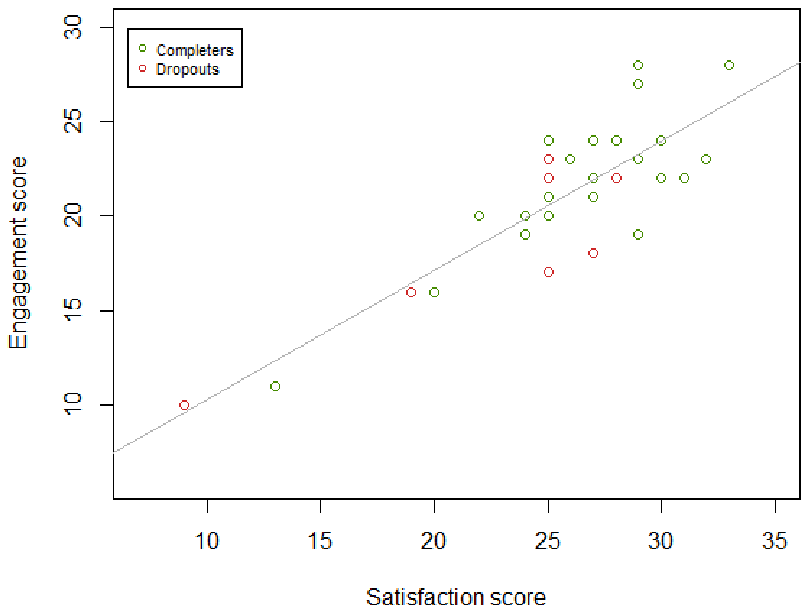


Figure 6.2: Scatterplot of Satisfaction and Engagement

To complement this finding, a linear regression was conducted to predict engagement based on satisfaction. This regression also yielded significant results ($F(1, 34) = 79.6, p = 0.00$). These results suggest that a high learner satisfaction is strongly related to increased learner engagement. Hence, it can be stated that a highly satisfied learner is also more engaged, since these learners perceive the system and its contents as positive and more fun to complete. The opposite is also true; highly engaged learners show higher levels of satisfaction with the system.

6.2.3 The Effect of Adaptive Learning

Hypotheses 2 and 3 of this paper concern the influence of adaptive learning on satisfaction and engagement. It was hypothesised that both satisfaction and engagement would be increased by implementing adaptive learning. In this section, the analyses of these hypotheses are discussed.

Hypothesis 2: Adaptive learning has a positive effect on engagement

Hypothesis 2 was based on the assumption that learner engagement would be increased by only providing the learner with material that is relevant to them at that specific point in time. To test whether or not this hypothesis is true, learner engagement scores were compared between the linear and adaptive conditions. As can be seen in Figure 6.3, learners in the linear condition scored themselves slightly better on engagement.

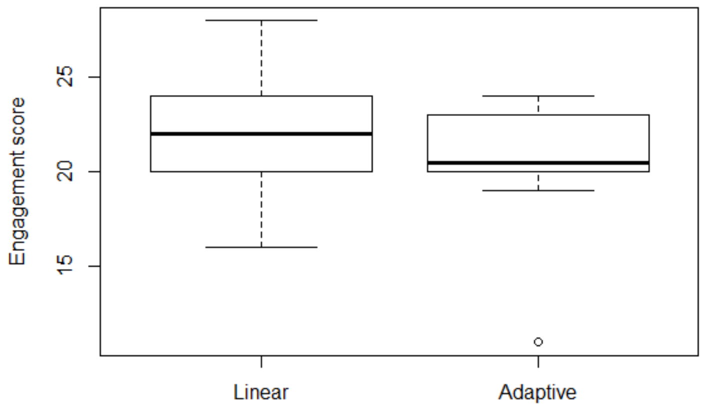


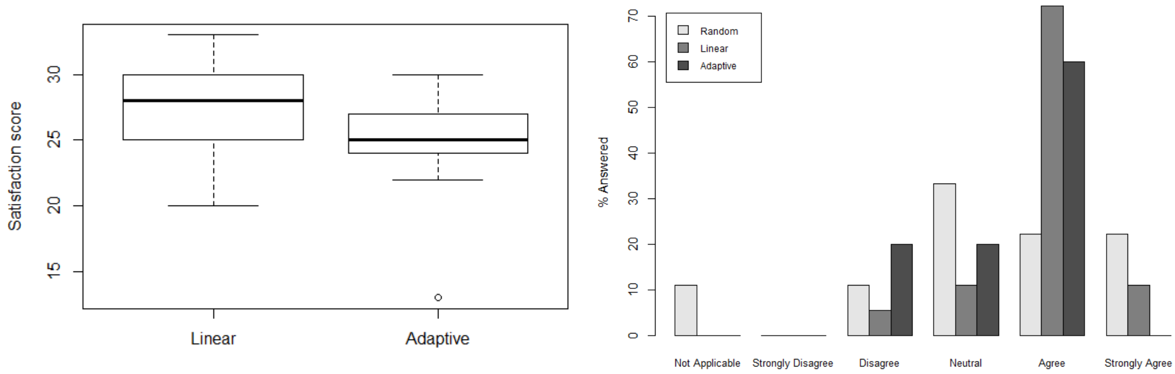
Figure 6.3: Boxplots of Engagement in Linear and Adaptive Conditions

Levene's test for homogeneity of variance was performed, but returned no significant results ($F(1, 26) = 0.04, p = 0.85$). Furthermore, a two-sample t-test was conducted, but no significant difference between the linear ($M = 78.22, SD = 10.02$) and adaptive ($M = 68.10, SD = 12.41$) conditions was found ($t(26) = 1.42, p = 0.17$). Therefore hypothesis 2, "Adaptive learning has a positive effect on engagement", was rejected and its null hypothesis ("Adaptive learning does not have a positive effect on engagement") was retained.

Hypothesis 3: Adaptive learning has a positive effect on satisfaction

This hypothesis was based on the assumption that adaptive learning would increase satisfaction by providing the right level of challenge and saving the user time. Again, the self-reported satisfaction values for participants in the linear and adaptive conditions were compared to test this hypothesis. This comparison is shown in Figure 6.4a.

A two-sample t-test was conducted for the linear ($M = 80.82, SD = 8.70$) and adaptive ($M = 70.20, SD = 11.31$) conditions. This test was found to be significant at the



(a) (b)

Figure 6.4: (a) Boxplots of Satisfaction in Linear and Adaptive Conditions, and (b) Survey results for Question 3: "The difficulty of the challenges was at the right level for me".

.05 confidence level ($t(26) = 2.64, p = 0.01$). However, this result directly refutes the hypothesis as learners in the linear condition score significantly higher than learners in the adaptive condition. Therefore, null hypothesis 3 ("Adaptive learning does not have a positive effect on satisfaction") was retained.

To find out the cause for this decrease in satisfaction, post-survey results from both conditions were analysed and compared per question. No notable differences were found between the conditions, with the exception of Question 3: "The difficulty of the challenges was at the right level for me". Learner answers for this question are shown in Figure 6.4b.

As can be seen, learners in the adaptive condition disagree with this question more than learners in the other conditions. A two-sample t-test was performed on learner response in the linear and adaptive conditions. It was found that the difference in response between the linear condition ($M = 3.76, SD = 0.83$) and the adaptive condition ($M = 3.00, SD = 0.82$) was significant at the .05 confidence level ($t(19) = 2.33, p = 0.03$).

It is not surprising that learners in the adaptive condition score lower in this question, as adaptive tests are often perceived as more difficult than traditional tests (for more refer to Section 4.2.2). However, it was not expected that this perceived difficulty has such a large impact on the overall satisfaction of learners.

It was also expected that users would be more satisfied because of the increased learning efficiency and the decrease in completion times. However, even though completion times were drastically decreased (refer to Section 6.3), further analysis of post-survey questions revealed no increase in satisfaction with the system's duration (both in total and per module).

6.3 Completion and Dropout

The main research question of this thesis concerns MOOC completion rates and the influence of adaptive learning on this metric. In this section, the analysis of research questions regarding completion and dropout will be discussed. This includes the analysis of hypotheses 1, 4 and 5.

There are various possible ways to analyse MOOC completion. The main term used in this thesis is 'completion rates'. However, to analyse MOOC completion in all possible ways, several sub-definitions were used during analysis. These are 'Completion', 'Dropout' and 'Dropoff'. These terms are defined in the table below and were each analysed separately.

Term	**Definition** ***(Measure)***
Completion	Whether or not a user answered all questions in the system, completing it entirely. *(Boolean)*
Dropout	Whether or not a user explicitly dropped out of the system through the 'Dropout' feature. *(Boolean)*
Dropoff	At the population level, dropoff denotes the amount of users that were active at a certain point in the system. Dropoff is similar to dropout, except that dropoff indicates users that stopped with the system non-explicitly. *(Percentage of active users)*

Table 6.3: Definition of Completion Types

6.3.1 The Effect of Adaptive Learning

Completion time

One variable that was expected to play a key role in the system's completion numbers was the time needed to complete the system. Logically, this time would be decreased in the adaptive condition since approximately 40% fewer questions are provided to the user compared to the linear condition. The actual decrease in average cumulative completion time per condition is shown in Figure 6.5.

As becomes apparent from this figure, the total cumulative completion time for each condition (*i.e.* where the lines intersect with the right side of the figure) is a lot lower for the adaptive condition, as expected. The difference in participation time (in minutes) was analysed by means of a two-sample t-test. As expected, a significant difference between the linear ($M = 54.39, SD = 57.80$) and adaptive ($M = 33.43, SD = 32.18$) conditions was found ($t(90) = 2.15, p = 0.03$), showing that the adaptive condition significantly reduced the required time to achieve the same learning results (as discussed in Section 4.4.6). It should be noted that this analysis includes users who did not complete the entire system, and participation times are therefore much lower than the completion times shown in the figure.

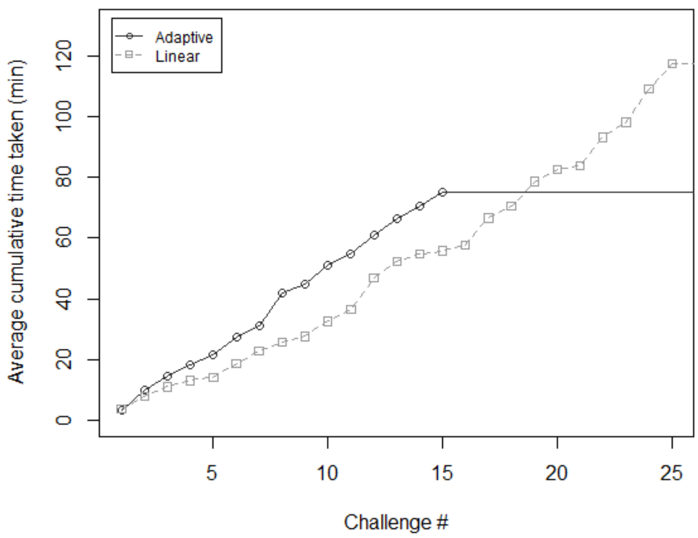


Figure 6.5: Average Cumulative Completion Time Per Challenge

Furthermore, the slope of the time needed to complete challenges in the adaptive condition (*i.e.* the average time needed per challenge) is slightly steeper, indicating that completing learners needed slightly more time to complete a challenge on average. This is to be expected, however, since users are served more difficult questions right from the start. The slope of the linear condition shows a fluctuating pattern, which coincides with the increasing difficulty levels per module.

To complement the above, the difference in linear ($M = 4.98, SD = 3.48$) and adaptive ($M = 4.73, SD = 3.33$) average question completion times was analysed, based on data from all questions (rather than only completing learners) averaged per question ID. No significant difference was found ($t(47) = -0.26, p = 0.79$). However, these numbers do indicate that learners in the adaptive condition answered questions quicker than those in the linear condition. This difference becomes even bigger when you look at the average time per question answered *correctly* (Linear $M = 4.66$, Adaptive $M = 5.27$). This is an interesting finding, because it suggests that learners in the adaptive condition were on average more efficient at answering questions, even though the questions they received were harder than those of learners in the linear condition.

Completion

Following the analysis of completion time, the difference in the amount of completing learners was analysed per condition. For this analysis, completion was converted to boolean (true/false) and learners with zero completion (*i.e.* users that did not actually answer any questions) were discarded from the data set.

To analyse the difference in completing and non-completing learners per condition, a Chi-square test with Yates' continuity correction was performed. This test revealed no significant difference in the proportion of completing learners per condition ($\chi^2(1, N = 92) = 0.51, p = 0.48$). Therefore, even though the proportion of completing learners is higher in the linear condition than in the adaptive condition (30% versus 22%, respectively), the odds of completing the system are not statistically higher in this condition.

Dropout

Following the analysis of system completion, data on learner dropout was analysed. As discussed, users are only considered dropouts if they *explicitly* made use of the aforementioned 'dropout' button, which skipped all remaining questions and redirected the learner to the post-survey. In total, this button was only used 14 times: 7 times in the random condition, 6 times in the linear condition and once in the adaptive condition. Figure 6.6 shows the division of dropped out per condition learners over challenge IDs.

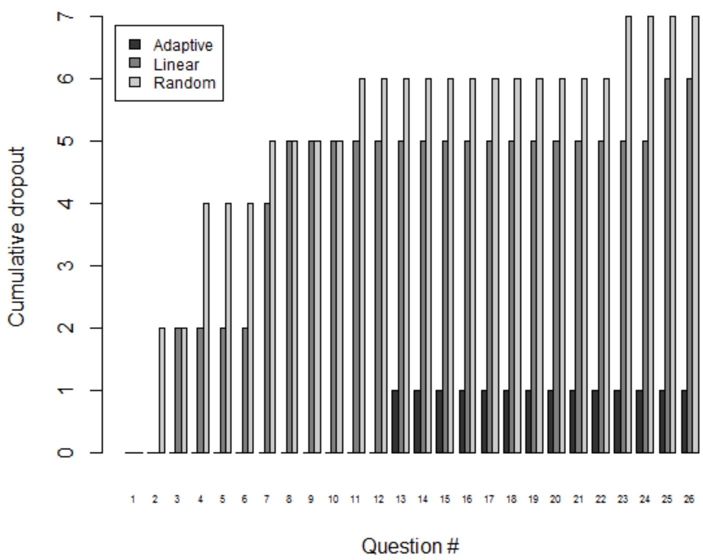


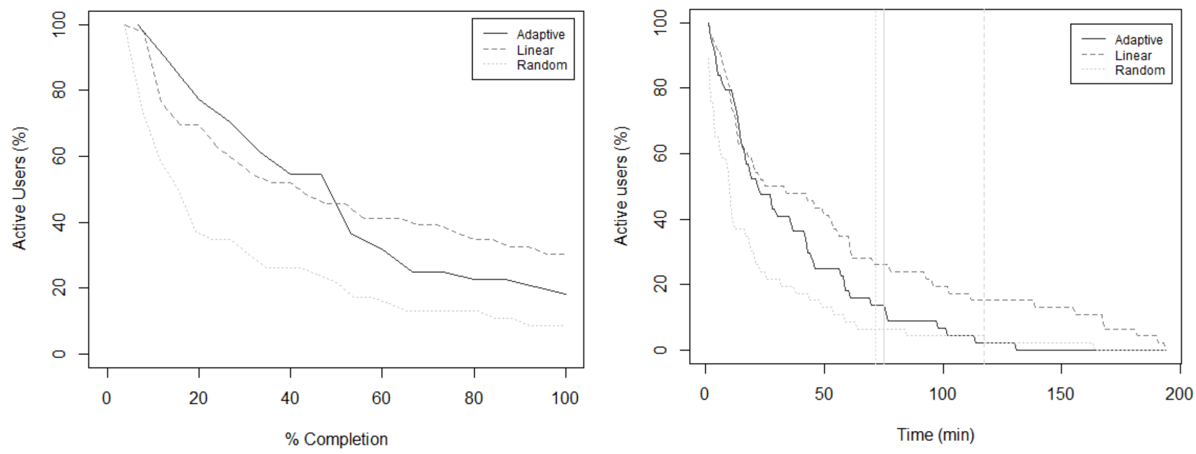
Figure 6.6: Dropout Per Condition

To test whether or not this difference in dropout usage was statistically significant, a two-sided Fisher's exact test was computed. This test was chosen over the chi-square test as the data involved small samples, violating the sample size assumption of the chi-square test. The Fisher's exact test returned a significant result at the .05 confidence level ($p = 0.046$), indicating that the odds of dropping out are significantly lower in the adaptive condition.

It is hard to discern the exact reason for this decrease in dropout in the adaptive condition, especially since adaptive completion rates are not higher than linear completion rates. Nevertheless, it can be concluded that adaptive learning has a significant positive effect on the reduction of dropout.

Dropoff

Looking at completion not as a boolean but as a continuous metric is interesting because it provides more information about the completion of learners, and can be measured over



(a) (b)

Figure 6.7: (a) Relative Dropoff Per Condition, and (b) Dropoff Per Condition Over Time. Vertical Lines Represent Mean Completion Times Per Condition.

time. (Non-)Completion over time is called 'Dropoff' in this thesis to discern it from the binary completion metric used before. Comparing dropoff between conditions is useful as it provides an indication of how many learners stop answering questions after a certain point in time. This can be interpreted as a non-explicit form of dropout.

A comparison of dropoff per condition is shown in Figure 6.7a. It is important to note that the x-axis of this figure denotes *relative* completion, *i.e.* that it encompasses more challenges for the linear condition than for the adaptive condition.

As can be seen, the dropoff curve for the adaptive condition is less steep than that of the linear condition, indicating a higher amount of completing learners. However, many learners in the adaptive condition drop off at approximately the 50% mark. This point translates back to the 7th challenge, the last challenge of the first module of the adaptive condition. Many learners in the adaptive condition apparently stopped answering questions after the first module, an effect which is not as visible in the linear condition (after challenge 12, the approximate 45% mark). It is uncertain whether this discrepancy was caused by the adaptive condition or by the experimental setup.

When looking at dropoff over time (Figure 6.7b), the dropoff curve for the adaptive condition becomes steeper than that of the linear condition. This is to be expected, since learners in the adaptive condition have to solve harder challenges from the start. Furthermore, learners in the adaptive condition reach the end sooner, which logically contributes to a steeper dropoff curve.

Both figures show that the dropoff curve of the random condition is a lot steeper, indicating a higher level of dropoff. This is presumably caused by the fact that the randomised condition consisted of a smaller participant group who received less continuation reminders.

Overall, it can be concluded that adaptive learning steepens the dropoff curve for learners, which suggests that learners require more motivation to continue. However, even though the dropoff curve is steeper for the adaptive condition due to the increased challenge difficulty and pace, this does not cause a significant decrease in overall completion, as discussed in Section 6.3.1.

Hypothesis 1: Adaptive learning has a positive effect on completion rates

Based on the analyses performed in the sections above, it can be concluded that adaptive learning does not have an explicit, positive effect on achieved completion rates in the context of the performed experiment. Therefore, null hypothesis 1 ("Adaptive learning does not have a positive effect on completion rates") was retained. However, some interesting observations were made. These include that adaptive learning significantly decreased both completion time and dropout, despite an increased slope of learner dropoff. Furthermore, a stronger effect may be introduced by different adaptive learning implementations in another context.

6.3.2 The Effect of Satisfaction and Engagement

It was hypothesised that a high learner satisfaction as well as a high learner engagement would contribute to higher completion rates. To measure the actual effect of satisfaction and engagement on completion, two groups were distinguished within the participants that completed the post-survey: test completers and dropouts. Unfortunately, these groups were rather small due to the limited response to the post-survey. The test completer group consisted of 28 participants, where the dropout group consisted of 14 participants.

Hypothesis 4: A high learner satisfaction has a positive effect on completion rates

First, the relation between satisfaction and completion was assessed by comparing the satisfaction reported by completing users ($M = 26.62, SD = 4.13$) to that of the dropouts ($M = 23.9, SD = 5.88$), shown in Figure 6.8a. A two-sample t-test was performed, but it returned no significant result ($t(34) = 1.57, p = 0.13$). In addition, no significant correlation ($r(34) = 0.32, p = 0.054$) or regression ($p = 0.054$) was found between these values. The data suggest that completing users are somewhat more satisfied with the system, but unfortunately it cannot be determined that satisfaction has a positive effect on completion as it is also possible that completing users are more satisfied with the system as a result of completion. Therefore, null hypothesis 4 ("A high learner satisfaction does not have a positive effect on completion rates") was retained.

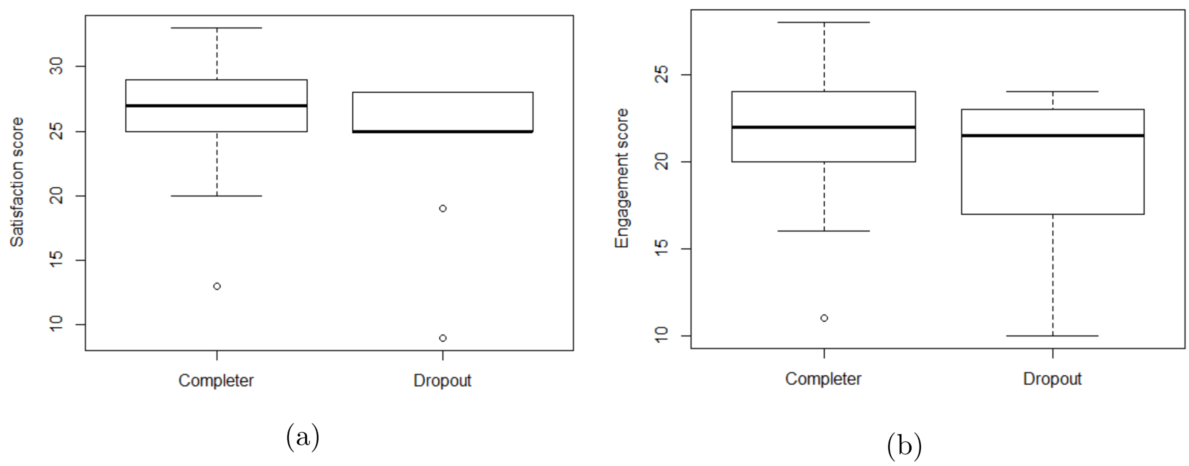


Figure 6.8: (a) Boxplots of Completer and Dropout Satisfaction, and (b) Boxplots of Completer and Dropout Engagement.

Hypothesis 5: A high learner engagement has a positive effect on completion rates

To assess the effect of engagement on completion rates, the same distinction was made between participants that completed the post-survey. As can be seen in Figure 6.8b, completing users ($M = 21.69, SD = 3.53$) scored somewhat higher on engagement than dropouts ($M = 19.7, SD = 4.45$). However, a two-sample t-test revealed that this difference was not statistically significant ($t(34) = 1.41, p = 0.17$). As with engagement, no significant correlation ($r(34) = 0.24, p = 0.16$) or regression ($p = 0.16$) was found between engagement and completion. Therefore, null hypothesis 5 ("A high learner engagement does not have a positive effect on completion rates") was retained.

7 Discussion

In this thesis, an attempt was made to increase MOOC completion rates through the implementation of adaptive learning in the context of MOOCs. Adaptive learning was successfully implemented in a prototype MOOC about cyber security, the "Hacklab MOOC". The development and implementation of this MOOC was succesful, showing that it is possible to create MOOCs with adaptive learning at their core.

It was found that the implementation of adaptive learning significantly reduced MOOC completion time and user dropout whilst retaining measuring accuracy. Furthermore, learners in the adaptive condition were more effective at answering questions. Despite these results, the realised implementation of adaptive learning did not successfully increase MOOC completion rates. It was found that the dropoff curve was slightly steeper in the adaptive learning system, and that overall completion rates were comparable to those in the control system.

To further analyse the effect of adaptive learning, the variables learner satisfaction and engagement were also investigated. These variables were found to be highly correlated, as was expected. No effect of either variable on completion rates was found, and engagement was not affected by the implementation of adaptive learning. However, satisfaction was significantly lower for learners in the adaptive condition, which was most likely caused by the non-traditional form of testing and increased difficulty, as described by Linacre [51].

The findings of this study per hypothesis are summarised in Table 7.1. Figure 7.1 shows the findings mapped onto the conceptual model of this research provided earlier. In addition to the listed hypotheses, several research questions were formulated in Chapter 3 to guide the execution of this research project. First, the formulated sub-questions will be answered based on the findings described above. Following that, the answer to the main research question will be provided in the next section.

SQ1. What is engagement?

As discussed in Section 2.1.4, most studies look at motivations behind and patterns of engagement rather than its role in the bigger picture of learning like this research project does. To clearly define engagement within the boundaries of this research project, a construct for engagement within the operational context of this study was defined in Section 6.2.1. This construct was based on participation, time, and usage of the help function. However, no concrete evidence was found that this construct accurately represented learner engagement. Therefore, the resulting information was only used for exploratory data analysis. Instead, a user's self-reported engagement was used as a metric. This metric was based on six questions, listed in Table A.2.

In sum, there are various approaches to measuring a user's engagement. In the broad

	Hypothesis	Accepted?	Remarks
H1	Adaptive learning has a positive effect on completion rates	Not Accepted	No significant effect on completion rates was found. Dropout was significantly reduced in the adaptive condition, but dropoff and completion were not.
H2	Adaptive learning has a positive effect on learner engagement	Not Accepted	No significant effect on engagement was found.
H3	Adaptive learning has a positive effect on learner satisfaction	Rejected	A significant *negative* effect on satisfaction was found. This was found to be largely due to the increase in difficulty.
H4	A high learner satisfaction has a positive effect on completion rates	Not Accepted	No significant effect on completion rates was found. Completing learners were slightly more satisfied.
H5	A high learner engagement has a positive effect on completion rates	Not Accepted	No significant effect on completion rates was found. Completing learners were slightly more satisfied.

Table 7.1: Hypothesis Outcomes

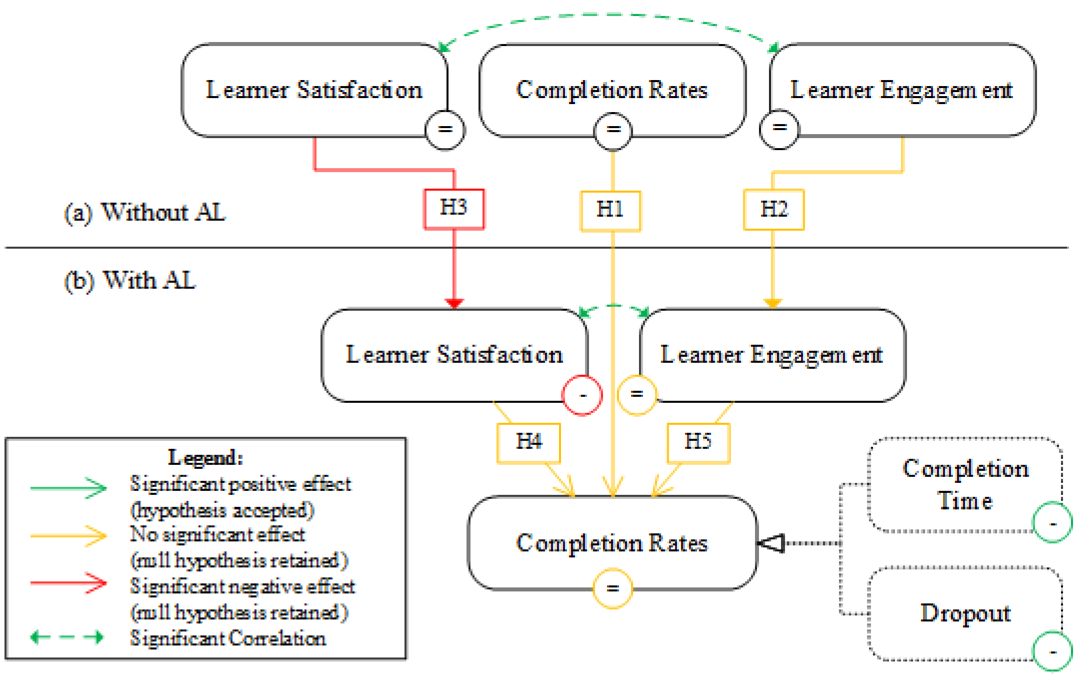


Figure 7.1: Conceptual Model With Research Results

sense it encompasses the type and level of interaction shown by the user with regard to the (learning) material provided by a system. The best way of measuring engagement may differ depending on the context. This study worked with both a quantified as well as a qualitative approach.

SQ2. What is an effective method for the implementation of adaptive learning?

The method for designing, developing, and implementing the adaptive learning system was described in Chapter 4 and the design specification supplied with this thesis[1]. Due to the limited duration of this research project, a minimal design approach was applied so that most of the effort could be spent on the implementation of the system itself. The design approach was based on the principles of design science [47], to ensure that both the developed prototype as well as the experimental environment were developed in a (scientifically) sound way.

The design approach proved effective for this project. Though it was mostly linear in nature, enough flexibility was kept to switch between phases. This allowed for the smooth implementation and testing of for example the subsequent experimental phases. The elaborate validation step of the design process helped check the system's validity and avoid errors during the system's deployment.

Using the Concerto Platform greatly helped with implementation of adaptive learning within the system. Since Concerto has built-in support for the implementation of adaptive selection and skill assessment functions, most functionality was available 'out of the box'. The platform was also highly modifiable and therefore allowed for implementation of additional functionality with relatively little development effort.

SQ3. What is an effective implementation of adaptive learning for the domain of cyber security?

The case study and practical execution of the designed cyber security MOOC prototype was described in Chapter 5. In the context of this case study, it was shown that the described implementation of adaptive learning was effective. Though the implemented system worked with challenges rather than theoretical questions, accurate difficulty estimations were made and the real-time challenge selection worked smoothly.

One major downside to this implementation of adaptive learning was the distinct lack of challenges within the system. Due to time and resource constraints the amount of challenges was limited to 26. Because these challenges were also divided over three modules, the adaptive learning system was greatly constrained in the selection of new challenges. It goes without saying that adaptive learning functions better if more challenges are

[1] If, for any reason, the design specification was not provided with your copy of this thesis, it can be downloaded separately here: `https://bit.ly/MScDesignDoc`

available for selection. Therefore, care needs to be taken that enough challenges (or questions) of varying difficulty levels are available if one were to implement this system in a production environment.

SQ4. To what extent does this adaptive learning implementation increase completion rates?

As shown in Section 6.3 in the Results chapter, the implemented adaptive learning system did not increase completion rates. However, learner dropout was significantly reduced through the implementation of adaptive learning. This indicates that learners in the non-adaptive conditions made the choice to explicitly stop with the test more often than learners in the adaptive condition. However, this did not affect dropoff (users stopping with the test non-explicitly) or completion rates at large.

SQ5. To what extent is the increase in completion rates affected by an increase in learner engagement?

The effect of learner engagement on completion rates was discussed in Section 6.3.2 of the Results chapter. Unfortunately the data supporting this question was limited due to the relatively low response on the system's post-survey. Even though completing users had a slightly higher engagement overall, this effect was shown not to be statistically significant. Furthermore, a causal relationship between engagement and completion was not fully established. Therefore, it is hard to provide a conclusive answer to this sub-question.

SQ6. To what extent is the increase in completion rates affected by an increase in learner satisfaction?

The effect of learner satisfaction on completion rates was also discussed in Section 6.3.2 of the Results chapter. As with engagement, it was hard to establish the effect of satisfaction on completion due to the low amount of observations. Again, satisfaction values were slightly higher for completing users, but this difference was found not to be statistically significant. Therefore, this questions can not be answered with a high level of certainty based on the collected data.

It is interesting to note however that even though satisfaction and completion did not seem to be strongly correlated, satisfaction was significantly lower for users in the adaptive condition. This can be explained by the fact that learners are not used to the adaptive way of testing, which immediately confronts them with harder questions in comparison to classical tests. This was confirmed by the post-survey results, which showed that learners in the adaptive condition were significantly less content about the difficulty level of the system.

7.1 Conclusion

In this research project it was found that it is possible and viable to implement adaptive learning within MOOCs. Doing so has substantial benefits, but it also has an effect on the form, delivery and learner perception of this MOOC.

The main benefit provided by adaptive learning is a higher learning efficiency and, by extension, lower completion times. It has been established in this case study that these benefits also apply when implementing adaptive learning in a MOOC: learners in the adaptive condition completed the MOOC faster and with the same measuring accuracy, despite a 40% decrease in the amount of challenges provided.

The benefits of adaptive learning do come at a cost, however. In this case study, learners were significantly less satisfied with the adaptive system as questions were harder and the pace was higher.

The main problem of MOOCs that this research project attempted to solve with adaptive learning was completion rates. The research question for this study was formulated as follows:

> *"How can a MOOC's completion rates be increased, directly, or indirectly by increasing learner satisfaction or learner engagement, through the implementation of an adaptive learning system?"*

Based on the performed case study and the answered sub-questions listed above, this question can now be answered. Unfortunately, the implementation of adaptive learning within the context of this case study did not increase completion rates, neither directly nor through the increase of satisfaction or engagement. In addition, no relationship between satisfaction or engagement and completion was found. However, dropout was significantly decreased for learners in the adaptive condition. This proves that adaptive learning does influence the way learners think about, participate in, and interact with massive open online courses.

Whether or not these effects are desirable in the context of a specific MOOC is hard to infer based on the conducted case study. MOOC designers and developers should carefully consider whether or not the implementation of adaptive learning in their MOOC supports their mission, recommendations to support this are provided in Section 7.2.

7.2 Recommendations

As described in this and previous chapters, implementing adaptive learning in MOOCs is a viable option which may yield great benefits for MOOCs in practice. However, an increased and more complex implementation effort is required, and as shown in this study, adaptive learning may also have negative effects on (the perception of) a MOOC. To support decision making, concrete recommendations on adaptive learning in MOOCs will be made in this section.

There are various advantages to implementing adaptive learning in a MOOC. Adaptive learning increases the system's measurement efficiency for skill levels by choosing questions that provide as much information about the user as possible. Furthermore, the system increases learning efficiency by always choosing questions that are at the right difficulty level for the learner. In doing so, the system is able to skip approximately 40% of questions while maintaining the same measuring and learning accuracy. This in turn leads to a decrease in completion times for the end user. Finally, this study has shown that the introduction of adaptive learning significantly reduced learner dropout. This is a greatly beneficial effect, and it may have a huge impact on the success and potentially the profitability of large-scale MOOCs.

Contrarily, this study has also revealed some potential reasons not to implement adaptive learning in a MOOC. One important reason is the increased development effort and costs. For an adaptive learning system to work with high accuracy it needs to be calibrated, which requires at least a basic trial implementation of the system. Furthermore, extra effort should be put into designing the MOOC so that it is optimised for the adaptive serving of questions, and enough questions or challenges which cover a broad difficulty spectrum should be available. Another important reason to decide against the implementation of adaptive learning within a MOOC is user perception. As shown in this thesis, adaptive learning is perceived as more difficult when compared to classical (linear) ordering of questions. Furthermore, the order of questions is different for each user so it is hard to order questions in a logical manner.

In sum, every MOOC is unique and the decision whether or not to implement adaptive learning in a MOOC should be made explicitly and consciously, based on its audience, mission, and other factors. Following the results of this study and the advantages and disadvantages of adaptive learning discussed in this section, the following recommendations are made:

Do implement adaptive learning in your MOOC if:

- Questions are used for testing or measuring skill
- Learning time should be minimised
- Dropout needs to be minimised

Do not implement adaptive learning in your MOOC if:

- Questions are used to guide learning or the order of questions is important
- Questions are similar in difficulty
- Learner perception is important
- Resources are limited

7.3 Limitations

This research project was conducted as part of a Master's Thesis, and was therefore bound to strict time and resource constraints. Because of this, some imperfections and limitations had to be accepted within the project. This section discusses these limitations and their implications on achieved results. Note: Technical limitations are discussed in Section 5.4 and will not be discussed further in this section.

One of the main limitations encountered during the execution of this research project was the low post-survey response rate and limited use of the dropout feature: Only 42 participants, of which 14 dropouts, filled in the post-survey. Since two research variables (satisfaction and engagement) were measured through this survey, the low rate of response on the post-survey lowered the quality of results that include these variables. Furthermore, most users simply stopped participating in the system rather than using the dropout feature, which made it hard to identify motivations for dropout. This issue has been partly resolved in the analysis phase by discerning dropoff from (explicit) dropout.

Another major limitation concerned the implementation of the adaptive learning system itself. Initially, the predicted success rate of the adaptive learning algorithm was set to a higher value to increase engagement. Following the observations listed in Section 4.2.2, the predicted success rate determined by the algorithm could potentially be increased to a maximum of 70% ($p = 0.7$). To override the logic for next item selection, the NextItem function from the CatR package needed to be overwritten. This was achieved by redefining said function in the Concerto source code, and altering the logic for next item selection in such a way that the predicted success rate (p) becomes 0.7 rather than 0.5. However, when the system was tested with this value of p, it was not possible to reach all questions due to the positive bias in item selection and the limited number of questions. Therefore, this change was rolled back.

If this measure was implemented successfully, the negative impact of adaptive learning on satisfaction (which was established to be caused by an increase in difficulty) could have most likely been negated with only a small impact on the amount of questions or measuring accuracy in the adaptive condition. Furthermore, this measure could have potentially improved engagement following the observations of Jansen *et al.* [69], who found that children with a higher success rate practised more using their adaptive system, suggesting a higher engagement with learning materials.

Another limitation regarded the engagement variable. It turned out to be very hard to operationalise the engagement variable in such a way that it was measurable from the captured data. An attempt was made to model this variable, but this model was not strong enough in practice. This issue was resolved by using an expert construct instead, but due to the lack of validation this construct provided little value in practice. This limitation lowers the usability of the results regarding engagement.

7.3.1 Threats to Validity

In addition to limitations of the study itself, it is important to address the possible threats to the scientific validity of this study that were introduced by the experiment and applied methods. In this section, possible threats to validity will be discussed and classified in accordance with the types of validity threats as defined by Wohlin *et al.* [70]. Threats to validity are also prioritised based on the importance Wohlin *et al.* [70] define in the context of applied research. Threats to internal validity are considered the most important, followed by threats to external validity, threats to construct validity, and threats to conclusion validity. In this section, the threats are listed in order of decreasing importance (most important threats first).

Interaction of selection and treatment (external validity). It is possible that the population of learners selected for this study is not completely generalisable to the targeted population of learners at large. Due to the selection of institutions based on existing partnerships (convenience sampling), a bias in age or experience may have been introduced which affected the results of the study. Another factor which may have influenced the generalisability of the studied population is the provided incentive for participating (*i.e.* the prizes for the best participants). This incentive could have shifted learner motivation from intrinsic (the desire to learn) to extrinsic (a desire to win the prize). Hence affecting the results of the learners. Though not immediately critical, this effect should be kept in mind when drawing conclusions based on this study.

Inadequate preoperational explication of constructs (construct validity). Due to the lack of clear definitions in existing literature, the engagement variable was not clearly defined when the research project was initially planned. It was clear that this variable played an important role in the process of learning, but it was unclear how this variable was to be measured during the experiment which can be classified as "inadequate preoperational explication of constructs". As a result of this, an expert construct for engagement was created during the analysis phase which was unfortunately not valid enough to be used for analysis. Following this, the post-survey observations of engagement had to be relied on for analysis, which greatly reduced the statistical power of the variable.

Reliability of measures (conclusion validity). Since the developed engagement construct was not used for analysis, both the variables satisfaction and engagement were self-reported by participants based on a single post-survey. These variables were therefore subjective, which is considered less reliable than objective measures. Furthermore, the questions used in this post-survey were only partly based on literature, and therefore for the most part unvalidated. Hence, the reliability of the satisfaction and engagement measures depends on the validity of the formulated post-survey questions (listed in Table A.2 in Appendix A). Though there is no immediate reason to assume that these questions are invalid, this is an important consideration when reviewing the results regarding satisfaction and engagement.

Reliability of treatment implementation (conclusion validity). Lastly, it is important to acknowledge that the implementation of the treatment (*i.e.* adaptive learning) condition can not be considered optimal in this study. As mentioned several times, the system was limited by the number of available challenges. Furthermore, the implementation of the three modules further restricted the adaptivity of the system by constraining the item selection process. Had more questions been available or no modules been implemented, the adaptive learning system would have performed better, most likely increasing effect sizes and possibly affecting the results of the study.

7.4 Further Research

The implementation of adaptive learning in MOOCs is a relatively unexplored concept. Therefore, many relevant and interesting opportunities for further research exist. In this section some of the research opportunities that are most relevant in the context of this thesis will be discussed.

First of all, the influence of adaptive learning on learning results has not been investigated in the context of this project due to the limited scope. It goes without saying that learning results are one of the key metrics in massive open online courses. Therefore, it is very important to establish that learning results are not or marginally affected by the implementation of adaptive learning. This would be an interesting research question for a future study.

It would also be interesting to extend the research in this paper beyond a single case study. This could be done by conducting case studies in other disciplines or by analysing data from existing MOOCs. Doing so would help validate and generalise the findings listed in this thesis.

Furthermore, it would be interesting to further investigate specific variables used in this research project. For example, efforts could be made to create and validate an operational construct for engagement, or to further distinguish the various forms of dropout and their interrelationship.

It would also be interesting to further explore user rationale for dropping out, specifically in the context of adaptive learning. Some studies have explored learner rationale for participating in MOOCs in general, but these studies could be further specified to gain more insight into user behaviour. A practical research question would be why explicit dropout is decreased when adaptive learning is implemented.

In sum, many interesting opportunities for MOOC and adaptive learning research exist. Combining these fields of research could not only contribute on a scientific level, as results can be used in practice to further improve methods of (online) learning.

Bibliography

[1] A. McAuley, B. Stewart, G. Siemens and D. Cormier, 'The MOOC model for digital practice', *Massive Open Online Courses Digit. Ways Knowing Learn.*, pp. 1–64, 2010, ISSN: 0378-7206.

[2] G. Siemens, *MOOCs are really a platform*, 2012. [Online]. Available: `www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform` (visited on 07/12/2016).

[3] O. C. Rodriguez, 'MOOCs and the ai-stanford like courses: two successful and distinct course formats for massive open online courses', *Eur. J. Open, Distance E-Learning*, p. 13, 2012, ISSN: 1027-5207.

[4] D. Clow, 'MOOCs and the funnel of participation', in *Proc. Third Int. Conf. Learn. Anal. Knowl. (LAK '13)*, vol. 8, New York, NY: ACM Press, 2013, pp. 185–189, ISBN: 9781450317856. DOI: `10.1145/2460296.2460332`.

[5] J. Daniel, 'Making sense of MOOCs: musings in a maze of myth, paradox and possibility', *J. Interact. Media Educ.*, vol. 2012, no. 3, p. 18, 2012. DOI: `10.5334/2012-18`.

[6] S. Jaschik and D. Lederman, *Affirmative action, innovation and the financial future: a survey of presidents*, Washington, DC, 2013. [Online]. Available: `http://bit.ly/2f628GP`.

[7] A. Margaryan, M. Bianco and A. Littlejohn, 'Instructional quality of massive open online courses (MOOCs)', *Comput. Educ.*, vol. 80, pp. 77–83, 2015, ISSN: 03601315. DOI: `10.1016/j.compedu.2014.08.005`.

[8] K. Jordan, 'Initial trends in enrolment and completion of massive open online courses massive open online courses', *Int. Rev. Res. Open Distance Learn.*, vol. 15, no. 1, pp. 133–160, 2014, ISSN: 14923831. DOI: `10.19173/irrodl.v15i1.1651`.

[9] D. F. O. Onah, J. Sinclair and R. Boyatt, 'Dropout rates of massive open online courses: behavioural patterns', in *EDULEARN14 Proc.*, IATED Academy, 2014, pp. 5825–5834, ISBN: 9788461705573.

[10] R. F. Kizilcec, C. Piech and E. Schneider, 'Deconstructing disengagement : analyzing learner subpopulations in massive open online courses', in *Proc. Third Int. Conf. Learn. Anal. Knowl. (LAK '13)*, New York, NY: ACM Press, 2013, pp. 170–179, ISBN: 9781450317856. DOI: `10.1145/2460296.2460330`.

[11] K. F. Hew and W. S. Cheung, 'Students' and instructors' use of massive open online courses (MOOCs): motivations and challenges', *Educ. Res. Rev.*, vol. 12, pp. 45–58, 2014. DOI: `10.1016/j.edurev.2014.05.001`.

[12] H. Khalil and M. Ebner, 'MOOCs completion rates and possible methods to improve retention - a literature review', in *Proc. World Conf. Educ. Multimedia, Hypermedia Telecommun.*, Chesapeake, VA: AACE, 2014, pp. 1236–1244, ISBN: 978-1-939797-08-7.

[13] Á. Fidalgo-Blanco, M. L. Sein-Echaluce and F. J. García-Peñalvo, 'Methodological approach and technological framework to break the current limitations of MOOC model', *J. Univers. Comput. Sci.*, vol. 21, no. 5, pp. 712–734, 2015, ISSN: 09486968.

[14] L. Yuan, S. Powell and B. Olivier, *Beyond MOOCs: sustainable online learning in institutions*, 2014. [Online]. Available: `http://publications.cetis.org.uk/2014/898`.

[15] J. Daniel, E. V. Cano and M. G. Cervera, 'The future of MOOCs: adaptive learning or business model?', *RUSC. Univ. Knowl. Soc. J.*, vol. 12, no. 1, p. 64, 2015, ISSN: 1698-580X. DOI: `10.7238/rusc.v12i1.2475`.

[16] D. Clark, *Adaptive MOOCs: from traditional, flat MOOCs to adaptive MOOCs*, 2013. [Online]. Available: `https://www.cogbooks.com/white-papers/`.

[17] A. Savi, *How can we tailor massive education to each individual learner?*, 2015. [Online]. Available: `http://www.alexandersavi.nl/blog/how-to-tailor-massive-education/` (visited on 24/11/2016).

[18] A. Savi, H. L. J. van der Maas and G. K. J. Maris, 'Navigating massive open online courses', *Science*, vol. 347, no. 6225, p. 958, 2015, ISSN: 10959203. DOI: `10.1126/science.347.6225.958`.

[19] S. Downes, 'Places to go : connectivism & connective knowledge', *J. Online Educ.*, vol. 5, no. 1, pp. 1–6, 2008, ISSN: 1552-3233.

[20] J. Reich, 'Rebooting MOOC research', *Science*, vol. 347, no. 6217, pp. 34–35, 2015, ISSN: 1095-9203. DOI: `10.1126/science.1261627`.

[21] R. Kop and A. Hill, 'Connectivism: learning theory of the future or vestige of the past?', *Int. Rev. Res. Open Distance Learn.*, vol. 9, no. 3, 2008, ISSN: 14923831.

[22] J. Wulf, I. Blohm, W. Brenner and J. M. Leimeister, 'Massive open online courses', *Bus. Inf. Syst. Eng.*, 2014, ISSN: 0937-6429. DOI: `10.1007/s12599-014-0313-9`.

[23] M. Sharples, A. Adams, N. Alozie, R. Ferguson, E. FitzGerald, M. Gaved, P. McAndrew, B Means, J Remold, B. Rienties, J Roschelle, K Vogt, D Whitelock and L Yarnall, *Innovating pedagogy 2015: open university innovation report 4*, Milton Keynes, England, 2015. DOI: `10.13140/RG.2.1.4727.6885`.

[24] J. DeBoer, A. D. Ho, G. S. Stump and L. Breslow, 'Changing course: reconceptualizing educational variables for massive open online courses', *Educ. Res.*, vol. March, no. 43, pp. 74–84, 2014, ISSN: 0013-189X. DOI: `10.3102/0013189X14523038`.

[25] J. Kay, P. Reimann, E. Diebold and B. Kummerfeld, 'MOOCs: so many learners, so much potential', *IEEE Intell. Syst.*, vol. 8, no. 3, pp. 70–77, 2013. DOI: `10.1109/MIS.2013.66`.

[26] A. D. Ho, I. Chuang, J. Reich, C. A. Coleman, J. Whitehill, C. G. Northcutt, J. J. Williams, J. D. Hansen, G. Lopez and R. Petersen, 'HarvardX and MITx: two years of open online courses fall 2012-summer 2014', *SSRN*, vol. 10, no. 10, pp. 1–37, 2015, ISSN: 1556-5068. DOI: `10.2139/ssrn.2586847`.

[27] Utrecht University, *Eerste MOOC UU live [first UU MOOC live]*, 2016. [Online]. Available: `https://educate-it-uu.sites.uu.nl/eerste-mooc-uu-live/` (visited on 06/12/2016).

[28] F. G. Martin, 'Will massive open online courses change how we teach?', *Commun. ACM*, vol. 55, no. 8, pp. 26–28, 2012, ISSN: 00010782. DOI: `10.1145/2240236.2240246`.

[29] K. Jordan, *MOOC completion rates: the data*, 2015. [Online]. Available: `http://www.katyjordan.com/MOOCproject.html` (visited on 06/12/2016).

[30] R. M. Carini, G. D. Kuh and S. P. Klein, 'Student engagement and student learning: testing the linkages', *Res. High. Educ.*, vol. 47, no. 1, pp. 1–32, 2006, ISSN: 03610365. DOI: `10.1007/s11162-005-8150-9`.

[31] G. D. Kuh, 'What we're learning about student engagement from NSSE: benchmarks for effective educational practices', *Chang. Mag. High. Learn.*, vol. 35, no. 2, pp. 24–32, 2003, ISSN: 0009-1383. DOI: 10.1080/00091380309604090.

[32] K. M. Alraimi, H. Zo and A. P. Ciganek, 'Understanding the MOOCs continuance: the role of openness and reputation', *Comput. Educ.*, vol. 80, pp. 28–38, 2015, ISSN: 03601315. DOI: 10.1016/j.compedu.2014.08.006.

[33] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé and L. Getoor, 'Modeling learner engagement in MOOCs using probabilistic soft logic', *NIPS Work. Data Driven Educ.*, vol. 21, pp. 1–7, 2013.

[34] T. Sinha, N. Li, P. Jermann and P. Dillenbourg, 'Capturing "attrition intensifying" structural traits from didactic interaction sequences of mooc learners', *EMNLP 2014 Work. Model. Large Scale Soc. Interact. Massively Open Online Courses*, pp. 42–49, 2014. arXiv: 1409.5887v1.

[35] J. He, J. Bailey and B. I. P. Rubinstein, 'Identifying at-risk students in massive open online courses', *Proc. 29th AAAI Conf. Artif. Intell.*, pp. 1749–1755, 2015.

[36] A. Newman, G. Bryant, B. Fleming and L. Sarkisian, 'Learning to adapt 2.0: the evolution of adaptive learning in higher education', Tyton Partners, Boston, MA, Tech. Rep., 2016.

[37] P. Brusilovsky, 'Adaptive and intelligent technologies for web-based education web-based educational systems: a review of technologies', *Künstliche Intelligenz, Spec. Issue Intell. Syst. Teleteaching*, vol. 4, pp. 19–25, 1999.

[38] A. Newman, P. Stokes and G. Bryant, 'Learning to adapt: a case for accelerating adaptive learning in higher education', Tyton Partners, Boston, MA, Tech. Rep., 2013.

[39] J. Lowendahl, T. Thayer and G. Morgan, 'Top 10 strategic technologies impacting higher education in 2016', Gartner Inc., Stamford, CT, Tech. Rep. January, 2016.

[40] T. W. V. Malone, 'What makes things fun to learn? heuristics for designing instructional computer games', in *Proc. 3rd ACM SIGSMALL Symp. first SIGPC Symp. Small Syst.*, vol. 162, 1980, pp. 162–169, ISBN: 0897910249. DOI: 10.1145/800088.802839.

[41] S. Klinkenberg, M. Straatemeier and H. L. J. van der Maas, 'Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation', *Comput. Educ.*, vol. 57, no. 2, pp. 1813–1824, 2011, ISSN: 03601315. DOI: 10.1016/j.compedu.2011.02.003.

[42] J. Torrente, P. Moreno-Ger, B. Fernández-Manjón and Á. del Blanco, 'Game-like simulations for online adaptive learning: a case study', in *Int. Conf. Technol. E-Learning Digit. Entertain.*, Springer, 2009, pp. 162–173, ISBN: 978-3-642-03363-6. DOI: 10.1007/978-3-642-03364-3_21.

[43] E. Martín and R. M. Carro, 'Supporting the development of mobile adaptive learning environments: a case study', *IEEE Trans. Learn. Technol.*, vol. 2, no. 1, pp. 23–36, 2009, ISSN: 1939-1382. DOI: 10.1109/TLT.2008.24.

[44] D. Szafir and B. Mutlu, 'ARTFul: adaptive review technology for flipped learning', in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2013, pp. 1001–1010. DOI: 10.1145/2470654.2466128.

[45] N. Sonwalkar, 'The first adaptive MOOC: a case study on pedagogy framework and scalable cloud architecture', *MOOCs Forum*, pp. 22–29, 2013, ISSN: 23258330. DOI: 10.1089/mooc.2013.0007.

[46] T. Mullaney and J. Reich, 'Staggered versus all-at-once content release in massive open online courses: evaluating a natural experiment', in *Proc. Second ACM Conf. Learn. @ Scale*, 2015, pp. 185–194, ISBN: 978-1-4503-3411-2. DOI: 10.1145/2724660.2724663.

[47] A. R. Hevner, S. T. March and J. Park, 'Design science in information systems research', *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004. DOI: 10.2307/25148625. [Online]. Available: http://dl.acm.org/citation.cfm?id=2017217.

[48] W. W. Royce, 'Managing the development of large software systems', in *Proc. IEEE WESCON*, vol. 26, Los Angeles, CA, 1970, pp. 1–9.

[49] D. Clegg and R. Barker, *Case Method Fast-Track: A RAD Approach.* Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1994, ISBN: 020162432X.

[50] D. J. Weiss, 'Computerized adaptive testing for effective and efficient measurement in counseling and education', *Meas. Eval. Couns. Dev.*, vol. 37, no. 2, p. 70, 2004.

[51] J. M. Linacre, 'Computer-adaptive testing: a methodology whose time has come.', in *Dev. Comput. Middle Sch. Achiev. Test*, Seoul: Komesa Press, 2000.

[52] T. J. Eggen and A. J. Verschoor, 'Optimal testing with easy or difficult items in computerized adaptive testing', *Appl. Psychol. Meas.*, vol. 30, no. 5, pp. 379–393, 2006, ISSN: 0146-6216. DOI: 10.1177/0146621606288890.

[53] W. P. Vispoel, T. Wang and T. Bleiler, 'Computerized adaptive and fixed-item testing of music listening skill: a comparison of efficiency, precision, and concurrent validity', *J. Educ. Meas.*, vol. 34, no. 1, pp. 43–63, 1997, ISSN: 1745-3984. DOI: 10.1111/j.1745-3984.1997.tb00506.x.

[54] J. W. Peltier, W. Drago and J. a. Schibrowsky, 'Virtual communities and the assessment of online marketing education', *J. Mark. Educ.*, vol. 25, no. 3, pp. 260–276, 2003, ISSN: 02734753. DOI: 10.1177/0273475303257762.

[55] University of Cambridge Psychometrics Centre, *Concerto: the open-source online adaptive testing platform*, 2017. [Online]. Available: http://concertoplatform.com/.

[56] R Development Core Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008. [Online]. Available: http://www.r-project.org.

[57] D. Magis and R. Gilles, 'Random generation of response patterns under computerized adaptive testing with the R package catR', *J. Stat. Softw.*, vol. 48, no. 8, 2012, ISSN: 15487660. DOI: http://dx.doi.org/10.18637/jss.v048.i08. [Online]. Available: http://www.jstatsoft.org/v48/i08/.

[58] D. Magis and J. R. Barrada, 'Computerized adaptive testing with R: recent updates of the package catR', *J. Stat. Software, Code Snippets*, vol. 76, no. 1, pp. 1–18, 2017. DOI: 10.18637/jss.v076.c01.

[59] Amazon, *Amazon elastic compute cloud*, 2017. [Online]. Available: www.aws.amazon.com/ec2.

[60] P. Mair and R. Hatzinger, 'Extended Rasch modeling: the eRm package for the application of IRT models in R', *J. Stat. Softw.*, vol. 20, no. 9, 2007. [Online]. Available: http://www.jstatsoft.org/v20/i09.

[61] P Mair, R Hatzinger and M Maier, 'Extended Rasch modeling: the R package eRm', *PDF-Dateianhang zum Program. eRm*, 2009. [Online]. Available: ftp://cygwin.uib.no/pub/cran/web/packages/eRm/vignettes/eRm.pdf.

[62] P. Mair, R. Hatzinger and M. J. Maier, *Erm: Extended Rasch modeling*, 2016. [Online]. Available: http://erm.r-forge.r-project.org/.

[63] J. M. Linacre, 'Rasch dichotomous model vs. one-parameter logistic model (1PL)', *Rasch Meas. Trans.*, vol. 19, no. 3, p. 1032, 2005. [Online]. Available: http://www.rasch.org/rmt/rmt193h.htm.

[64] V. W. Urry, 'A monte carlo investigation of logistic mental test models', *ProQuest Inf. Learn.*, 1971.

[65] International Organization for Standardization, *ISO/IEC 25010:2011 - systems and software engineering – systems and software quality requirements and evaluation (square) – system and software quality models*, 2011. [Online]. Available: https://www.iso.org/standard/35733.html (visited on 27/03/2017).

[66] E. B. Andersen, 'A goodness of fit test for the Rasch model', *Psychometrika*, vol. 38, no. 1, pp. 123–140, 1973.

[67] R. F. Martin, 'General Deming regression for estimating systematic bias and its confidence interval in method-comparison studies', *Clin. Chem.*, vol. 46, no. 1, pp. 100–104, 2000.

[68] Deloitte, *Hacklab high school*. [Online]. Available: https://www2.deloitte.com/nl/nl/pages/risk/events/hacklab-high-school.html.

[69] B. R. Jansen, J. Louwerse, M. Straatemeier, S. H. van der Ven, S. Klinkenberg and H. L. Van der Maas, 'The influence of experiencing success in math on math anxiety, perceived math competence, and math performance', *Learn. Individ. Differ.*, vol. 24, pp. 190–197, 2013. DOI: 10.1016/j.lindif.2012.12.014.

[70] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell and A. Wesslén, *Experimentation in Software Engineering*. New York, NY: Springer Heidelberg, 2000, ISBN: 978-3-642-29043-5. DOI: 10.1007/978-3-642-29044-2.

Appendices

A Survey Questions

Trait	Question	Answer type
Age	How old are you?	Free input (Positive integer)
Gender	What is your gender?	Dropdown (Male or Female)
Experience	How much experience do you have with Cyber Security?	Dropdown (None, A Little, Some, Quite Some, A lot)
Interest	Which module are you most interested in?	Dropdown (Web Hacking, Network Hacking & Forensics, Encryption)
Institution	Which school are you from?	Dropdown (Participating Schools + Other / Not Applicable)

Table A.1: Pre-survey Questions

Measure	Trait	Statement
Satisfaction	*Content*	This course supplied me with an effective range of challenges.
Satisfaction	*Content*	This course included applied learning and problem-solving experiences.
Satisfaction	*Difficulty*	The difficulty of the challenges was at the right level for me.
Satisfaction	*Duration*	The duration of the total test was good.
Satisfaction	*Duration*	The length of the individual modules was well balanced.
Satisfaction	*Theory*	The provided theory (documentation) was interesting.
Satisfaction	*Theory*	The provided theory (documentation) was helpful.
Engagement	*Content*	Course materials stimulated my desire to learn.
Engagement	*Content*	This course effectively challenged me to think.
Engagement	*Theory*	I interacted with the provided theory (documentation) a lot.
Engagement	*Theory*	I used a lot of external help sources.
Engagement	*Motivation*	I felt motivated to find the answer to challenges.
Engagement	*Motivation*	I felt motivated to finish the complete course.

Table A.2: Post-survey Statements