Large-Scale Learning Analytics
Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses

Davis, Daniel

DOI
10.4233/uuid:b8be8302-84a0-4b29-a6fe-761a3f872420

Publication date
2019

Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses

Daniel John Davis

Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Tuesday 7 May 2019 at 12:30 o'clock

by

Daniel John DAVIS

Master of Arts in Communication, Culture & Technology
Georgetown University, USA
born in Worcester, MA, USA

This dissertation has been approved by the promoters.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr.ir. G.J.P.M. Houben	Technische Universiteit Delft, promotor
Dr.ir C. Hauff	Technische Universiteit Delft, co-promotor

Independent Members:

Prof.dr. M.M. Specht	Technische Universiteit Delft
Prof.dr. W.F. Admiraal	Universiteit Leiden
Prof.dr. M. Kalz	Heidelberg University of Education, Germany
Prof.dr. K. Verbert	KU Leuven, Belgium
Prof.dr. A. van Deursen	Technische Universiteit Delft, reserve member

SIKS Dissertation Series No. 2019-14

The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Published and distributed by: Daniel John Davis
E-mail: dan7davis@gmail.com

ISBN: 978-94-028-1475-0

Keywords: learning analytics, web information systems, learning science, educational data mining

Cover design by: Daniel Davis

Printed and bound in The Netherlands by IPSKamp Printing.

Acknowledgments

Mom
Dad
Katie
Allie

Katya

Bentley
John
Nathan

Joe
Nick
Sonny
Brett
Matt
Snax
John

Ron
Barry

Taylor
Freddie
Dragos

Guanliang
René
Daniel
Zach
Ioana
Jorge
Mar
Jonah
Vasileios
Tim
Jacqueline

CEL
Geert-Jan
Henk
Timo
JP
Janine

Yianna
Kelly
Dr. Lang
Dr. Land

Daniel John Davis
April 2019
Atlanta, Georgia, USA

Contents

Chapter 1

Introduction

1.1 Motivation

Online learning environments have been blossoming in popularity and use over recent years with the dawn of large-scale learning technologies. In the past, learning management systems (LMS) were traditionally used simply as file management systems for instructors to store files for convenient student access, as was common in the early stages of e-learning [216, 49]. This was an evolution from the original distance education age where hard copies of course materials were physically mailed to learners [228]. Today, however, entire degrees are being delivered exclusively online. This evolution has generated a demand for a massive reevaluation of teaching methods across various media. These courses and degree programs have been translated from the traditional, physical classroom settings to the online context without critical evaluation, which has led to systematic issues of attrition and passivity in the learning process [129].

The promise and excitement behind large-scale online education lies in its scale and reach [233]—Massive Open Online Courses (MOOCs) offered by TU Delft have more than 2 million enrollments at the time of writing. Learners from all around the globe can participate in courses from top academic institutions as long as they have an internet connection. With this scale, however, come some issues. One of the key issues plaguing the large-scale online education community thus far is its perceived inferior effectiveness compared to traditional, classroom-based learning environments and methods of delivery [114]. Instructors do not have the benefits of face-to-face contact with learners, and this lack of immediate, real-time visual feedback makes managing a classroom more difficult, especially when enrollments count in the tens of thousands for a given course [228].

To tackle this problem of delivering quality online instruction, this thesis presents a series of research efforts which investigate measures to improve online learning by developing technology to incorporate active learning strategies (defined as any process which enables learners to be critically engaged in thinking about and carrying out their learning process [86]) in the course design and online learning platform. For example, the early paradigm of MOOCs in making courseware freely available to anyone in the world was concerned with the delivery of materials [180]. While there is certainly lots of value in that, what this paradigm fails to consider is the fact that learning is a skill, and just because someone has access to such materials does not ensure that they possess the necessary *learning skills* to successfully engage with and learn the content and achieve improved learning outcomes [86]. The learning skills which we focus on encouraging in the systems we deployed are those concerned with self-regulated learning (SRL), or a learner's proactive engagement with the learning process through various personal management strategies in order to control & monitor cognitive and behavioral processes towards a learning outcome [278, 280, 256].

In the series of research efforts presented here, we tackle this problem by first referencing the research literature to discover what types of interventions or improvements can be made to large-scale online learning environments to address the identified issues; conducting exploratory studies on the natural behavior of learners (how they engage with course resources) and strategies of instructors (how they design and build courses); and finally we design, develop, deploy, and evaluate a series of course augmentations (in the form of web applications embedded within the edX MOOC platform[1]) aimed at improving learning outcomes in MOOCs, which suffer from systematic rates of high attrition (only around 5% of learners typically go on to finish a course [129]) by supporting learners in self-regulating their learning process—equipping them with the tools necessary to succeed in what is a foreign learning context for many of them.

The two key SRL skills we aim to support learners in practicing in the interventions deployed in this thesis are *study planning* and *retrieval practice*. Research on study planning has found that students who spend time thinking about, explicitly stating, and reflecting on their goals on a daily, weekly, or even yearly level show increases in both engagement and academic performance [196, 221, 259]. Retrieval practice, also known as the testing effect, is

[1] edx.org

the process of reinforcing prior knowledge by actively and repeatedly recalling relevant information. This strategy is more effective in facilitating robust learning—the committing of information to long-term memory [142]—than passively revisiting the same information, for example by going over notes or book chapters [4, 49, 212, 97, 166, 118, 117]. Both retrieval practice and study planning are examples of *active learning* [86].

The work presented in this thesis highlights the importance and value of the field of *learning analytics*, which investigates computationally modeling learning & teaching behaviors in online education and develops technical solutions to support and improve this modeling. To offer a holistic contribution to the learning analytics field, the present research is a convergence of computer science and educational science—both being imperative to the practical implications of this work. For example, innovations in learning and instruction in the past have largely been conducted at small scales (e.g., classrooms of 40 or fewer students) where personal attention could be paid to each student or participant, and these types of small scale studies are the basis upon which the scientific literature about learning is founded [96]. However, with today's large-scale online classroom with thousands of learners, this type of manual, face-to-face attention is implausible for one instructor (or even a team of teaching assistants) to offer. Accordingly, thanks to new technologies, we here leverage and build new scalable tools (able to serve MOOC learners with no manual work required from an instructor) to deliver personalized learning experiences in large-scale learning environments, thus enabling personalized feedback or attention from the system rather than an instructor. Likewise, learning analytics and large-scale online learning environments also allow for large-scale randomized field experiments for learning interventions [127, 245]. By carefully combining web technologies with a strong theoretical underpinning of the science of learning, we are here able to advance the field of learning analytics.

1.2 Objectives

The primary, overarching research question that this thesis aims to address is:

RQ How does the design of Massive Open Online Courses affect learner success and engagement?

We break it down into two sub-questions to better guide each individual research contribution along the way:

RQ1 To what extent do teaching and learning strategies that have been found to be effective in traditional learning environments translate to MOOCs?

RQ2 How can MOOC environments be improved to advance the possibilities of experimentation?

RQ1 derives from the assumption that has been widely applied in the early years of large-scale online education (2011–present): what works in the controlled laboratory or classroom should also work online at scale [210]. While this is a sensible approach to enact when first exploring the new educational medium, we see it as imperative that the transference of learning and teaching strategies from traditional learning environments to large-scale online learning environments be empirically evaluated and tested. We are particularly interested in how such strategies are realized at scale—from both technological and design perspectives. In this body of research we address this assumption by translating learning and teaching strategies from small to large-scale learning contexts—and the highly heterogeneous learner population demographics therein—and measure the extent to which they do or do not hold in their impact on learning outcomes. In large-scale and automated environments data is available to help us observe and assess these effects, and therefore raising this questions of this transference is opportune; the data traces from the learning environment/behavior contains key information for us to better answer the driving, underlying questions.

We propose **RQ2** to address the technical challenges of delivering high quality education to the masses. Relating back to **RQ1**, in the past, many scientific interventions (treatments to learners to observe a causal effect) have relied on manual labor (such as feedback or personalized support), but with class enrollments ranging in the thousands (TU Delft MOOC enrollments for a single course have reach as high as over 70,000 individual learners), new approaches to instruction must be conceived that rely on system automation rather than manual labor. This line of inquiry is an integral part of the scientific process to both understand and create better solutions for large-scale online education. The other side to this technical challenge is the constraint of platforms: edX and Coursera have emerged as the most popular environments for delivering Massive Open Online Courses—each attracting more than 10 million learners and each offering over 2,000 courses. Each of the software systems presented in this thesis have been developed especially for the edX platform and are deployable on any course on the platform.

In each of the learner support systems we developed and deployed in the following chapters, we discovered high rates of noncompliance, as MOOC learners tend not to engage with materials which are not required in order to pass a course. In randomized experiments, a participant is considered noncompliant when they opt not to engage with the intervention, thus removing the possibility to measure a causal effect of the treatment. From the very first study described below, we observe high rates of noncompliance and set out to address it through more engaging interventions that are most likely to be used by and beneficial to the learner.

1.3 Thesis Outline

This thesis contains four thematic parts. The main body of the thesis details the overall trajectory of the work presented in this thesis—beginning with a literature review identifying the key problems of online learner behavior that past researchers have addressed, followed by quantitative studies further examining & assessing these problems as they manifest in MOOCs, and ending with a series of instructional intervention experiments to measure their effect on learner success and engagement. We then conclude the thesis with a summary of findings and contributions.

1.3.1 Part I: Improving Learning & Teaching Strategies

Chapter 2: Review of Large-Scale Learning Intervention Studies

This chapter considers the rich history in the learning sciences which has evaluated how different teaching strategies can effect positive changes in learner behavior. We conducted a review of the research literature in this domain while only considering interventions that are able to be implemented at scale. Given that the main problem plaguing MOOCs at the time (which persists to the present time of writing) was that of attrition, we sought out to identify which *active learning* strategies would be the most promising to apply and test in a large-scale learning setting such as MOOCs, in posing the following research question:

RQ2.1 Which active learning strategies for digital learning environments have been empirically evaluated, and how effective are they?

To this end we make the following contributions regarding **RQ1**: (i) due to their large scale and heterogeneity of learners and topics, MOOCs were the most difficult environment to generate significant experimental results from, and (ii) we identified the three most promising (and previously successful) types of interventions from the literature[2].

1.3.2 Part II: Teaching & Learning Paths

One of the key affordances of the large-scale data logged through MOOCs is that it is fine-grained (every user action is logged) enough to offer the ability to capture various patterns throughout the learning and teaching process that cannot be identified through surveys or questionnaires [245]. For example, since it is only possible for a learner to meaningfully engage with one task (such as taking a quiz, watching a video, or posting on the discussion forum) at a time, we can model a learner's path through the course showing the order in which he or she engaged with each learning activity. Likewise, a key contribution in this work is the consideration of course structure as a valuable data source for learning analytics insights. Historically, learning analytics studies often only considered the learners' log traces and activities without contextualizing the results or findings within the structure/design of a given course [233]. We here offer a method to contextualize online learning behavior within the unique traits of individual courses by computationally modeling the structure of a course so that any observed trends in learning behavior can be interpreted in the context of the course's design/structure.

Chapter 3: Adherence to the designed Learning Path

Given that the massive attrition rates of MOOCs had gained wide recognition, we wanted to gain a better understanding of where all these learners were going astray and how this behavior might be rectifiable. To address this issue of learner pathways and attrition with regard to **RQ2**, in this study we posed the following guiding research question:

RQ3.1 To what extent do learners adhere to the designed learning path set forth by the instructor?

[2]Chapter 2 is published as "Activating Learning at Scale: A Review of Innovations in Online Learning Strategies" [67], by D. Davis, G. Chen, C. Hauff, and G.J. Houben in *Computers & Education*, 125:327. 344. 2018.

We explored the extent to which learners deviate from the designed path and the extent to which this is related to their eventual success in the course. The key contribution from this work is the finding that high levels of deviance are related to not passing the course and that high levels of adherence to the designed learning path are more likely to lead to passing the course[3].

Chapter 4: Modeling the Anatomy of a Course

Prior to this study, course structures/designs were rarely taken into consideration in MOOC research [63, 260, 131]. However, with the rapidly growing literature of learning analytics research coming from a massive variety of courses, one must question how well insights generated from one context might transfer to another. In line with **RQ2**, we addressed this issue by developing a method to model the design of a course based on its structure, leading to the research questions:

RQ4.1 To what extent can we model the design of a MOOC by employing principles from the learning design literature?

RQ4.2 How can we quantitatively compare and contrast the design of MOOCs?

RQ4.3 Are there structural components that differentiate a MOOC's design?

In this study, we contribute a method to quantify, model, and cluster the structure of online courses using learning design theory to abstract course content from its underlying structure. We were also able to identify some cases of statistically different passing rates between clusters of course structures—thus indicating that this method could be used to arrive at "best practices" indicating which course structures are most likely to lead to certain learning outcomes (or patterns of learner engagement)[4].

[3]Chapter 3 is published as "Gauging MOOC Learners' Adherence to the Designed Learning Environment" [63], by D. Davis, G. Chen, C. Hauff, and G.J. Houben in Proceedings of the Ninth International Conference on Educational Data Mining, 2016.

[4]Chapter 4 is published as "Toward Large-Scale Learning Design: Categorizing Course Designs in Service of Supporting Learning Outcomes" [69], by D. Davis, D. Seaton, C. Hauff, and G.J. Houben in Proceedings of the Fifth Annual ACM Conference on Learning at Scale, 2018.

1.3.3 Part III: Study Planning

Chapter 5: Social Comparison Learner Dashboard

Due to the high levels of noncompliance in the previously mentioned study, we set out to design a study planning intervention that would require less engagement from the learner but is still effective in increasing engagement and passing rates. Whereas the treatment in the previous study encouraged learners to type in an open text field, in this follow-up study, we opted for a dashboard approach to study planning. This way, the learners could simply view a social comparison (the phenomenon of people establishing their social and personal worth by comparing themselves to others [200]) visualization that shows their behavior compared to a previously-successful learner, make any mental notes they choose, and continue in the course accordingly. To this end, regarding **RQ2**, we designed, developed, and deployed the Learning Tracker, a learner dashboard visualization that shows the current learner's behavior compared to previously successful learners taking the same course. We asked the following guiding research questions:

RQ5.1 Do learners benefit from a tool that allows them to engage in a social comparison of their behavior in the course?

RQ5.2 Which learners benefit most from the `Feedback System`?

RQ5.3 Does feedback framed in line with a learner's cultural context lead to increased achievement and self-regulatory behavior compared to a culturally mismatched framing?

To this end we contribute a method to leverage the scale of a MOOC (aggregating hundreds of successful learners behaviors into a digestible dashboard visualization) to be highly effective in increasing passing rates. Across all four of the randomized controlled trials we ran, we observed significant increases in passing rates between groups who received the learning tracker and those who did not[5].

[5]Chapter 5 is published as "Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale" [66], by D. Davis, I. Jivet, R. Kizilcec, G. Chen, C. Hauff, and G.J. Houben in Proceedings of the of 7th International Conference on Learning Analytics and Knowledge, 2017.

Chapter 6: Study Planning Interfaces

We conducted a pilot study focused on study planning which was modeled after numerous previous works [221, 67, 272] which found that in traditional classroom settings, prompting learners to state their plans and intentions for the course led to significant increases in passing rates and drastic reductions in the achievement gap (race and gender).

In this randomized controlled trial, we did not find any significant differences between the treatment and control groups (there were high levels of noncompliance with the intervention), thus indicating that simply encouraging learners to engage with such interventions is insufficient. We contribute the following recommendation that more active measures need to be taken to (i) get learners to meaningfully engage with such interventions and (ii) reap the cognitive benefits they have on their study success in the long term[6].

To further improve study planning mechanisms in MOOCs, built an interactive study planning system which provided real-time learner feedback called SRLx, intended to give learners more autonomy in defining their own goals and intentions for the course. To evaluate **RQ2**, we set out to address the following research questions:

RQ6.1 To what extent do MOOC learners adopt and take advantage of a personalized SRL support tool?

RQ6.2 Does `SRLx` support MOOC learners in promoting effective self-regulated learning behavior?

We also addressed the technical challenge of building an advanced SRL tool with real-time learner feedback. We were able to use the data generated from SRLx to gain insights about learner study planning habits among those who engaged. From these results we conclude that MOOC learners may not desire such interactivity in their SRL process—on the contrary, we contribute the finding that providing learners the right information at the right time, while requiring no immediate/explicit action on their behalf, is enough to elicit significant improvements in learning outcomes and engagement.

[6]Chapter 6 is based on two full conference papers, published as "Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-Regulated Learning Strategies at Scale" [65], by D. Davis, G. Chen, van der Zee, Tim, C. Hauff, and G.J. Houben in Proceedings of the 11th European Conference on Technology-Enhanced Learning, 2016. and "SRLx: A Personalized Learner Interface for MOOCs" [70], by D. Davis, V. Triglianos, C. Hauff, and G.J. Houben in Proceedings of the 13th European Conference on Technology-Enhanced Learning, 2018.

1.3.4 Part IV: Retrieval Practice

Chapter 7: Knowledge Retention and Retrieval Practice

We first conducted a pilot study with an intervention designed based on the SRL strategy of retrieval practice. Previous research [4] has found retrieval practice, or the active recall of information from memory, to be among the most effective strategies for promoting long-term memory.

In this randomized controlled trial we observed high levels of noncompliance—that is, the vast majority of learners ignored or opted not to engage with the intervention. This was yet another indication that out-of-the-box approaches to translating traditional learning strategies to scale would be insufficient.

Whereas the pilot study used simple open text field prompts after lecture videos, this system would automatically deliver retrieval cues to learners based on their history within the course. We used this system to explore the following research questions:

RQ7.1 How does an adaptive retrieval practice intervention affect learners' academic achievement, course engagement, and self-regulation compared to generic recommendations of effective study strategies?

RQ7.2 How does a push-based retrieval practice intervention (requiring learners to act) change learners' retrieval practice behavior?

RQ7.3 To what extent is robust learning facilitated in a MOOC?

In order to produce the system, we also addressed a technical challenge of building a personalized system that encourages retrieval practice within the edX platform. The system needed to be push-based to address the issue of noncompliance—we could not rely on learners to seek out interventions and study materials—and the system needed to be personalized, as an understanding/model of the learners current knowledge state is integral to the retrieval practice strategy. While we observed null results in the causal analysis of this study, the data generated by the system allowed us to model the deterioration of learners' knowledge over time in plotting a forgetting curve, the first analysis of this kind from a MOOC[7].

[7]Chapter 7 is based on published peer-reviewed work as "Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-Regulated Learning Strategies at Scale" [65], by D. Davis, G. Chen, van der Zee, Tim, C. Hauff, and G.J. Houben in Proceedings of the 11th European Conference on Technology-Enhanced Learning, 2016. and "The Half-Life of MOOC Knowledge: A Randomized Trial Evaluating Knowledge

Through each of these research efforts detailed in the following chapters, we gain a deeper understanding of how the design of online learning environments affects learner success and engagement. After drawing insights from the literature on what has been found the be the most effective instructional strategies in traditional learning environments, we describe the extent to which these translate to MOOCs through a series of randomized experiments.

Retention and Retrieval Practice in MOOCs" [68], by D. Davis, R. Kizilcec, C. Hauff, and G.J. Houben in Proceedings of 8th International Conference on Learning Analytics and Knowledge, 2018.

Part I

Improving Learning Outcomes

This part serves **RQ1** (*To what extent do teaching and learning strategies that have been found to be effective in traditional learning environments translate to MOOCs?*) by focusing on measures taken to improve student behavior and learning outcomes in large-scale learning environments. To this end we go to the literature and seek out past examples of innovations in teaching and learning strategies that could be applied at scale with the goal of improving large-scale learning environments.

Chapter 2 asks the question: which active learning strategies for digital learning environments have been empirically evaluated, and how effective are they? By surveying the literature in service of this question, we contribute the identification of a number of trends in this space and highlight recommendations for future research as well. We found that (i) experiments conducted in large-scale environments (more than 500 participants) were the least likely to generate significant results and (ii) cooperative learning, simulations & gaming, and interactive multimedia are the most effective and promising strategies for driving positive change in learner behavior in large-scale learning environments.

Chapter 2

Activating Learning at Scale: A Review of Innovations in Online Learning Strategies

Taking advantage of the vast history of theoretical and empirical findings in the learning literature we have inherited, this research offers a synthesis of prior findings in the domain of *empirically evaluated active learning strategies in digital learning environments*. The primary concern of the present study is to evaluate these findings with an eye towards scalable learning. Massive Open Online Courses have emerged as the new way to reach the masses with educational materials, but so far they have failed to maintain learners' attention over the long term. Even though we now understand how effective active learning principles are for learners, the current landscape of MOOC pedagogy too often allows for passivity — leading to the unsatisfactory performance experienced by many MOOC learners today. Through our systematic search we found 126 papers meeting our criteria and categorized them according to Hattie's learning strategies. We found large-scale experiments to be the most challenging environment for experimentation due to their size, heterogeneity of participants, and platform restrictions, and we identified the three most promising strategies for effectively leveraging learning at scale as Cooperative Learning, Simulations & Gaming, and Interactive Multimedia.

This chapter is published as "Activating Learning at Scale: A Review of Innovations in Online Learning Strategies" [67], by D. Davis, G. Chen, C. Hauff, and G.J. Houben in *Computers & Education*, 125:327. 344. 2018.

2.1 Introduction

In the dense landscape of scalable learning technologies, consideration for sound pedagogy can often fall by the wayside as university courses are retroactively translated from a classroom to the Web. Up against the uncertainty of how to best rethink and conceive of pedagogy at scale, we here synthesize the previous findings as well as highlight the possibilities going forward with the greatest potential for boosting learner achievement in large-scale digital learning environments.

Now that the initial hype of Massive Open Online Courses has passed and the Web is populated with more than 4,000 of these free or low-cost educational resources, we take this opportunity to evaluate and assess the state-of-the art in pedagogy at scale while identifying the best practices that have been found to significantly increase learner achievement.

This study conducts a review of the literature by specifically seeking innovations in *scalable* (not requiring any physical presence or manual grading or feedback) learning strategies that aim to create a more active learning experience, defined in Freeman et al. [86] as one that “engages students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert. It emphasizes higher-order thinking and often involves group work." By limiting the selection criteria to empirical research that can be applied at scale, we aim for this survey to serve as a basis upon which future MOOC design innovations can be conceived, designed, and tested. We see this as an important perspective to take, as many learning design studies provide design ideas, but do not contain a robust empirical evaluation. We certainly do not intend to discount the value of observational or qualitative studies in this domain; rather, for the following analyses we are primarily concerned with results backed by tests of statistical significance because this offers a more objective, quantitative measure of effectiveness.

2.2 Method

The driving question underpinning this literature survey is:

RQ2.1 Which active learning strategies for digital learning environments have been empirically evaluated, and how effective are they?

To begin the literature search we utilized John Hattie’s *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* [96] as a basis. It provides a comprehensive overview of findings in the domain of empirically tested learning strategies in traditional classroom environments. As Hattie's work was published in 2008, we used that as a natural starting point for our review, working forward to July 2017. It creates a narrow enough scope (nine years: 2009-2017) and temporally relevant (MOOCs went mainstream in 2012) time constraints for the review. We manually scanned all publications released from our selected venues in this time period and determined for each whether or not they met our criteria: (1) the learning strategy being analyzed must have been **scalable** — it must not require manual coding, feedback, physical presence, etc., (2) the evidence must come from empirical analyses of **randomized controlled experiments** with a combined **sample size of at least ten** across all conditions, and (3) the subjects of the studies must be **adult learners**, i.e. at least 18 years old. We included the age criterion based on the profile of the typical MOOC learner — aged 25-35 according to [247], which aligns with our own institution's data as well.

From Hattie's synthesis of meta-analyses we identified the 10 core learning strategies that best apply to open online education — only selecting from those which Hattie found to be effective. With these learning strategies fixed, we systematically reviewed all publications in five journals and eight conferences (listed in Table 2.1) that have displayed a regular interest in publishing work on testing these categories of innovative online learning strategies. These venues were identified and selected based on an exploratory search through the literature—we began with a sample of studies we were previously familiar with that fit the scope of the present review and perused the references of each to identify more potential venues worth exploring. This process was repeated for each identified study thereafter. The lead author also reached out to experts in the field to assure that this method did not overlook any potential venues. The thirteen venues used for the final review are those which showed the most consistent interest in publishing studies that meet our criteria. We employed this method over a search/-query term method because our criteria (namely that of being a randomized controlled trial among adult populations) are not reliably gleanable from standard search engine indexing.

We acknowledge there are other journals and conference proceedings that may have been applicable for this survey, but given our search criteria, we found these thirteen venues to be the most appropriate based on our initial exploratory search.

Table 2.1: Overview of included venues. The most recent included issue from each publication is indicated in parentheses. Unless otherwise indicated with a †, the full proceedings from 2017 have been included.

Computers & Education (Vol. 114)
Journal of Learning Analytics (Vol. 4, No. 2)
Journal of Educational Data Mining (Vol. 8, No. 2)
The Open Education Journal eLearning Papers (Issue 43)
IEEE Transactions of Learning Technologies (Vol. 10, Issue 1)
ACM Learning @ Scale (L@S)
Learning Analytics & Knowledge (LAK)
European Conference on Technology-Enhanced Learning (EC-TEL) †
International Conference on Educational Data Mining (EDM)
ACM Conference on Computer-Supported Cooperative Work (CSCW)
European MOOCs Stakeholders Summit (EMOOCs)
European Conference on Computer-Supported Collaborative Work (ECSCW)
Human Factors in Computing Systems (CHI)

Table 2.2: Overview of considered learning categories. The selected papers per category are shown in parentheses. The sum of the numbers is 131 and not 126, as five papers apply to two categories.

Mastery Learning (1)
Meta-Cognitive Strategies (24)
Questioning (9)
Spaced vs. Massed Practice (1)
Matching Learning Styles (3)
Feedback (21)
Cooperative Learning (17)
Simulations & Gaming (18)
Programmed Instruction (6)
Interactive Multimedia Methods (31)

Of the **7,706** papers included in our search, we found **126** (1.6%) to meet our criteria. The criterion requiring randomized controlled trials proved to be a strong filter with many studies not reporting randomization or a baseline condition to compare against. Overall, these 126 papers report on experiments with a total of **132,428** study participants. We then classified each work into one of the ten learning strategy categories (listed in Table 2.2).

Figure 2.1 illustrates the number of studies that met our selection criteria organized by the year published. It shows the increasing frequency of such experiments in recent years, with the most notable increase from 2014 to 2015.

We could propose any number of explanations for the decrease in studies from 2015 to 2016, but it would be purely speculation. However, when examining the studies themselves, we do notice a prominent trend with some explanatory power. With the dawn of MOOC research emerging around 2013 and 2014, the experiments carried out in this window can be viewed now, in hindsight, as foundational. Such interventions in this era included sending out emails to learners [137] or dividing the course discussion forum and controlling instructor activity [247]. However, in 2016 and 2017 we begin to see an elevated level of complexity in interventions such as the *adaptive and personalized* quiz question delivery system [215] implemented and evaluated at scale in a MOOC. It is also worth noting that a number of journal issues and conference proceedings from 2017 had not yet been released at the time of this writing (indicated in Table 2.1).

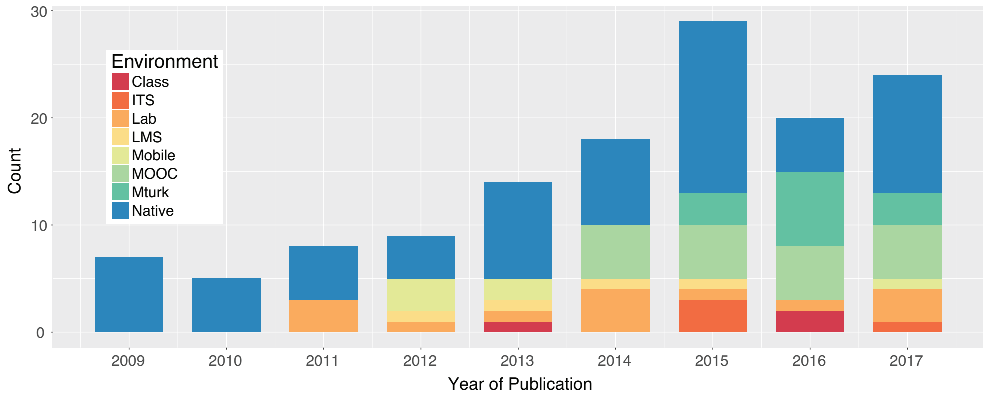


Figure 2.1: The number of papers by year and learning environment meeting our selection criteria. Each environment is defined in detail in Section 2.3.1. Best viewed in color.

Figure 2.2 shows the proportion of results (positive, null, or negative) with respect to the experimental environment employed by the selected articles/studies. Noting the difference between MOOCs and native environments (those designed and implemented specifically for the study), we see native environments yielding positive results at a much stronger rate than MOOCs (59% vs. 42% respectively). We see two main factors contributing to this difference: (i) native environments can be modeled specifically for the experiment/tested concepts, whereas experiments done in MOOCs must adapt to the existing platforms and (ii) no MOOC studies provide participants any incentive to participate, whereas this is common to experiments in native environments.

Figure 2.4 further visualizes this discrepancy in illustrating the proportion of positive, negative, and null results across three subject pool sizes: small-scale studies with between 10 and 100 participants, medium-sized studies with 101– 500 participants and large-scale studies with more than 500 study participants. We here find a statistically significant difference in the proportion of reported positive findings in large (42% in studies with 500+ participants) and small (60% in studies with 10–100 participants) studies using a χ^2 test ($p < 0.05$). As the focus of this study is on large-scale learning, we specifically ran this analysis to evaluate the impact that scale and, in turn, sample size and heterogeneity can have on an experiment.

We registered this project with the Center for Open Science[2], and the registration which includes all data gathered as well as scripts used for analysis & visualization are available at https://osf.io/jy9n6/.

2.3 Terminology

We now define the terminology used in the reporting of our results. Not only is this explicit terminology elucidation important for the clarity of this review, it can also serve as a reference for future experiments in this area to ensure consistency in how results are reported and replicated. In discussing each study, we refer to “learners", “students", or “participants" as the authors do in the referenced work.

2.3.1 Environment

The first dimension by which we categorize the studies is the environment wherein the experiment/intervention took place. We distinguish between the following:

- **Intelligent Tutoring System (ITS)**: Digital learning systems that monitor and adapt to a learner’s behavior and knowledge state.
- **Laboratory Setting (Lab)**: Controlled, physical setting in which participants complete the experimental tasks.
- **Learning Management System (LMS)**: Software application used to host & organize course materials for students to access online at any time.
- **Mobile Phone Application (Mobile)**: Participants must download and use an application on their mobile phone to participate in the experiment.
- **MOOC**: Online course which offers educational materials free of cost and with open access to all.

[2] https://cos.io

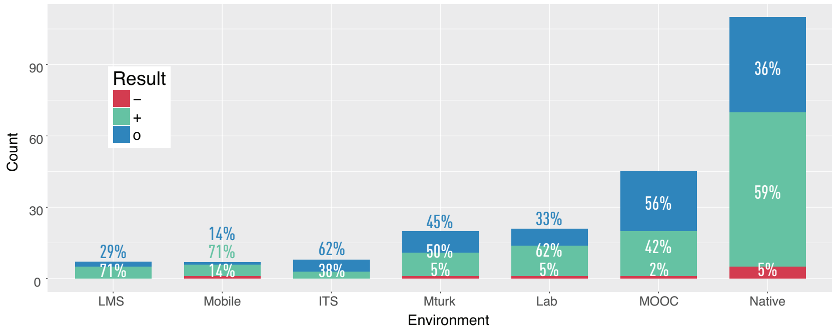


Figure 2.2: Reported results (y-axis) from papers (partitioned by environment) meeting our selection criteria partitioned by environment, incentive, and size. The red "-" indicates studies reporting a significant negative effect of the intervention; the green "+" indicates a significant positive effect of the intervention; and the blue "o" indicates findings without a statistically significant effect. Best viewed in color.

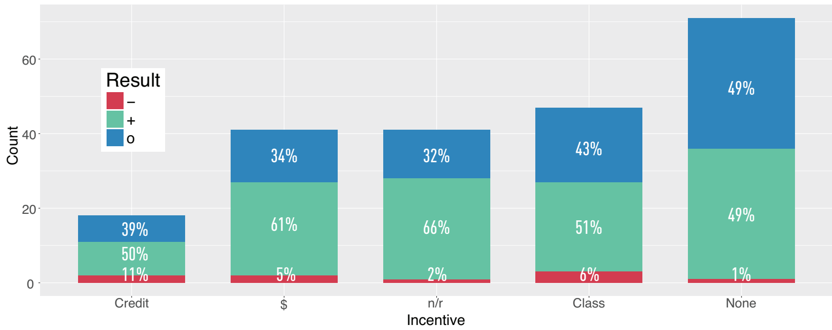


Figure 2.3: Papers partitioned by incentive.

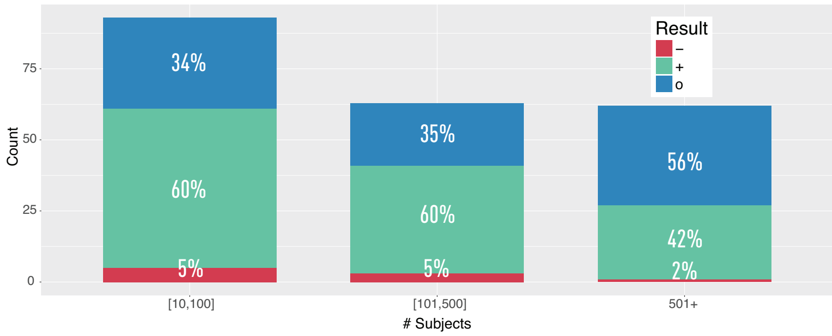


Figure 2.4: Papers partitioned by number of study participants.

- **Amazon Mechanical Turk (Mturk)**: Online marketplace used to host low-cost micro-payment tasks for crowdsourcing and human computation. Participants are recruited and paid through MTurk and often redirected to an external application.

- **Native**: A piece of software designed and implemented specifically for the study.

Figure 2.1 shows the breakdown of our studies with respect to the environment. Note that despite the widespread availability of MOOC and LMS environments in 2015, native environments still dominated that year. We speculate that this may be because researchers find it more efficient to build their own environment from scratch rather than adapt their study to the limitations of a pre-existing platform—which is the case with all MOOC experiments included in this study; each intervention had to be designed within the confines of either the edX or Coursera[3] platforms. We also note a sudden spike in popularity for studies using Mturk from 2015 to 2016. While it is more expensive to carry out research with Mturk compared to MOOCs (which provide no incentive or compensation), Mturk ensures a certain level of compliance and engagement from the subjects in that they are rewarded for their time with money.

2.3.2 Incentive

The second dimension we distinguish is the incentive participants in each study received for their participation:

- **Monetary Reward ($)**: Participants receive either a cash reward or a gift certificate.
- **Required as part of an existing class (Class)**: An instructor conducts an experiment in her own course where all enrolled students are participants.
- **Class Credit (Credit)**: By participating in the study, participants receive course credit which can be applied to their university degree.
- **None**: Participants were not provided any incentive or compensation.
- **n/r**: Not reported.

2.3.3 Outcome Variables

As experiments on learning strategies can evaluate a multitude of outcomes, here we provide an overview of all learning outcomes reported in the included studies.

[3] www.coursera.org

- **Final Grade**: the cumulative score over the span of the entire course which includes all graded assignments.
- **Completion Rate**: the proportion of participants who earn the required final passing grade in the course.
- **Learning Gain**: the observed difference in knowledge between pre-treatment and post-treatment exams
- **Exam Score**: different from the final grade metric in that this only considers learner performance on one particular assessment (typically the final exam).
- **Long-Term Retention**: measured by assessing a learner's knowledge of course materials longitudinally, not just during/immediately after the experiment.
- **Learning Transfer**: measuring a learner's ability to apply new knowledge in novel contexts beyond the classroom/study.
- **Ontrackness**: the extent to which a learner adheres to the designed learning path as intended by the instructor.
- **Engagement**: a number of studies measure forms of learner activity/behavior and fall under this category. Specific forms of engagement include:
  - **Forum Participation**: measured by the frequency with which learners post to the course discussion forum (including posts and responses to others' posts).
  - **Video Engagement**: the amount of actions (pause, play, seek, speed change, toggle subtitles) a learner takes on a video component.
  - **Revision**: the act of changing a previously-submitted response.
  - **Persistence/Coverage**: the amount of the total course content accessed. For example, a learner accessing 75 out of the 100 components of a course has 75% persistence.
- **Self-Efficacy**: a learner's self-perceived ability to accomplish a given task.
- **Efficiency**: the rate at which a learner progresses through the course. This is most commonly operationalized by the amount of material learned relative to the total time spent.

2.4 Review

In the following review we synthesize the findings and highlight particularly interesting aspects of certain experiments. Unless otherwise indicated, all results presented below come from intention-to-treat (ITT) analyses, meaning all participants enrolled in each experimental condition are considered without exception. Each category has a corresponding a table detailing the total sample size ("**N**"), experimental environment ("**Env.**"), incentive for participation ("**Incentive**"), and reported results ("**Result**"). In the **Result** column, statistically significant positive outcome variables as a result of the experimental treatment are indicated with a +; null findings where no significant differences were observed are indicated with a ∘; and negative findings where the treatment resulted in an adverse effect on the outcome variable are indicated with a -.

2.4.1 Mastery Learning

Teaching for mastery learning places an emphasis on learners gaining a full understanding of one topic before advancing to the next [20]. Given that students' understanding of new topics often relies upon a solid understanding of prerequisite topics, teaching for mastery learning only exposes students to new material once they have mastered all the preceding material, very much in line with constructivist thinking as outlined by [60]. In the traditional classroom, teaching for mastery learning presents a major challenge for teachers in that they must constantly monitor each individual student's progress towards mastery over a given topic—a nearly impossible task in a typical classroom with 30 students, never mind 30,000. However, with the growing capabilities of education technologies, individualized mastery learning pedagogy can now be offered to students at scale.

While mastery learning is so frequently found to be an effective teaching strategy in terms of student achievement, it often comes at the cost of time. This issue of time could be a reason behind there being only one paper in this category. [186] implemented a data-driven knowledge tracing system to measure student knowledge and release content according to their calculated knowledge state. Students using this system were far more engaged than those using the default problem set or that with on-demand hints. A strict implementation of mastery learning — as in [186], where learners in an ITS are required to demonstrate concept mastery before advancing in the system

— would be useful to understand its effect on the heterogeneous MOOC learner population.

Table 2.3: Mastery Learning

Mastery Learning: + : 1				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Mostafavi et al. [186]	302	ITS	Class	+Engagement

2.4.2 Metacognitive Strategies

Metacognitive behavior is defined by [96] as "higher-order thinking which involves active control over the cognitive processes engaged in learning." Metacognition is an invaluable skill in MOOCs, where learners cannot depend on the watchful eye of a teacher to monitor their progress at all times. Instead, they must be highly self-directed and regulate their own time management and learning strategies to succeed. The papers in this category explore novel course designs and interventions that are intended to make learners more self-aware, reflective, and deliberate in the planning of (and adherence to) their learning goals.

[65] conducted two experiments: in study "A" they provided learners with retrieval cue prompts after each week's lecture, and in study "B" they provided study planning support to prompt learners to plan and reflect on their learning habits. Overall, neither intervention had any effect on the learners in the experimental conditions, likely because the learners could ignore the prompts without penalty. However, when narrowing down to the very small sample of learners who engaged with the study planning module, the authors found desirable significant increases in learner behavior. [172] also ran an experiment testing support for retrieval practice. They found that (i) retrieval prompts increase learning gain and (ii) the complexity of the retrieval prompt had a significant impact on the prompts effect, with deeper prompts leading to better learning gains. In contrast, the retrieval prompts used by [65] assessed shallow, surface-level knowledge, which could be a reason for the lack of a significant effect.

Even though the education psychology literature suggests that boosting learners' metacognitive strategies is highly effective for increasing learning outcomes [96], 23 of the 38 results (61%) in this category report null or negative findings. Furthermore, with the reporting of a negative impact of an intervention, [137] found a certain form of participation encouragement (collectivist-framed prompting) to actually decrease learners' participation in the course discussion forum.

[188] conducted a study evaluating the effect of framing a group learning activity in different ways. Compared to a "group processing" frame of mind (where group members are asked to assess the contribution of each group member), the "positive interdependence" frame of mind (where group members are reminded that boosting one's individual performance can have a great impact on the overall group achievement) group had higher post assessment scores.

In lieu of an actual learning platform, crowdworker platforms are also beginning to be used for learning research. One example is the study by [87], who evaluated the effect of achievement priming in information retrieval microtasks. While completing a crowdworker task aimed at teaching effective information retrieval techniques, the participants were also assessed on their learning through a test at the end of the task. By providing achievement primers (in the form of inspirational quotes) to these crowdworkers, the authors observed no significant difference in persistence or assessed learning. Given the ease with which these experiments can be deployed, more work should go into exploring the reproducibility of findings from a crowdworker context to an actual learning environment.

In summary, the current body of work in supporting learners' metacognitive awareness indicates how difficult it is to affect such a complex cognitive process, as more than half of the reported results from this category led to non-significant results. While some studies do indeed report positive results, the overall trend in this category is an indication that we have not yet mastered the design and implementation of successful metacognitive support interventions that can effectively operate at scale. Setting this apart from other categories is the difficulty to measure metacognition; compared to other approaches such as questioning (where both the prompt and response are easily measurable), both eliciting and measuring responses to metacognitive prompts is far more challenging.

2.4.3 Questioning

[96] found questioning to be one of the most effective teaching strategies in his meta-analysis. Questioning is characterized by the posing of thoughtful questions that elicit critical thought, introspection, and new ways of thinking. The studies in this category explore new methods of prompting learners to retrieve and activate their prior knowledge in formative assessment con-

Table 2.4: Metacognitive Strategies

Metacognitive Strategies: + : 15 / ○ : 20 / − : 3				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Kizilcec et al. [132]	653	MOOC	None	○Persistence ○Final Grade
Lang et al. [153]	950	ITS	None	○Learning Gain ○Engagement
Lamb et al. [151]	4,777	MOOC	None	+Forum Participation
Sonnenberg and Bannert [238]	70	Native	$	+Learning Gain
Dodge et al. [77]	882	LMS	Class	○Final Grade
Tabuenca et al. [244]	60	Native	$	○Exam Score
Kizilcec et al. [137]	11,429	MOOC	None	○Forum Participation −Forum Participation
Margulieux and Catrambone [177]	120	Native	Credit	+Exam Score
Xiong et al. [267]	2,052	Native	None	+Learning Gain +Completion Rate
Noroozi et al. [195]	56	Native	n/r	+Learning Gain
Davis et al. [65]A	9,836	MOOC	None	○Final Grade ○Engagement ○Persistence
Davis et al. [65]B	1,963	MOOC	None	○Final Grade ○Engagement ○Persistence
Maass and Pavlik Jr [172]	178	Mturk	$	+Learning Gain
Kizilcec et al. [133]	1,973	MOOC	None	+Final Grade +Persistence +Completion Rate
Yeomans and Reich [272]A	293	MOOC	None	−Completion Rate
Yeomans and Reich [272]B	3,520	MOOC	None	−Completion Rate ○Engagement
Gadiraju and Dietze [87]	340	Mturk*	$	○Final Grade ○Persistence
Kim et al. [124]	378	Mturk	$	+Final Grade
Hwang and Mamykina [105]	225	Native	n/r	+Learning Gain
De Grez et al. [71]	73	Native	Class	○Learning Gain
Nam and Zellner [188]	144	Native	Class	○Engagement +Final Grade
Huang et al. [102]	60	Mobile	None	+Final Grade
Poos et al. [207]	80	Lab	None	○Final Grade ○Learning Transfer
Gamage et al. [88]	87	MOOC	n/r	+Learning Gain

texts. [271] evaluated the effectiveness of a two-tier questioning technique, described as "...a set of two-level multiple choice questions [in which the] first tier assesses students' descriptive or factual knowledge...while the second tier investigates the reasons for their choices made in the first tier." They found this questioning technique to be highly effective in their experiment, with learners in the two-tier condition achieving 0.5 standard deviations better learning gains than learners receiving standard one-tier questions.

Instructional questioning was explored in the Mturk setting by [263] who compared the effectiveness of different questioning prompt wordings. They found prompts that directly ask the learner to provide an explanation of why an answer is correct leads learners to revise their answers (to the correct one) more than a prompt asking for a general explanation of the answer.

[72] conducted a study where half of the learners were cued to generate their own inferences through self-explaining and half were provided prewritten instructional explanations. Taking place in the context of a course about the human cardiovascular system, results show that learners prompted to self-explain performed better on the final test, but did not show any difference in persistence or learning transfer from the given explanation group.

Given its effectiveness and relative simplicity to implement, two-tier questioning should be further investigated in the MOOC setting to stimulate learners critical thought beyond surface-level factual knowledge.

Related to the tactic of questioning is the learning strategy known as retrieval practice, or the testing effect, which is characterized by the process of reinforcing prior knowledge by actively and repeatedly recalling relevant information [4]. Recent work has found retrieval practice to be highly effective in promoting long-term knowledge retention [4, 49, 212, 97, 166, 118, 117]. Accordingly, we recommend that future research interested in questioning tactics is designed to stimulate learners to engage in retrieval practice.

2.4.4 Spaced vs. Massed Practice

[96] describes the difference between spaced learning (sometimes referred to as distributed practice) and massed practice as "the frequency of different opportunities rather than merely spending more time on task." In other words, distributing one's study sessions over a long period of time (e.g., 20 minutes per day for 2 weeks) is characteristic of high spacing, whereas studying in intense, concentrated sessions (one four-hour session) is characteristic of massed practice [266]. Historically, studies have found that the desired effect

Table 2.5: Questioning

Questioning: + : 7 / ○ : 7				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Yang et al. [271]	43	Native	n/r	+Learning Gain
Thompson et al. [246]	43	Native	Class	+Learning Gain
Williams et al. [263]	659	Mturk	$	+Revision
Şendağ and Ferhan Odabaşı [225]	40	Native	Class	○Learning Gain ○Final Grade
Chen [42]	84	Native	Class	+Learning Gain
de Koning et al. [72]	76	Native	Credit	+Final Grade ○Persistence ○Learning Transfer
Yang et al. [270]	79	Native	Class	+Final Grade ○Engagement
Attali [8]	804	Mturk	$	+Learning Gain
Attali and van der Kleij [10]	2,445	Native	None	○Persistence ○Final Grade

of spaced learning (long-term knowledge retention) is found most commonly in tasks of low difficulty, and the effect decreases as the difficulty increases [214].

[73] developed a mobile phone "Vocabulary Wallpaper" which aimed to implicitly teach (through the learners mobile phone background) learners new vocabulary in a second language in highly spaced microlearning sessions. Their findings show that, compared to learners receiving the lessons at less distributed rates, learners with highly-spaced exposure showed a significant increase of second language vocabulary learned.

As evidenced by the lone study in the category, it is difficult to design and implement experiments that effectively get learners to commit to high spacing (ideally enacted as a learned self-regulation skill). Even still, given its proven effectiveness elsewhere in the learning literature [96], practitioners and researchers should tackle this design challenge in creating and evaluating environments that encourage spaced practice.

Table 2.6: Spaced vs. Massed Practice

Spaced vs. Massed Practice: + : 1				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Dearman and Truong [73]	15	Mobile	$	+Learning Gain

2.4.5 Matching Learning Styles

[31] conducted an experiment testing the efficacy of "learning style-adapted e-learning environments." In the study, where students self-proclaimed learning styles were either matched or unmatched, yielded no significant differences in terms of learner achievement between conditions. Consistent with the current popular literature on the topic [126, 125], the authors found that adapting the courses to students' learning styles did not result in any significant benefit.

[236] employed a game-based learning environment to evaluate the impact of adapting instruction to learning styles in a computer programming learning context. The authors report that compared to the groups using a non-adaptive version of the SQL language tutor software, the adaptive system yielded no difference in final grades [236].

However, there does still exist some evidence in favor of this learning strategy. [44] created a online learning environment where the teaching strategy was adapted to each of the learners' individual thinking styles. With three teaching strategies (constructive, guiding, or inductive) either matched or unmatched to three thinking styles (legislative, executive, or judicial, respectively), the authors found that the group who had their thinking style matched accordingly outperformed those who did not.

Instead of adapting to a single modality that a learner prefers (such as being a "visual learner"), the literature on learning styles emphasizes that while one modality may be preferred by the learner (and can lead to positive experimental results in certain contexts), providing them instruction in a variety of modalities will provide the greatest benefit overall [126].

Table 2.7: Matching Learning Styles

Matching Learning Styles: + : 2 / ○ : 2				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Brown et al. [31]	221	Native	Class	○Exam Score
Chen et al. [44]	223	Native	n/r	+Final Grade
Soflano et al. [236]	120	Native	None	○Final Grade +Efficiency

2.4.6 Feedback

[96] defines feedback as "information provided by an agent (e.g., teacher, peer, book, parent, or one's own experience) about aspects of one's performance or understanding." Strategically providing students with feedback

offers them the chance to reflect and reassess their approach to a given situation. Feedback can best be thought of as a mirror for learners; it serves to encourage them to stop and mindfully evaluate their own behavior or learning processes — which are otherwise unconscious or unconsidered — and make them readily visible. However, this act of mindfully evaluating and altering one's behavior should not be taken for granted. Self-regulating one's own learning processes (especially in response to feedback) is a skill which is highly correlated with and caused by prior education [265]. Especially in the MOOC context, where the learners come from many diverse backgrounds, it is imperative that feedback offered to the learner is adaptive and aligned to their ability to process, understand, and act upon it.

While [96] finds feedback to be the most effective teaching strategy in his entire meta-analysis, we find very mixed results in our selected studies in terms of its effectiveness. Of the 38 results reported within the 21 papers of this category, only 14 (37%) are positive findings.

Zooming in on two of the MOOC studies in this category, [50] and [247] evaluated the effectiveness of feedback in the context of the discussion forum. [50] tested the effectiveness of implementing a reputation system in a MOOC discussion forum — the more you post to the forum, the more points you accumulate (this paper also applies to the Simulations & Gaming category for this reason). The authors found that providing this positive feedback did indeed lead learners to post more frequently in the forum, but this did not have any impact on their final course grade. [247] ran an experiment in which learners were divided into one of two course discussion forums — in one forum the instructor was active in providing individualized feedback to learners and engaging in discussion, and in the other no instructor feedback was provided. The authors report no differences in either completion rate or course engagement between the two conditions.

To address the challenge of providing in-depth feedback on students' learning in a coding context, [261] tested the effectiveness of a code style tutor which offered adaptive, real-time feedback and hints to students learning to code. Compared to a control group receiving a simplified feedback system consisting of a single unified score assessing the code, students who used the adaptive feedback system did not show any difference in the extent to which they improved their coding style [261].

[18] developed and evaluated an animated pedagogical agent which was able to provide different types of emotional feedback to participants in a simulated environment. They found that positive emotional feedback (expressing happiness and encouragement in response to desirable behavior) led

to significantly higher test scores than negative feedback (where the agent expressed anger and impatience to undesirable behavior). Also taking place in a simulated environment, the experiment carried out by [224] evaluated the effectiveness of a feedback-enabled simulation learning environment. Compared to students in the intervention group who interacted with the feedback-enabled simulation environment, those in the control condition, who did not have access to the simulation, performed more poorly on a final assessment.

While navigational feedback (support for learners in optimizing their learning path through a course) like that introduced by [23] is common in ITS to help learners through problems, the challenge now arises to provide personalized feedback at scale on other factors such as learner behavior patterns. This way, feedback can be used as a mechanism to make learners more aware of their learning habits/tendencies and, in turn, better at self-regulating. However, with only 37% of results reported in this category being positive, this highlights the fact that simply providing feedback is insufficient in promoting positive learning outcomes—these results are an indication that, even though we now have developed the technology to enable the delivery of feedback at scale, attention must now be shifted towards understanding the nuance of what type of feedback (and with what sort of frequency) will help the learner in a given context or state.

2.4.7 Cooperative Learning

Interventions targeting cooperative learning explore methods to enable learners in helping and supporting each other in the understanding of the learning material. Cooperative learning is one of the major opportunity spaces in MOOCs for their unprecedented scale and learner diversity, as evidenced by the prevalence of reported positive findings (71%). The studies in this category develop and test solutions which try to find new ways to bring learners together no matter where they are in the world to complete a common goal.

One successful example of this is the study by [147] where MOOC learners were divided into small groups (between 2 and 9 learners per group) and allowed to have discussions using real-time video calls over the internet. Each group was given prompts encouraging the learners to both discuss course materials and share general reflections of the course experience. The authors found that learners in groups with a larger diversity of nationalities performed significantly better on the course final exam than learners in groups with low diversity. This result shows promise that the scale and diversity of MOOC

Table 2.8: Feedback

Ref.	N	Env.	Incentive	Result
Feedback: + : 14 / ∘ : 22 / - : 2				
Kulkarni et al. [148]	104	Native	None	+Exam Score ∘Revision
Williams et al. [262]	524	MTurk	$	+Exam Score ∘Self-Efficacy
Eagle and Barnes [78]	203	Native	Class	+Persistence
Kardan and Conati [116]	38	Native	n/r	+Learning Gain ∘Exam Scores ∘Efficiency
Fossati et al. [84]	120	Native	Class	+Learning Gain
Borek et al. [23]	87	Native	Class	∘Exam Score ∘Learning Transfer +Learning Gain
Coetzee et al. [50]	1,101	MOOC	None	∘Final Grade ∘Persistence +Forum Participation
Tomkin and Charlevoix [247]	20,474	MOOC	None	∘Completion Rate ∘Engagement
Beheshitha et al. [16]	169	Class	None	∘Forum Participation
Rafferty et al. [209]	200	Mturk	$	∘Learning Gain
Bian et al. [18]	56	Lab	None	+Self-Efficacy +Exam Score
Nguyen et al. [193]	205	Mturk	$	+Final Grade - Revision
Wiese et al. [261]	103	Native	Class	∘Learning Gain ∘Revision
Davis et al. [66]	33,726	MOOC	None	+Completion Rate ∘Engagement
Mitrovic et al. [181]	41	Native	Class	∘Learning Gain ∘Engagement ∘Final Grade +Efficiency
Corbalan et al. [54]	34	Native	n/r	∘Engagement ∘Final Grade
González et al. [91]	121	Native	Class	+Final Grade
van der Kleij et al. [252]	152	Native	Class	∘Final Grade
Erhel and Jamet [80]	41	Lab	n/r	∘Final Grade
Christy and Fox [47]	80	Lab	Credit	-Final Grade
Sedrakyan et al. [224]	66	Native	n/r	+Final Grade

learners can actually bring something novel to the table in learners' apparent interest in cultural diversity.

On a similar note, [275] developed an algorithm which aimed to divide MOOC learners into small groups in a more effective fashion compared to randomization. This model took into consideration the following factors: collaboration preferences (local, email, Facebook, Google+ or Skype), gender, time zone, personality type, learning goal, and language. The authors found this algorithmic sorting of students into groups to not have any effect on overall engagement, persistence, or final grade. Whereas this algorithm grouped largely for *similarity* (for example, grouping learners in the same time zone together), the study presented by [147] suggests that *diversity* may be a better approach to automated group formation.

There are also possibilities for cooperative learning in which the learners do not meet face-to-face. In this light, [17] evaluated a cooperative learning system which crowd-sourced learner explanations. After answering an assessment question, learners were prompted to give an explanation/justification. These explanations were then accumulated and shared with their peers; the authors found that providing learners the explanations of their peers increased the likelihood of a learner revising their answer to the correct one. In this scenario, the prompting for explanations not only serves as a reflective activity for the individual learner, it also leverages the social aspect by allowing him or her to contribute to the larger course community and potentially help a peer in need.

[46] investigated the effectiveness of co-explanation (where learners are instructed to collaboratively explain worked examples) as compared to self-explanation (where learners work alone) in a Design Principles course context. The authors found that learners in the co-explanation condition were not as engaged with the assessment questions (identifying a design's strengths and weaknesses) as their counterparts in the self-explanation condition.

In the domain of peer review, [199] compared "free-selection" peer review (where students could freely choose which of their peers' work to review) against an "assigned-pair" design (where the peer review pairings are assigned by the instructor) in the context of a computer networking course. The authors found that students in the free-selection group achieved greater learning outcomes and provided better reviews than those in the assigned-pair group [199].

[150] developed a recommender system within a MOOC to provide each learner with a list of peers in the same course who they would likely work well with based on profile similarity modeling. Compared to the control group with no recommendations, the experimental group (receiving the list of peer recommendations) displayed significantly improved persistence, completion rate and engagement.

Given the consistency of positive results in this category (71%—the highest of any category), the above studies should be used as building blocks or inspiration for future work in finding new ways to bring learners together and increase their sense of community and belonging in the digital learning environment. Advances made in this vein would work towards harnessing the true power of large-scale open learning environments where learners not only learn from the instructor but from each other as well through meaningful interactions throughout the learning experience [232].

Table 2.9: Cooperative Learning

Cooperative Learning: + : 17 / ○ : 6 / − : 1				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Bhatnagar et al. [17]	144	Native	Class	+Revision
Lan et al. [152]	54	Native	Class	+Learning Gain
Tritz et al. [248]	396	Native	Class	+Final grade
Ngoon et al. [191]	75	Native	$	+Learning Gain
Kulkarni et al. [148]	104	Native	None	+Exam Score ○Revision
Kulkarni et al. [147]	2,422	MOOC	None	+Exam Score
Cambre et al. [37]	2,474	MOOC	None	+Exam Score
Culbertson et al. [58]	42	Native	$/Credit	+Learning Gain
Coetzee et al. [51]	1,334	MTurk	$	+Exam Scores
Zheng et al. [275]	1,730	MOOC	None	○Engagement ○Persistence ○Final Grade
Khandaker and Soh [123]	145	Native	Class	○Exam Score
Konert et al. [145]	396	Class	None	+Engagement +Persistence
Labarthe et al. [150]	8,673	MOOC	None	+Persistence +Completion Rate +Engagement
AbuSeileek [2]	216	Lab	n/r	+Final Grade
Papadopoulos et al. [199]	54	Native	Class	+Final Grade
Chang et al. [40]	27	LMS	Class	+Final Grade
Cho and Lee [46]	120	Class	None	−Engagement ○Learning Gain

2.4.8 Simulations & Gaming

[96] categorizes simulations and games together and defines them as a simplified model of social or physical reality in which learners compete against either each other or themselves to attain certain outcomes. He also notes the subtle difference between simulations and gaming in that simulations are not always competitive. The studies in this category are carried out predominantly in native environments. While understandable given the games could have been developed for purposes other than experimentation, this raises potential issues with an eye towards reproducibility. However, considering 19 of the 28 reported results (68%) in the category pertain to desirable benefits in learner achievement or behavior, this also indicates a very strong trend towards the generalizable effectiveness of using simulations and gamification to help learners.

While each game or simulation is unique in its own right, the underpinning theme in all of these studies is as follows: the learner earns and accumulates rewards by exhibiting desirable behavior as defined by the instructor/designer. While creating native educational games or gamifying existing learning environments (especially MOOCs as in [50]) is a complex, time-consuming process, based on the predominantly positive findings in the literature, we conclude that it is an area with high potential for boosting learning performance.

[62] ran a study evaluating the effect of choosing versus receiving feedback in a game-based assignment. Compared to the group which passively received feedback, the group which was forced to actively retrieve feedback was more engaged with the environment, but showed no difference in learning or revision behavior. Note that this study is not in the Feedback category, as both cohorts received the same feedback; the element being tested in the experiment was the manner in which it was delivered within the simulated environment.

[35] employed a serious game design and put students either in a competitive (showing a scoreboard and ranking of peer performance) or non-competitive environment. The authors found that the competitive environment led to significantly higher test scores and more time spent answering questions. On a similar note, [9] evaluated the effect of implementing a points system within a computer-based mathematics learning environment. Although participants in the conditions with the points system answered questions faster, there was no effect on the accuracy of their responses.

[101] created a formative assessment game for a computer programming learning task. The authors found that participants in a traditional, non-computer-based environment performed worse on problem solving tasks than those who received the computer-based formative assessment system. [14] also compared a game-based learning environment to a non-computer-based experience. In their experiment, the participants in the game-based learning condition displayed better scores on a post-test.

With 68% of the reported results being positive findings—the second highest among all categories—we see great potential for the effectiveness of learning experiences where learners are afforded the ability to interact with and explore simulated environments. Due to the substantial cost of developing such environments, future research is needed to evaluate whether this trend of positive findings continues so that institutions can be assured in justifying their investment in these instructional strategies.

Table 2.10: Simulations & Gaming

Simulations & Gaming: + : 19 / ○ : 9				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Bumbacher et al. [34]	36	Native	n/r	+Learning Gain
Cox et al. [55]	41	Native	$	+Engagement
Culbertson et al. [58]	42	Native	$/Credit	+Engagement +Exam Score
Ibanez et al. [106]	22	Native	Class	+Exam Score
Krause et al. [146]	206	Native	n/r	+Persistence +Exam Score
Li et al. [163]	24	Native	n/r	○Efficiency +Learning Transfer
Schneider et al. [222]	82	Native	Class	+Learning Gain +Engagement
Cutumisu and Schwartz [62]	264	Mturk	$	○Engagement ○Learning Gain
Coetzee et al. [50]	1,101	MOOC	None	○Final Grade ○Persistence +Forum Participation
Cheng et al. [45]	68	Native	n/r	○Learning Gain
Pozo-Barajas et al. [208]	194	Native	n/r	+Final Grade
Smith et al. [235]	57	Native	Class	+Learning Gain
Brom et al. [27]	75	Lab	$ or Credit	○Final Grade +Engagement
Attali and Arieli-Attali [9]	1,218	Mturk	$	○Final Grade ○Efficiency
Cagiltay et al. [35]	142	Native	n/r	+Final Grade
Hooshyar et al. [101]	52	Native	Class	+Learning Gain
Nebel et al. [190]	103	Native	$ or Credit	+Final Grade +Efficiency
Barr [14]	72	Lab	n/r	+Final Grade

2.4.9 Programmed Instruction

According to [96], programmed instruction is a method of presenting new subject matter to students in a graded sequence of controlled steps. Its main purposes are to (i) manage learning under controlled conditions and (ii) promote learning at the pace of the individual learner.

Programmed instruction is inherently adaptive–it presents material to the learner according to that learners unique set of previous actions. As they stand now, MOOCs are simply online course content resources that remain static irrespective of a learners behavior. Unlike the native and lab environments used in [26] and [115], the current MOOC technology has not yet accounted for a learners past behavior in delivering personalized content accordingly. By developing and implementing these types of systems in a MOOC, MOOCs could then become more adaptable and able to cater instruction based on the individual learner. To enable this would require a real-time tracking system for learners where their behavior could be collected, modeled/analyzed, and then acted upon (e.g., with the delivery of a personalized recommendation for a next activity or resource) in real time on a large scale.

Table 2.11: Programmed Instruction

Programmed Instruction: + : 4 / ○ : 7				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Brinton et al. [26]	43	Native	None	+Engagement
Karakostas and Demetriadis [115]	76	Lab	n/r	+Learning Gain
Rosen et al. [215]	562	MOOC	None	+Learning Gain +Persistence ○Final Grade
Zhou et al. [276]	153	ITS	Class	○Final Grade ○Engagement
Arawjo et al. [7]	24	Native	n/r	○Engagement ○Completion Rate ○Learning Transfer
van Gog [253]	32	Lab	n/r	○Final Grade

2.4.10 Interactive Multimedia Methods

As lecture videos are currently the backbone of MOOC instructional content, it is imperative that they effectively impart knowledge to learners in an engaging, understandable fashion. Also among the most effective strategies with 64% of reported results being positive, interactive multimedia methods, though not limited to video, test various methods of content delivery through multimedia application interfaces.

[136], for example, compared lecture videos which included a small overlay of the instructor's face talking versus the same lecture videos without the overlay. Results show that while learners preferred videos showing face and perceived it as more educational, there were no significant differences in the groups' exam scores.

A more interactive approach to lecture videos was explored by [184] who integrated several interactive components into lecture videos by integrating quiz, annotation, and discussion activities within a video player. Compared to a baseline interface, which separates the videos and assessments, the integrated interface was favored by learners and enabled them to learn more content in a shorter period of time.

Looking beyond video delivery methods, [149] compare four delivery methods of a tutorial on the topic of data visualization. The four conditions are: (i) a baseline which only included text, (ii) baseline plus static images, (iii) video tutorial, and (iv) interactive tutorial where learners worked with a Web interface to manipulate and create their own data visualizations. The authors found that learners with the interactive tutorial performed better on the exam and did so while spending less overall time in the platform — an indication of increased efficiency.

[5] evaluated the effectiveness of captioned animation with keyword annotation (a note explaining the meaning of a given word) in multimedia listening activities for language learning. Compared to participants who received either just animations or animations with captions, the captioned animations with keyword annotation condition performed significantly better on recognition and vocabulary tests. However, the participants just receiving the animations significantly outperformed the other conditions on listening comprehension and recall over time.

[179] deployed a “here and now" learning strategy (where learners have 24/7 access to learning activities on their mobile phones) to compare its effectiveness against computer-based instruction. While the “here and now" conditions expressed more positive attitudes towards the learning experience after the experiment, the computer-based learning cohort earned higher scores on a post-test.

[165] tested the impact of modality (text vs. audio+text) on learning outcomes. They found that the multimodal format (audio+text) led to better learning outcomes than receiving text alone. [206] ran a study to see the effect of compressing the time of instruction (decreasing time to train/learn the materials) on learning. They found that decreasing (accelerating) the time by 25% leads to similar learning outcomes, whereas decreasing by 50% causes a decrease in learning.

Pursuing new research in this category is important going forward in trying to truly leverage the Web for all of its learning affordances. The possibilities for digital interfaces, sensors, and devices are expanding rapidly, and more immersive, interactive, and intelligent environments promise to make a significant impact on online learning environments in the future. Even before these exciting technologies have become widely explored, we still observe an encouraging trend in this category in terms of positive results reported; we therefore recommend future research continue to explore the new possibilities in highly dynamic, interactive learning environments.

Table 2.12: Interactive Multimedia Methods

Interactive Multimedia Methods: + : 28 / ○ : 12 / − : 4				
Ref.	**N**	**Env.**	**Incentive**	**Result**
Lee et al. [159]	102	Native	n/r	+Learning Gain
Monserrat et al. [183]	15	Native	None	+Efficiency
Monserrat et al. [184]	18	Native	None	+Learning Gain
Nicholson et al. [194]	40	Native	None	+Efficiency
Trusty and Truong [249]	21	Native	$	○Learning Gain ○Engagement
Dearman and Truong [73]	15	Mobile	$	+Learning Gain
Kizilcec et al. [136]	22	Lab	Credit	○Exam Score
Kwon and Lee [149]	120	Mturk	$	+Exam Score +Efficiency
Kizilcec et al. [134]	104	Native	Credit	○Final Grade
Zhu et al. [277]	22	Native	n/r	+Learning Gain +Engagement
Pandey et al. [198]	44	Native	n/r	○Engagement - Final Grade
Culbertson et al. [59]	27	Native	$	○Larning Gain +Engagement
Austin [11]	75	Native	Credit	+Learning Transfer
Yamada [268]	40	Native	n/r	+Engagement ○Revision
Wang et al. [258]	123	Lab	n/r	+Final Grade
Pastore [206]	154	LMS	n/r	+Efficiency
Chen et al. [43]	81	LMS	n/r	+Final Grade
Chuang and Tsao [48]	111	Mobile	n/r	+Learning Gain ○Long Term Retention
AbuSeileek and Qatawneh [3]	30	Native	Class	−Engagement
Imhof et al. [107]	71	Native	$ or Credit	−Final Grade
Urquiza-Fuentes and Velázquez [251]	132	Native	Credit	○Final Grade ○Long Term Retention +Completion Rate
Aldera and Mohsen [5]	50	Native	Class	+Final Grade +Long Term Retention
Martin and Ertzberger [179]	109	Mobile	Class	−Final Grade
Chen and Wu [41]	37	Native	n/r	+Learning Gain
Song et al. [237]	144	Native	None	+Learning Transfer +Engagement
van Gog et al. [254]	43	Lab	$ or Credit	+Final Grade
Limperos et al. [165]	259	Lab	None	+Final Grade
Türkay [250]	621	Mturk	$	○Persistence +Final Grade
Jang et al. [108]	76	Native	n/r	+Final Grade
Jeno et al. [109]	71	Mobile	$	+Final Grade
van Wermeskerken and van Gog [255]	69	Lab	$ or Credit	○Final Grade
Sharma et al. [227]	n/r	MOOC	None	+Video Engagement

2.5 Conclusion

Based on both the quantitative and qualitative analyses from this review, we identify Cooperative Learning, Simulations & Gaming, and Interactive Multimedia as the three most promising strategies for most effectively activating learning at scale. We draw this conclusion from the proportion of positive results from each category: 71% for Cooperative Learning, 68% for Simulations & Gaming, and 64% for Interactive Multimedia—compared to all other groups with more than one study which have an average of 43% positive results.

According to Hattie's meta-analysis including over 50,000 studies, the ten learning strategies shown in Table 2.2 are among the most effective. And yet, in so many instances do we here find null results for the studies employing them. Based on Hattie's work and sheer volume of studies included, we cannot yet dismiss the strategies themselves as ineffective; rather, translating them to the digital age of scalable learning has emerged as the primary challenge for the future. We identify a key factor in meeting this challenge to be that of incentive. Compared to an even balance of positive and null findings in experiments without any incentive for the participant (49% + and 49% ∘), positive results are twice as likely as null results in experiments which provide a monetary incentive. Given that the main application area for the present review is concerned with self-directed, informal online learning environments such as MOOCs, we see this discrepancy as one that calls for thorough investigation in future research. This trend suggests that the applicability and effectiveness of instructional and/or learning strategies could potentially be context-dependent—that the same intervention might be highly effective in a context where participants have a monetary incentive and ineffective in a context where participants are rewarded with course credit or are intrinsically motivated, for example. This bears significant implications for the generalizability of online learning research in that (i) researchers must take great care in contextualizing findings and (ii) readers must be attentive in interpreting results and mindful of the study's full context.

We are in the beginning stages now of constructing this new narrative of pedagogy at scale, and would be naive to think this could have been perfected in just three years. Guided by the proven efficacy of Cooperative Learning, Simulations & Gaming, and Interactive Multimedia learning strategies, the community should now work through iterative cycles of designing, testing, and evaluating new solutions in formalizing this emergent body of theory and literature.

Part II

Teaching & Learning Paths

This part serves **RQ2** (*How can MOOC environments be improved to advance the possibilities of experimentation?*) by focusing on teaching and learning paths. The concept of a path is born out of the design of the edX platform as it has stood since its creation. In the platform, courses can only be structured in a linear fashion—that is, each activity/component must be built and ordered sequentially after another. The platform does not allow for multiple branches of a learning path, for example, that would enable the learner to choose from a selection of pre-determined (by the instructor or course designer) pathways.

While this approach to instruction does not encourage a great degree of exploration from the learner, it is still very much possible for learners to create their own trajectories through the course. Even though the course is designed and delivered in a linear fashion, the learners are still free to deviate from this prescribed learning path and navigate the course at their discretion. For learners who are highly effective at self-regulating their own learning processes, this may turn out to be beneficial, as they might benefit from the flexibility and find the pathway through the course that best serves their own interests and goals. However, given that MOOCs are intended for those without access to high-quality education from disadvantaged populations, one cannot operate under the assumption that such deviations from the prescribed learning path will always have resulted from a thoughtful, effective, intentional decision from the learner.

This line of inquiry gave rise to **Chapter 3**, which explores the extent to which learners adhere to (or deviate from) the designed learning path, and the subsequent effect that such deviations have on a learner's eventual course outcome. We find that there is indeed a substantial amount of deviation from the designed learning path that occurs in the four courses considered in the study. The key finding from this line of inquiry is that there is a strong relationship between the linearity of a learner's path through the course and their likelihood of earning a passing grade—learners with fewer deviations (a more linear path) are more likely to pass the course. The next key contribution is the identification of behavioral *motifs*, or common strings/patterns of behavior carried out by learners across the four courses. We then divided these motifs by passing and non-passing learners to uncover which motifs were common to the most successful learners in the course.

This study on learning pathways gave rise to the following study which offers a deeper dive into the designed learning path of MOOCs. In **Chapter 4**, based on the insights gained from Chapter 2 that highlight the importance and value of taking the course context into account when analyzing learner behavior, we take a deep dive into understanding trends in the design of learning trajectories in MOOCs. By applying methodology and theory from the learning design literature, we encode the course design of 177 MOOCs from two institutions, computationally identify similarly-designed courses, and explore the relationship between course designs and learning outcomes. This work serves as a key step in the movement towards understanding the causal effect of a designed learning path on eventual learner achievement.

In summary, this part contributes a new understanding of the way MOOCs are designed and the patterns by which learners engage with them.

Chapter 3

Gauging MOOC Learners' Adherence to the Designed Learning Path

Massive Open Online Course platform designs, such as those of edX and Coursera, afford linear learning sequences by building scaffolded knowledge from activity to activity and from week to week. We consider those sequences to be the courses' *designed* learning paths. But do learners actually adhere to these designed paths, or do they forge their own ways through the MOOCs? What are the implications of either following or not following the designed paths? Existing research has greatly emphasized, and succeeded in, automatically predicting MOOC learner success and learner dropout based on behavior patterns derived from MOOC learners' data traces. However, those predictions do not directly translate into practicable information for course designers & instructors aiming to improve engagement and retention — the two major issues plaguing today's MOOCs. In this work, we present a three-pronged approach to exploring MOOC data for novel learning path insights, thus enabling course instructors & designers to adapt a course's design based on empirical evidence.

This chapter is published as "Gauging MOOC Learners' Adherence to the Designed Learning Environment" [63], by D. Davis, G. Chen, C. Hauff, and G.J. Houben in Proceedings of the Ninth International Conference on Educational Data Mining, 2016.

3.1 Introduction

MOOCs can deliver a world-class education on virtually any academic or professional development topic to any person with access to the Internet. Millions of people around the globe have signed up to courses offered on platforms such as edX, Coursera, FutureLearn and Udacity. At the same time though, only a very small percentage of these learners actually complete a MOOC successfully [144], an issue that continues to plague massive open online learning. Keeping MOOC learners engaged and improving the dismal retention rates are major concerns to instructional designers and MOOC instructors alike. Considerable research efforts have been dedicated to the automatic prediction of learners' (imminent) dropout in MOOCs, e.g. [95, 138, 197, 269], under the assumption that once learners under the threat of attrition are identified, an automated intervention can be staged to (re)engage those learners with the course material. While the accuracy of these usually machine-learning-based predictors is high, their *explanatory power* is often low. Model features that have the strongest impact on prediction purely based on statistical grounds may not provide course designers & instructors with enough information to adapt the design or content of a MOOC in response.

In this work we aim to provide a more holistic view of learners' progression through a MOOC in order to enable more practicable insights to instructors and designers. Our approach to educational data mining as presented here is a very literal realization of Graesser's vision for the field by illustrating and "*look[ing] at unique learning trajectories of individuals*" [239]. We make use of the concept of *learning paths* (a learner's route through course activities) and investigate how the learning paths of successful and unsuccessful MOOC learners differ.

The design of MOOCs on the edX platform[2] implies a *linear* trajectory through the learning material. Most courses are broken up into weeks (Week 1, Week 2, etc.) and released one week at a time. Within these weeks, the standard instructional approach is to first provide a brief introduction to the week's material, followed by the weekly video lectures (the main source of content delivery), then the assessments that evaluate learners' knowledge of the preceding video lectures, and, finally, courses may offer bonus material. This cycle is repeated each course week (and sometimes multiple cycles comprise a single week). But do learners *actually* adhere to this cycle, and thus the *designed* learning path? Does it matter if they do not? These are the central issues that we focus on in this paper. While the concept of *executed*

[2]Our empirical work is based on edX MOOCs, but the same principles apply to other major MOOC platforms.

learning paths (i.e., the paths students actually take through a course) has received substantial attention in the e-learning and intelligent tutoring communities [139, 226], in the MOOC setting this concept has so far garnered little attention. First empirical evidence that learners do not always follow the designed sequence through a MOOC has been observed in [94], however, to our knowledge no in-depth investigation of this phenomenon in the MOOC context exists as of yet. We aim to close this knowledge gap and investigate the following research question:

RQ3.1 To what extent do learners adhere to the designed learning path set forth by the instructor?

We develop three approaches to characterize learning paths, thus providing three different views on a MOOC's *designed learning path* (created by the course instructor or designer) and the *executed paths* (created by the learners of the MOOC). We apply our approaches on the log traces of more than $113,000$ learners who participated in one of four edX-based MOOCs in the domains of computer science, political debates and business ethics.
We show that (1) our approaches shed light on the deviations between designed and executed learning paths, and, (2) successful and unsuccessful learners differ considerably in the paths they follow. We believe that our work can provide instructional designers a valuable analysis tool to improve the design of both online courses and MOOC platforms in the future as they provide data-driven insights into the actual behavior of learners and the impact of their behaviors on learning outcomes.

3.2 Related Work

In this section, we elaborate on existing research in learner modeling [76], focusing on works that investigate learning activity sequences and their impact on learning outcomes.
The problem solving behavior of learners in the context of e-learning and intelligent tutoring systems has been explored in [110, 139, 140, 226]. In contrast to our work, which considers a range of activities learners perform throughout a course (and compares them to the designed learning path), these works have explored learners' exhibited behavior within only one activity type — problem solving. Specifically, Köck and Paramythis [140] performed activity sequence clustering (an application of sequential pattern mining [240]) to model the learners' behavior, while in [139] automated clustering

and human synthesis of the generated clusters were combined to identify patterns of problem solving. Shanabrook et al. [226] introduced a semi-automatic approach to identify a student's state while problem solving (including: gaming the system, guessing out of frustration, abusing hints, being on-task) in a high school-level intelligent tutoring system employing sequence-based motif discovery. Jeong and Biswas [110] developed a Hidden Markov Model to describe how different middle school student behavior trends lead to different learning processes & outcomes when problem solving.

In the context of MOOCs, sequences of learning activities have been explored by Wen and Rosé [260], who investigated the most common two-step activity sequences learners exhibit across two MOOCs. These patterns were then manually checked and analysed for interesting learning habits. A similar analysis of two-step chains was performed in Guo and Reinecke [94] who found that learners generally progress through the course content in a non-linear, "exploratory" manner [164]. Guo and Reinecke [94]'s observation of learners frequently performing "backjumps" (moving from a quiz to a lecture video previously introduced) can be considered as one of the first comparisons of executed and designed learning paths in MOOCs. Kizilcec et al. [131] (replicated in [81]) have also taken steps in this direction, by utilizing the assessment submission times (either on track, late or never) in MOOCs as indicators of learner engagement groups (completing, auditing, disengaging or sampling learners). Our work can be considered a significant expansion to these approaches, as we explore longer activity sequences (eight-step chains), thus enabling the discovery of more high-level and complex patterns and making designed vs. executed paths the focal point of our investigation.

Video interactions in MOOCs were the focus of Sinha et al. [234], who categorized the most prominent chains of video interactions (pause, play, speed, and skipping) and analyzed them with respect to learner dropout.MOOC discussion patterns have been investigated by Brooks et al. [29] who found that MOOC students exhibit markedly different discussion patterns than were expected based on blended learning environments. This finding can also be considered as a motivation for our work; MOOCs may not always be used by learners the way the instructors or course designers intended.

The concepts of process mining and conformance checking, in particular, are also employed in areas such as business process execution; [218] explains how business processes can be monitored (process mining) and then compared to the intended model (conformance checking) via a measure of fitness.

3.3 Subjects & Data

Table 3.1: Overview of the MOOCs in our study. The #Chains column contains the number of events observed throughout the MOOC (cf. Table 3.2). The “Passing Grade" shows the percentage of quiz questions to answer correctly to receive a course certificate. “Tries” indicates how many attempts a learner has per question. “Videos Accessed" shows the average % of course videos watched by certificate-earning learners. “Missing" is the % of certificate-earning learners who streamed zero video lectures.

	N	Pass Rate	Chains Pass/Non	Wks	Vids.	Quiz Q's	Pass. Grade	Tries	Vids. Access	Miss.
FP	37,485	5.3%	1.06M/807k	14	41	288	60%	1	67.5%	4.3%
RI	8,850	4.3%	66k/30k	7	47	75	59%	1-3	49.7%	19.6%
FR	34,017	2.4%	95k/141k	6	55	26	50%	2	51%	3.8%
EX	33,515	6.5%	1.02M/855k	8	59	136	60%	2	76.3%[3]	3.6%

We explore our research question in the context of four MOOCs: `Functional Programming` (FP) (teaching the functional programming paradigm), `Data Analysis` (EX) (teaching spreadsheet and basic Python skills for data analysis), `Framing` (FR) (the art of political debates), and `BusinessX` (RI) (a MOOC on the ethics and safety of new technologies). All MOOCs were offered on the edX platform in 2014/2015 and designed as xMOOCs.

Overview of MOOCs Table 3.1 provides an overview of the four MOOCs in this study. The learner enrollment varies between ≈9k and ≈37k. While the four MOOCs are comparable in their video material offerings (between 41 and 59 videos), they differ significantly in the number of summative assessment questions (between 26 and 288 quiz questions). We also observe considerable differences in the percentage of video material watched by certificate-earning learners (replicating [94]) — less than half of the videos are accessed by successful learners in `Data Analysis`, while more than two thirds of the videos are accessed by successful learners in `Functional Programming`. Lastly, we note that the `BusinessX` MOOC is an outlier with respect to the percentage of learners that passed the course *without* streaming any video material,[4] with nearly 20% of successful learners falling into this category; the same applies for only ≈4% of learners in the other three MOOCs.

Translating Log Traces into a Semantic Event Space The edX platform provides a great deal of timestamped log traces, including clicks, views,

[3]This number was incorrectly reported in the published version of the paper as 45%.
[4]Note that the log traces did not capture video downloads and subsequent offline watching.

quiz attempts, and forum interactions. We adapted the `MOOCdb`[5] toolkit to our needs and translated these low-level log traces into a data schema that is easily query-able.

Table 3.2: Overview of events considered in this work.

Video	**Quiz**	**Progress**	**Forum**
WATCH	START	VIEW	START
	SUBMIT		SUBMIT
	END		END

For this work, we focus on four event types as listed in Table 3.2: events related to videos, quizzes, progress pages, and discussion forums. Videos can be watched - this event is generated whenever a user clicks the video ‘play’ button. Quizzes are identified through the beginning of the quiz session (the user enters the quiz page), the submission of one or more answers[6], and the ending of the quiz session (the user leaves the quiz page). Those quizzes are typically summative in nature. If a user views his or her progress page, the `VIEW` event is elicited. Finally, we condense discussion forum events into three kinds of items: the start of a forum session (the user first enters the forum), the submission of content (question, comment or reply) and the end of the forum session (the user leaves the forum page).
All *executed* learning paths that we extract from the learner log traces consist of the events listed in Table 3.2. The rationale for choosing these events comes from the designed learning path by which xMOOCs are typically formed: first watch one or more lecture videos, and then move on towards the quiz and/or forum section for assessment and knowledge building & verification respectively. In Figure 3.2 we visualize a week’s designed learning path for each of the four MOOCs we study (this pattern is repeated in every course week). Video lectures form a common denominator, starting the path. `Functional Programming` and `Data Analysis` rely on videos and quizzes only (with `Data Analysis` exhibiting multiple video-quiz “cycles” within a week), whereas `BusinessX` and `Framing` make use of the forums as well. The learning path shown for `Framing` does not include quizzes as they are posed only in the final week (in the form of an exam).

[5]http://moocdb.csail.mit.edu/

[6]Note that on the edX platform answers to individual quiz questions are submitted (instead of all answers at once).

3.4 Approach

Having introduced the subjects of our work and the events we consider, we now describe the three distinct approaches to the visualization & exploration of executed learning paths (that is, learners' sequential movement *over time* through the activities offered in a MOOC) we developed.

3.4.1 Video Interactions

As shown in Figure 3.2, videos are a focal point of xMOOCs. Accordingly, in a first analysis, we focus exclusively on video interactions and explore to what extent learners adhere to the designed video watching learning path. Therefore, in this study we only make use of `WATCH` events.

We transform the `WATCH` events generated by a set of learners L across the duration of a MOOC $\mathcal{M}$ into a directed graph $G_{\mathcal{M},L} = (V_{\mathcal{M}}, E_{\mathcal{M},L})$ — as the subscripts indicate, with $\mathcal{M}$ fixed, the set V is independent of the subset of learners chosen, while E is dependent on the learners in L. All lecture videos contained in $\mathcal{M}$ form the set of vertices $V_{\mathcal{M}}$. The vertices are labelled chronologically, that is, for any vertex pair (v_i, v_j) with $i < j$, the corresponding lecture video i must appear in the designed learning path before video j.

The edges are directed and weighted according to the number of `WATCH` events by the learners L: an edge between v_{i-1} (source) and v_i (target) presents the learners' transition between these videos, i.e. the number of times learners watching video v_{i-1} watch v_i next, before any other video. We disregard self-loops (watching the same video again) as we are focusing on the progression of the learners through the set of lecture videos.

Having generated $G_{\mathcal{M},L}$, we now turn to its visualization (to aid instructors and course designers): the vertex layout is sequential and governed by the designed learning path through the videos (represented as vertices). For MOOCs with thousands of participants it is likely that every single video pair combination possible is contained in at least one learning path. To avoid visual clutter, we filter out the most *infrequent* edges: we bin the edges according to the week their source vertex appears in and remove the 10% of edges that occur most infrequently in this course week.

To discover whether or not there are marked differences in the way different groups of learners behave, we generate the video interaction graph for different sets of learners, such as successful (certificate earning) vs. unsuccessful learners.

3.4.2 Behavior Pattern Chains

Having considered the transitions between lecture videos, we now turn to the exploration of transition patterns among all eight events identified in Table 3.2. Previous works [260] have viewed MOOC learner patterns either in terms of one-step directed pairs of events (such as *watch video* → *begin quiz*) or based on video click chains only [234].
One-step chains can only provide limited insights into more high-level behavioral patterns — we may, for instance, be interested to understand how many learners are "binge watchers" (watching many videos in a row) or "strategic learners" (looking at quiz questions before watching the corresponding lecture video). In order to contribute insights to our research question we need to consider longer chains. We have settled on eight-step chains, as they provide insights into more high-level concepts but are still numerous enough in our log traces to make claims about their general usage. We consider all events of Table 3.2 and create event chains by sliding a window of size eight over each learner's chronologically ordered learning path through a MOOC. An example eight-step chain this procedure yields is shown in Figure 3.1.

$$\mathtt{QUIZ}_{START} \rightarrow \mathtt{QUIZ}_{END} \rightarrow \mathtt{WATCH} \rightarrow \mathtt{WATCH}$$

$$\rightarrow \mathtt{WATCH} \rightarrow \mathtt{WATCH} \rightarrow \mathtt{WATCH} \rightarrow \mathtt{WATCH}$$

Figure 3.1: An example eight-step chain.

To identify the underlying trends in the chains, we employed the *open card sort* approach [83]. After printing out two sets of the thirty most frequently occurring chains on paper, two authors independently sorted them into (non-predefined) like-groups by hand and afterwards discuss the differences in each sort, creating a composite of the two results. The outcome of this method is a synthesis of similar chain types into groups sharing the same *motif*, or recurring theme. Based on the motifs, we created a rule-based system that assigned a MOOC's entire set of chains to the identified motifs (chains that do not fit into any motif are left "unassigned"). This process is repeated for each of the MOOCs we investigate. The advantage of this approach over the automatic clustering of the chains is the infusion of our domain knowledge into the clustering process.

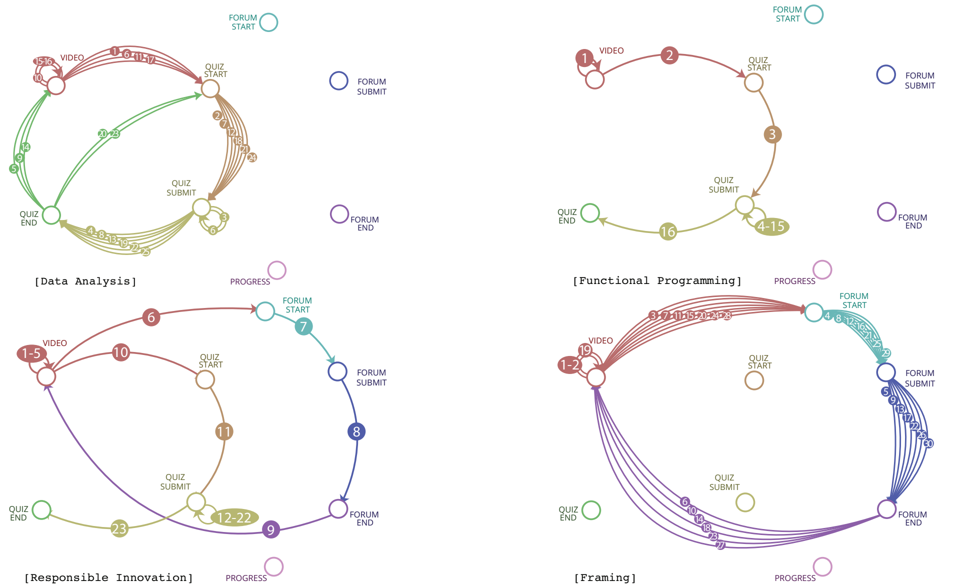


Figure 3.2: The designed learning path for a standard week (Week 4) of each MOOC. The circled numbers indicate the step number of each transition in that week's sequence. Notice the diversity in course designs that characterize these four MOOCs.

3.4.3 Event Type Transitions

Lastly, we explore event type transitions, or how likely learners are to move from one event type to another. Inspired by the methods employed in [110, 139, 140] we use discrete-time Markov chains (a memory-less state transitioning process encoding how often learners move from one event type to another) in order to chart the likelihood that a learner will transition from one engagement activity to another. Whereas the prior works employ these methods in the context of problem solving (knowledge assessment), we focus on the larger process of *knowledge building*, which transpires over the span of an entire course. While it may be self-evident that non-passing learners answer less quiz questions than their certificate-earning peers (and thus the transition probabilities to $\texttt{SUBMIT}_{QUIZ}$ are likely to be lower for non-passers), the visualization of the Markov chains enables designers to pinpoint the differences in transitions between different types of learners (e.g. passers vs. non-passers) across all events in one coherent plot.

3.5 Findings

To answer our research question (do learners adhere to the designed learning path?), we apply the three approaches outlined in Section 3.4 to the datasets described in Section 3.3.

3.5.1 Video Interactions

We visualize the video interactions across the first three weeks (these are where the most deviations occur; the later weeks are more in line with the designed path) of each MOOC in Figures 3.3 to 3.6, distinguishing two sets of learners: those that eventually earn a certificate ("Passing") and those that do not ("Non-Passing"). The *designed* video interaction learning path is exhibited by the left-to-right flow of the vertices (one per video). The edges correspond to the *executed* learning paths — with edge thickness indicating the (normalized) number of learners having taken that path (only the 90% most frequently occurring transitions each week are shown); the set of red edges represent the executed transitions that follow the designed transitions.
A number of observations can be made based on the visualizations: (i) passing learners deviate considerably less from the designed learning path than non-passing learners across all four MOOCs, (ii) passing learners are more likely to skip video lectures introducing the platform (the first three videos in the `Framing` MOOC) than non-passing learners, indicating a higher level of seniority in MOOC-taking, (iii) towards the end of week three, the deviations among the sets of passing and non-passing learners are negligible (i.e. the non-passing learners still active exhibit a similar video watching behavior as the passers), and (iv) skipping videos — jumping ahead — is much more common than backtracking — jumping backwards — for both passers and non-passers.
An emerging object in the field of Design (and gaining some attention in the field of Software Design [57]) is that of *desire paths*, or paths not intended by the designer, but those which "arise due to off-[path] use ... for a variety of purposes such as access to places of interest and shortcutting" [25]. This research serves as a reminder that desire paths indeed exist in MOOCs (as evident in the skipping of introductory lecture material) — they just have not yet been made as visible as those brown stripes of beaten grass and dirt transecting public parks and trails. They are a reminder that humans can collectively communicate good design by their actions.

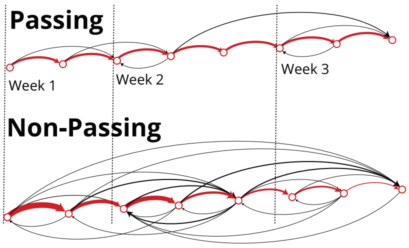


Figure 3.3: `Functional Programming` video interactions.

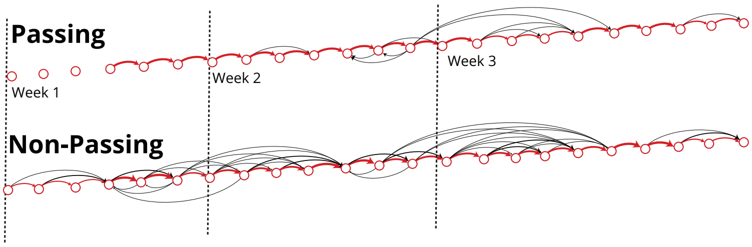


Figure 3.4: `Framing` video interactions.

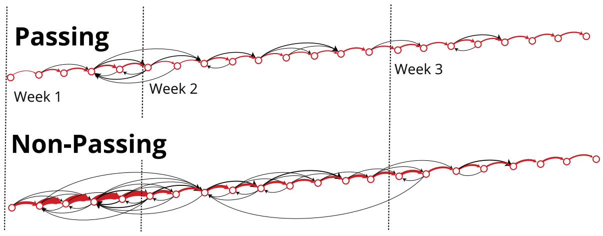


Figure 3.5: `Data Analysis` video interactions.

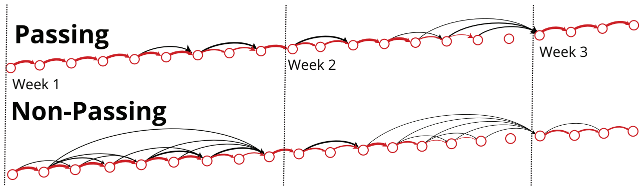


Figure 3.6: `BusinessX` video interactions.

3.5.2 Behavior Pattern Chains

Our second approach explores learners' behavioral patterns. As outlined in Section 3.4.2, we first manually clustered and labelled the most frequent eight-step pattern chains in order to determine what type of behaviors (or motifs) learners exhibit beyond a single-click transition, before automatically assigning the remaining chains into those motifs. Depending on the MOOC, this approach yielded between eight and 11 motifs, with some motifs appearing only in a subset of courses. For brevity reasons, in Tables 3.3 to 3.6 for

each MOOC we list its most frequent motifs (specifically those into which ≥2% of all chains are classified); as a comparison in Table 3.1 we also list the total number of chains generated by passing/non-passing learners in each MOOC — depending on the MOOC, the listed motifs capture between 42%–77% of the total number of chains. Whenever a motif is first introduced, we briefly describe which event types and event orderings characterize it[7].
Examining the results, we observe that (i) *Binge Watching* is a frequent motif in all MOOCs with non-passers always exhibiting more binge watching (i.e. watching videos uninterrupted by other activities) than passers, (ii) the *Lecture→Quiz Complete* motif, which captures the "classic" xMOOC idea of video watching with subsequent question answering is frequent in three of the four MOOCs[8], however no consistent divergent behavior for passers and non-passers is found, (iii) motifs with forum events occur in three of the four MOOCs — by course design in `Framing` and `BusinessX` (cf. Figure 3.2), but not in `Functional Programming`, indicating issues related to material clarity, and (iv) the *Quiz Check* motif, which is exhibited by learners checking the quiz questions without answering any of them (which is usually followed by video watching and subsequent quiz completion), is only found in one MOOC frequently; in `Data Analysis` 2% of the chains follow this motif, a smaller percentage than we expected, indicating that very few learners are gaming the system by "attempting to succeed in an educational environment by exploiting properties (quiz questions are posted alongside the video material) of the system (edX platform) rather than by learning the material and trying to use that knowledge to answer correctly," [12].

3.5.3 Event Type Transitions

The Markov models of our four MOOCs are visualized in Figures 3.7 to 3.10. Since we observe the same event types across the four MOOCs, the set of vertices, their placement in the visualization, and their semantics are identical. To minimize visual clutter, we only plot the transitions (i.e. the edges) that exhibit a probability of 0.2 or higher. Once more we make the distinction between passing and non-passing learners. The resulting visualizations show the behavioral differences not only between passing and failing students within a given course, but these also allow for cross-course analyses which shed light on what types of behavioral patterns define a course. For

[7]Note, that we implemented our rules for the automatic assignment of chains to motifs according to these characterizations.

[8]It does not appear among the frequent motifs in `Framing`, which has a final exam instead of weekly quizzes.

Table 3.3: Most frequent motifs (≥2% chains) in `Functional Programming`.

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Quiz Complete	552,363 (29.4%)	328,995 (30.8%)	223,368 (27.7%)
	$\texttt{X}_{\texttt{QUIZ}}$ events only with at least one X = *SUBMIT*			
2	Binge Watching	149,784 (8%)	59,498 (5.6%)	90,286 (11.2%)
	`WATCH` events only			
3	Lecture→Quiz Complete	100,179 (5.3%)	50,415 (4.7%)	49,764 (6.2%)
	`WATCH` event(s) followed by $\texttt{X}_{\texttt{QUIZ}}$ events; at least one X = *SUBMIT*			
4	Quiz Complete→Forum	99,828 (5.3%)	67,722 (6.3%)	32,106 (4%)
	$\texttt{X}_{\texttt{QUIZ}}$ events (at least one X = *SUBMIT*) followed by $\texttt{X}_{\texttt{FORUM}}$ events			
5	Quiz Complete→Progress	38,854 (2.1%)	26,126 (2.4%)	12,728 (1.6%)
	$\texttt{X}_{\texttt{QUIZ}}$ events (at least one X = *SUBMIT*) followed by $\texttt{X}_{\texttt{Progress}}$ events			

Table 3.4: Most frequent motifs (≥2% chains) in `BusinessX`.

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Quiz Complete	18,446 (16.6%)	11,377 (14.7%)	7,069 (21.1%)
2	Binge Watching	12,530 (11.3%)	8,461 (10.9%)	4,069 (12.1%)
3	Lecture→Quiz Complete	5,060 (4.6%)	3,752 (4.8%)	1,308 (3.9%)
4	Lecture→Forum→Lecture	3,910 (3.5%)	2,386 (3.1%)	1,524 (4.5%)
	`WATCH` events followed by $\texttt{X}_{\texttt{FORUM}}$ events followed `WATCH` events			
5	Quiz Complete→Progress	3,741 (3.4%)	2,898 (3.7%)	843 (2.5%)
6	Quiz Complete → Lecture → Quiz Comp.	2,277 (2.1%)	2,019 (2.6%)	258 (0.8%)

example, when comparing `Framing` (Figure 3.9) and `Data Analysis` (Figure 3.7), marked differences in their pedagogical structure are evident; `Fram-`

Table 3.5: Most frequent motifs (≥2% chains) in `Framing`.

	Motif	**Freq. Total**	**Freq. Passing**	**Freq. Non-pass.**
1	Binge Watching	64,822 (27.3%)	18,023 (18.9%)	46,726 (33.1%)
2	Lecture→Forum→Lecture	29,224 (12.3%)	11,651 (12.2%)	17,505 (12.4%)
3	Quiz Complete	12,984 (5.5%)	9,156 (9.6%)	3,781 (2.7%)
4	Forum→Lecture $\texttt{X}_{\texttt{FORUM}}$ events followed `WATCH` events	7,850 (3.3%)	3,035 (3.2%)	4,800 (3.4%)
5	Lecture→Forum	7,488 (3.2%)	3,008 (3.2%)	4,462 (3.2%)
6	Quiz Complete→Lecture→Quiz Comp.	5,551 (2.3%)	4,022 (4.2%)	1,501 (1.1%)

`ing` appears to foster a very social, collaborative environment, whereas `Data Analysis` learners mostly focus their attention on lectures and assessments, with little concern for discussion. The visualizations also reveal at which specific moments learners seek feedback on their progress (i.e. make a transition to the Progress vertex), such as after a Quiz or Forum in `BusinessX` and `Framing`. These movements are *not* included in any of the courses' designed paths; course designers can use this insight to proactively insert feedback in order to encourage more awareness and self-regulated learning. When comparing transitions of passing vs. non-passing learners, we observe that (i) non-passers make the transition to the video event from more diverse event types than passers (indicating that non-passers' executed paths follow the designed path to a lesser degree than passers' executed paths), (ii) video-to-video transitions are more prevalent among non-passers (in line with our findings on the binge watching motif), and (iii) passing learners are more likely to move from *Quiz Start* to *Quiz Submit*, while non-passing learners are more likely to move from *Quiz Start* to *Quiz End* (without answering a question).

Table 3.6: Most frequent motifs (≥2% chains) in `Data Analysis`.

	Motif	Freq. Total	Freq. Passing	Freq. Non-pass.
1	Quiz Complete	169,786 (9%)	116,878 (11.4%)	52,908 (6.2%)
2	Quiz Complete→Lecture→Quiz Comp.	145,596 (7.7%)	82,247 (8%)	63,349 (7.4%)
3	Binge Watching	87,760 (4.7%)	28,066 (2.7%)	59,694 (7%)
4	Lecture→Quiz Complete	78,790 (4.2%)	41,543 (4.0%)	37,247 (4.4%)
5	Quiz Complete→Lecture	43,612 (2.3%)	21,916 (2.1%)	21,696 (2.5%)
6	Quiz Check $\texttt{QUIZ}_{START}$ followed by $\texttt{QUIZ}_{END}$ events	37,406 (2%)	19,444 (1.9%)	17,962 (2.1%)
7	Quiz Complete→Forum	33,085 (1.8%)	22,126 (2.2%)	10,959 (1.3%)
8	Quiz Check→Lecture	29,079 (1.5%)	12,376 (1.2%)	16,703 (2%)
	Unassigned Chains	1.1M (58.3%)	631,251 (61.5%)	466,061 (54.5%)

3.6 Conclusion

Before adaptive learning systems can reach their potential, two important baselines must be established: (i) the precise learning path the instructor wants the student to follow and (ii) students' natural behavior within the course. Adaptive instruction will be most effective when the differences between these two baselines are both identified and addressed. The present research offers novel insights into the identification of those differences.

Specifically, in this work we have introduced three different approaches (the video interaction graph, behavior pattern chains and event type transitions) to explore and visualize MOOC log traces with respect to the designed and executed learning paths.

We have applied our approaches on the log traces of four different edX-based MOOCs (from different domains and different pedagogical structures) and have shown to what extent learners (as a whole group as well as partitioned into passing and non-passing learners) follow the prescribed path.

In future work, we will expand our analyses to a larger set of MOOCs to gain a greater understanding of the “classes” of xMOOCs that exist on the major MOOC platforms today. We also plan to consider more diverse sub-populations of learners in future analyses, beyond passing or not passing. We will also investigate semi-automatic approaches to the adaptation of MOOC learning paths, in order to minimize the gap between designed and executed paths as well as the impact this work has on engagement, retention, learner success and more fine-grained learner partitions (such as completing, auditing, and sampling learners [131]).

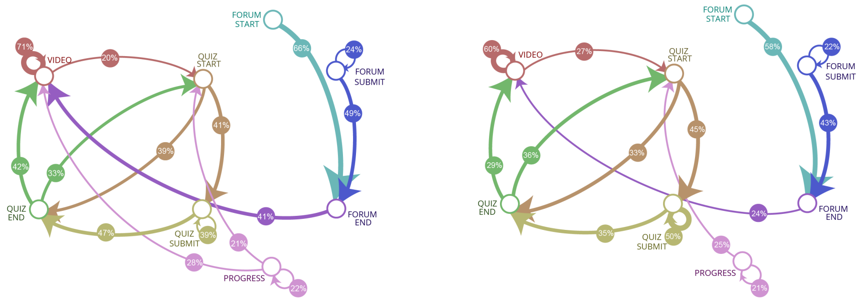


Figure 3.7: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Data Analysis` MOOC. Edges with weights below 20% are hidden from view.

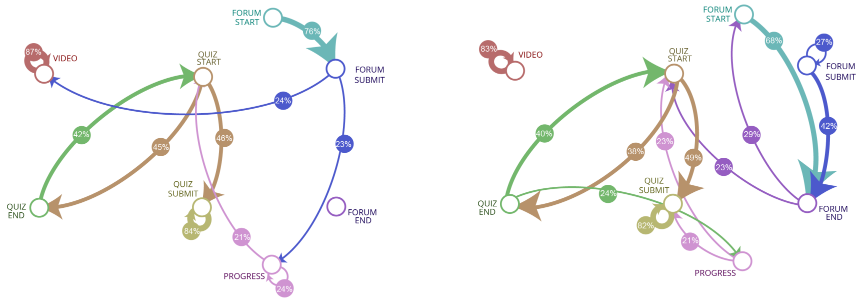


Figure 3.8: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Functional Programming` MOOC. Edges with weights below 20% are hidden from view.

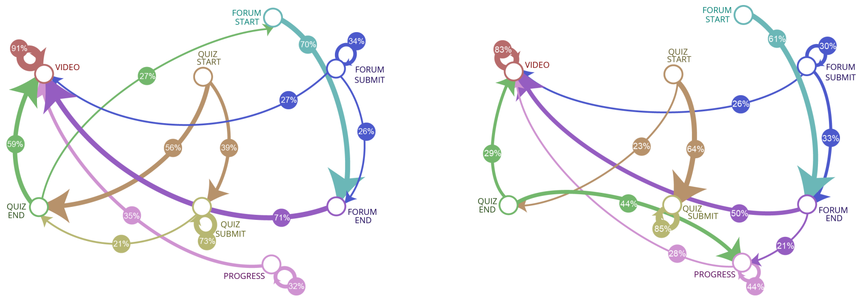


Figure 3.9: Markov Model state visualization of non-passing (left) and passing (right) learners in the `Framing` MOOC. Edges with weights below 20% are hidden from view.

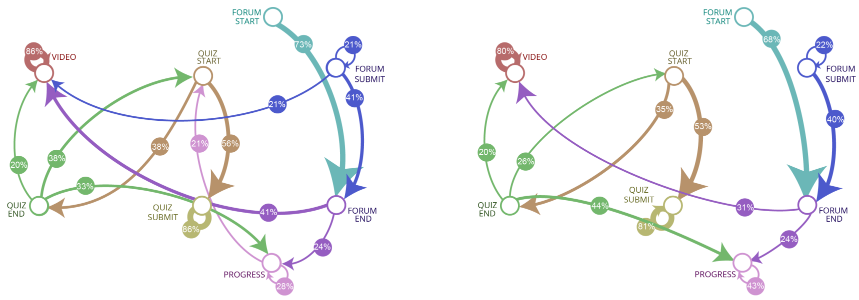


Figure 3.10: Markov Model state visualization of non-passing (left) and passing (right) learners in the `BusinessX` MOOC. Edges with weights below 20% are hidden from view.

Chapter 4

Toward Large-Scale Learning Design

This chapter applies theory and methodology from the learning design literature to large-scale learning environments through quantitative modeling of the structure and design of Massive Open Online Courses. For two institutions of higher education, we automate the task of encoding pedagogy and learning design principles for 177 courses (which accounted for for nearly 4 million enrollments). Course materials from these MOOCs are parsed and abstracted into sequences of components, such as videos and problems. Our key contributions are (i) describing the parsing and abstraction of courses for quantitative analyses, (ii) the automated categorization of similar course designs, and (iii) the identification of key structural components that show relationships between categories and learning design principles. We employ two methods to categorize similar course designs—one aimed at clustering courses using transition probabilities and another using trajectory mining. We then proceed with an exploratory analysis of relationships between our categorization and learning outcomes.

This chapter is published as "Toward Large-Scale Learning Design: Categorizing Course Designs in Service of Supporting Learning Outcomes" [69], by D. Davis, D. Seaton, C. Hauff, and G.J. Houben in Proceedings of the Fifth Annual ACM Conference on Learning at Scale, 2018.

4.1 Introduction

The ubiquity of digital learning platforms is leading to new ways of documenting and understanding course design. Even though online learning platforms often constrain instructors to design choices in the limited context of videos, text, and various assessment components, there still exists a vast and uncharted diversity in the way instructors choose to design and structure their digital learning materials. A recent example in scaled learning is the edX consortium, where over 1,700 courses[2] have been created by 118 institutions[3] across the globe. This makes for a truly massive possibility space that spans discipline, culture, and pedagogy.

MOOC researchers have begun analyzing course design and pedagogy in order to understand this diversity, but the work has been isolated and largely a process of human categorization based on broad interpretations of learning design. Recent applications of pedagogical inventories involving human classification on a number of scales exemplifies these efforts. The authors in [243] have compared the pedagogical structure of 17 MOOCs using an inventory called AMP (Assessing MOOC Pedagogies), and [176] applied a similar inventory across 78 MOOCs. Both found signs that many MOOCs are replicating traditional instruction tactics. Such work can potentially help address best practices in course design, but it has remained a manual task and not yet found widespread adoption by researchers in the MOOC community.

Furthermore, researchers in more traditional areas of learning design have only been able to conduct small-scale (usually on a single, course-by-course basis) mostly-qualitative analyses of course structures and their relationship with learning outcomes. And although the number of courses considered is small (typically ranging from 1–20 [75]), learning designers have developed methods for comparing and classifying courses' structures. This is achieved through a process of *abstraction*, or the separation of a course's topical content (such as math, engineering, history, etc.) and its internal structure (the sequence of activities used to teach the content).

In another area of research, learner behavior modeling has taken off for MOOCs [63, 89, 134, 260]. However, there has yet to be any large-scale or automated evaluation of the effectiveness of various learning design patterns using the tools from the learner-behavior community. So while there is a quickly emerging corpus of learner modeling research unfolding, there have

[2] https://www.edx.org/course?course=all
[3] https://www.edx.org/schools-partners

not yet been any empirical efforts to connect the findings to the design of the learning environment.

By understanding the theory and methodology from the learning design literature and applying it to large-scale learning environments, we are able to advance the field of learning at scale through a quantitative analysis of the structure/design of online learning environments. The MOOC community has primarily focused on learners so far. But the number of courses accessible to researchers is growing large enough to offer a new paradigm for teaching research at scale.

In this paper, we attempt to build a framework that can help aid classification of course design in an automated and scalable fashion. Our framework is largely built around the following ideas:

- A methodology to parse and abstract a course to enable quantitative analyses of its structure.
- Quantitative measurement of the difference between course designs.
- Identification of key structural components that differentiate courses with clustering and then gaining a deeper understanding through qualitative analysis.

Using a dataset made up of 177 MOOCs from two institutions of higher education, we abstract course design into a sequence of learner activities and apply two types of pattern mining, namely, (i) transition probability mining and (ii) trajectory mining. We explore both methods on an institution by institution basis. In addition, we explore the relationship between our classification (clusters) with a straightforward learning outcome — verified learner pass rates. This exploratory addition to the study is to further support whether our abstraction and automation can lend itself to goals of improving learning outcomes through better design.

4.2 Related Work

Below we describe the current state of the art in the domains of learning design and learner behavior. Our review of the literature finds a distinct common thread connecting learning design and learner behavior studies, namely, that of abstraction and complexity reduction. In addition, many of the methods in our work are inspired by research in the area of learner-behavior pattern mining [63, 89, 134, 260]; we find that many methodologies in this field have have potential applications to pattern mining of course structure and pedagogy.

4.2.1 Learning Design Patterns

The learning design literature offers a substantial body of research theorizing about the design of learning patterns and sequences. Reference [156] offers an exhaustive review of how Learning Designers have tackled the challenge of describing and synthesizing *patterns for learning*, defined by [185] as a semi-structured description of a strategy for teaching a given topic or skill. The primary purpose of patterns for learning is to externalize knowledge in a way that can be generalized and accessed by members of the teaching community.

Traditional classroom teaching environments do not require explicit documentation of a strategy or pattern for teaching. These are often proprietary and documentation standards vary across institutions [74, 156]. Reference [156] calls for a standardization to facilitate sharing of patterns for learning throughout the teaching community in having teachers "enact design science" as a normal part of the teaching practice so that, as a community, they can gain an understanding over which designs lead to which outcomes/achievements.

One effort to facilitate the comparison and standardization of teaching design patterns is found in [74] where the authors developed a "Teaching Method Template" which describes instruction primarily in terms of activity sequences—found to be the most effective method of depicting patterns for learning in terms of teacher preference and usefulness.

In this template, reference [74] represents activity sequences both textually and graphically: the *graphical representation* uses flow charts and activity diagrams to visualize patterns in a way users can quickly internalize and the text-based *sequence of activities* approach details the temporal sequence of activities and assessments in a given plan. The authors in reference [156] identify the "sequence of activities" approach (defined as a collection of teaching design patterns building towards an outcome) as the most interesting and promising in the age of digital learning and instruction. This includes the decisions of which activities to introduce at which point, but also the effective transition between activities so that each activity appropriately informs the next. Though this topic is not yet prevalent in the area of digital learning environments, the authors in [156] claim that "The origin or provenance of a pedagogical pattern is as important as citations are in research. Teachers considering adopting a new pattern need to know its origin, and should be able to track the way it has developed into alternative versions." There is not yet a widely accepted standard for patterns as of yet, however, digital learning environments present a tremendous opportunity to develop, track, and share pedagogical patterns due to how content is stored in digital-learning platforms, as was done in [192].

Teaching design patterns that have been identified as topic-specific are referred to as "signature pedagogies"[229].

An example of signature pedagogies is the contrast between hands-on (bedside) teaching for medical education and the inquisitive nature of a law school lecturer (firing off strings of question sequences to their audience members). To the best of our knowledge, no work as of yet has been done to evaluate the detailed patterns of such signature pedagogies in the context of digital-learning environments.

Reference [157] poses this question about the extent to which disciplines can be "disentangled" from their signature pedagogies. This leaves the question open about whether some strategies are best kept tied to a specific discipline, or perhaps through the sharing of such pedagogical wisdom, disciplines can benefit from a new perspective. [156] introduces a method of documenting instructional sequences in a structured and standardized manner so that patterns from one domain "can be replaced with entirely new topic content to generate the same pattern in a different subject area."

The key to this "disentangling" of pedagogies from their disciplines is a successful abstraction of the pedagogy to a form that is transferable to a new context. And by removing the content and only focusing on the activity type and transitions between activities, we arrive at a structured method of documenting patterns and sequences for learning [155, 158].

4.2.2 Learner Behavior Patterns

We next describe methods from research in learner behavior patterns and their applications to the above challenges of learning design patterns. There has recently been a surge in research exploring MOOC learners' navigational patterns throughout course activities. The impact of this research stems from our ability to see learner behavior in highly self-directed environments, i.e., without instructor oversight. However, while these methods continue to be evaluated and developed in the context of learner activity patterns and navigational events, we here propose that similar methods ought to be employed in evaluating course design patterns in digital learning environments. Doing so will allow us to better understand how course design and the sequencing of activities are related to learner behavior. Below we review work on learner behavior modelling while pointing out how previous work influences our methods for course structure pattern mining.

The research presented in [63, 89, 134, 260] characterizes MOOC learners through their clickstream data tracking their transition between activities. Reference [260] first identified common 2-gram event transitions; [63] next extended these to 8-gram event sequences and labeled the sequences as various motifs representing a study pattern; and [134] extended this by connecting these event transitions to self-regulated learning strategies using learner self-reported survey results as well. [89] builds on the Markov modeling technique in [63] by developing a two-layer Markov model which accounts for transitions between both micro and macro activity patterns. In the present research we apply this methodology of analyzing learner transition probabilities to course structure data—exploring the transitions between course components as defined by the instructor as opposed to the path executed by the learner.

Reference [24] builds upon the work in [63, 134, 260] by applying clustering techniques to MOOC learner behavior. Clustering in this case enables the automatic identification of similar trajectories to be identified at scale, whereas prior work in this area was done manually [176, 243]. We apply this scalable clustering approach to MOOC course structures in the present research. Reference [24] employed both pattern- and data-driven approaches for analyzing and clustering MOOC learner activity data. They correlated learner engagement patterns with course learning outcomes as well—final course grades earned and each cluster's overall passing rate. The authors first categorize learners into one of four categories (separated by behavior patterns preceding assessment) on a week-by-week basis to account for changes over time, and then they use hierarchical agglomerative clustering to group learners with similar week-by-week trajectories.

The authors in [24] also introduce a second method to track latent learner activity patterns with an unsupervised processing pipeline. The pipeline is comprised of four phases: (i) activity sequence modeling, where a transition matrix is generated and used as a learner model, (ii) distance computation, (iii) clustering, where the dissimilarity matrix is clustered with hierarchical agglomerative clustering using the Ward's method, and (iv) cluster matching, to identify temporal relationships between identified clusters. Based on this method, the authors enable a direct comparison of various types/patterns (clusters) of behavior and academic achievement, very similar to our method presented here of clustering course structures.

The primary methodology employed in [175, 187] is that of process mining. Such process mining techniques include (i) visualization, where processes are plotted in a variety of graph types in order to make trends and patterns visually apparent, (ii) conformance checking, where actual/executed processes are compared to the normative/intended model, and (iii) process discovery, where a process model is learned from event log data. Whereas the authors in [24] were motivated by connecting behavior to learning outcomes, the authors in [175, 187] are motivated to model learner behavior in order to develop targeted interventions to support learners in developing self-regulated learning skills.

After reviewing the state of the art and existing knowledge gaps in the literature of learning design patterns and learner behavior patterns, we arrive at the following three primary **R**esearch **Q**uestions:

RQ4.1 To what extent can we model the design of a MOOC by employing principles from the learning design literature?

RQ4.2 How can we quantitatively compare and contrast the design of MOOCs?

RQ4.3 Are there structural components that differentiate a MOOC's design?

In addition, we put forward an exploratory **RQ4.4** addressing the relationship between our abstraction of course design and students' learning outcomes.

4.3 Methods

Building upon the learning design methodology of abstracting a course's structure away from content, the present research methodology employs an exploratory approach in applying methods from research in learning design and learner behavior patterns to the topic of learning design patterns in digital learning environments. We next outline the methodology used with regard to each guiding research question.

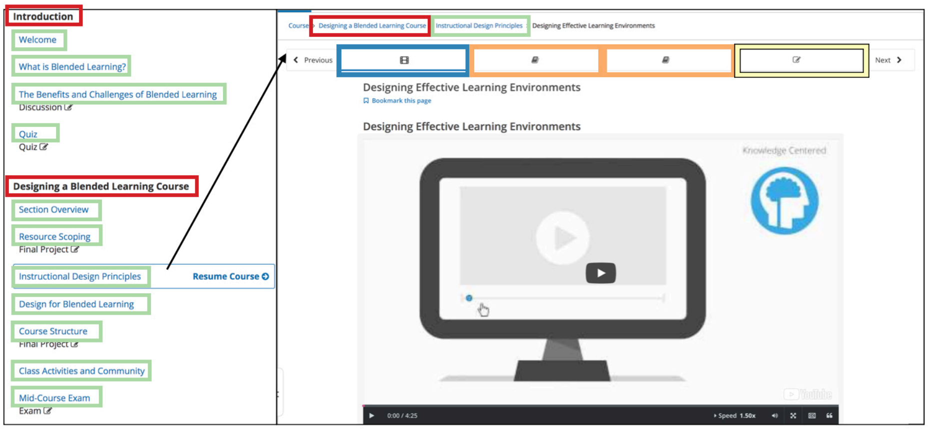


Figure 4.1: edX platform screenshot with components and containers associated with the OLX format marked by color: chapter (red), sequential (green), videos (blue), html (orange), and problems (yellow).

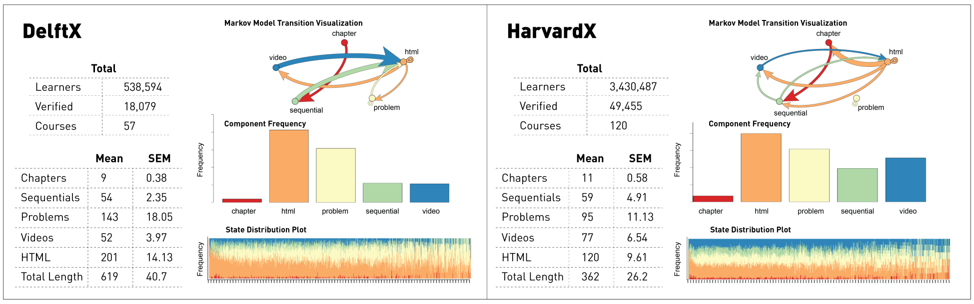


Figure 4.2: Course structure overview for each institution. Tables indicate the total number of enrolled and verified learners for each institution, along with summary statistics about the occurrence of course components (mean per course and standard error of the mean (SEM)). The Markov model transition visualization indicates the most common event type transitions across all courses for each institution; edge/line weights distinguish transition prominence. Component frequency bar graphs show how common each component type was across all courses. The state distribution plot – depicting the left to right occurrence of course components – is a trajectory mining visualization that accounts for the likelihood of component occurrence accounting for all courses in each institution.

4.3.1 Dataset

Our dataset consists of edX MOOCs from Delft University of Technology (or DelftX, as it is known on the edX platform) and Harvard University (HarvardX). Within this study, DelftX accounts for 57 MOOCs with a total of 35,283 course components, and HarvardX accounts for 120 MOOCs with a total of 43,514 components. In edX courses content can be broken up into components and collections. Components are stand-alone assets with which learners interact: videos, problems, html pages, and custom activities. Collections are containers that provide structure and navigation for learners: chapters, sequentials, and verticals. For clarity, all components and collections are illustrated in context in Figure 4.1.

In this study, we remove verticals from consideration to reduce complexity, namely, in our own ability to interpret results, as the institutions studied tend to have short verticals, leading to numerous verticals that act as delimiters between small numbers of resources. While future analyses can include verticals, we found here that verticals in DelftX courses typically include between 2 and 3 resources (avg. of 2.75 resources per vertical) and, for HarvardX courses, 1 to 3 resources per vertical (avg. 1.7 resources per vertical). We omit verticals to allow for an analysis of longer, more representative learning design sequences. We also omit custom components, which have extreme variation in students' interactions and in many cases evolve over time (i.e., may not have the same use case from course to course). In addition, the namespaces for these components do not remain consistent, making them difficult to track in this initial study.

4.3.2 Parsing edX Courses

All content authored for the edX platform is stored in the Open Learning XML (OLX) format[4]. OLX is a standard that allows the transfer of content between instances of the open source edX platform, authorship outside the platform, and extraction of information related to course design (like in this work). OLX contains the raw markdown (XML) for all authored content in a course, namely, all content tags, text associated with content, and relevant metadata. Courses are generally designed in edX Studio – a GUI for creating and structuring courses – masking the OLX from most users. OLX data can be exported through edX Studio and is also provided in regular data exports to edX consortium members through the edX research pipeline [5]. For each course in the present study, we download the OLX data and pass it through a parsing algorithm to structure the data in a more desirable format for analysis (colloquially referred to as the "course axis"). All OLX components are sorted in sequential order according to their placement in the course.

[4]http://edx.readthedocs.io/projects/edx-open-learning-xml
[5]https://github.com/edx/edx-analytics-pipeline

4.3.3 Abstracting Structure from Content

Research in learning design relies heavily on the process of abstracting a course structure into a standardized, comparable structure. Abstraction here is the process of stripping away the course topic materials from the underlying structure and components (**RQ4.1**). For example, in a course about Statistics, a given sequence of activities might include: a lecture about the difference between frequentist and Bayesian statistics → discussion about the benefits and drawbacks of each approach → exam assessing learners' ability to apply what they've learned. The abstracted version of this sequence would become: lecture → discussion → assessment. This method for abstraction is also commonly used when considering learner activity in courses as well [24, 63, 202, 260]. We view this abstraction as similar to processes like coarse-graining in physics, where microscopic structure is often approximated in order to measure macroscopic properties of a system.

1. Original Sequence

A	B	C	A	C	B	A	B

2. Edge List

	A	B	C	A	C	B	A	B
A	B	C	A	C	B	A	B	

3. Transition Matrix (P)

	A	B	C
A	0.0	0.66	0.33
B	0.5	0.0	0.5
C	0.5	0.5	0.0

4. Sample Transition Matrix (Q)

	A	B	C
A	0.25	0.5	0.25
B	0.0	0.7	0.3
C	0.5	0.5	0.0

5. Distance Matrix ($|P - Q|$)

	A	B	C
A	0.25	0.16	0.08
B	0.5	0.7	0.2
C	0.0	0.0	0.0

Figure 4.3: The process of calculating similarity using transition probability. (**1**) Original sequence of elements, read from left to right. (**2**) Edge list: showing the transitions from each source element to its target. (**3**) Transition matrix showing the probability of each source element's (y-axis) transition to each target element (x-axis). Each row sums to 1.0. (**4**) Sample transition matrix (Q) to compare to P. (**5**) Distance matrix for P and Q.

4.3.4 Computing Similarity

After abstraction of a course, we qualitatively measure the differences between course structures (**RQ4.2**) using two approaches: (i) clustering transition probability, and (ii) trajectory mining. Transition probability treats the course activity sequence as a Markov chain and considers the prominence of each of the possible transitions between activity types. The choice for this approach is based on the learning design principle which highlights the importance of the *consecutive* sequencing of learning activities. The trajectory mining approach takes the entire sequence into account by calculating differences in the order and position of all components, which allows for the analysis of learning design sequences over the span of entire courses beyond single transitions.

We employ both methods for computing dissimilarity between course structures because both have been used in prior research for learning path analysis [63, 134, 175], and both methods have their own advantages and drawbacks. For example, the main advantage transition probability has over trajectory mining is that the length of the sequence is not considered, whereas in trajectory mining the difference in sequence length imposes a significant bias/cost on the results. On the other hand, the main benefit trajectory mining has over transition probability is that it takes the entire course sequence into consideration and enables more macro-level course design insights.

Transition Probability

A transition matrix is a method of representing a sequence of transitions, or a Markov chain. Computing a transition matrix has been a prominent method for modeling learner behavior in online learning environments [63, 134, 175], but this method has not yet been applied to teaching or instructional behavior. By adopting a method focusing on transitions from one activity to the next, we are able to connect digital learning environments to the literature on learning design.

We compute transition matrices by first generating an edge list, as shown in Figure 4.3.2 which represents all origin→target pairings (sequential connections from one event type to the next) from the original sequence of elements from Figure 4.3.1.

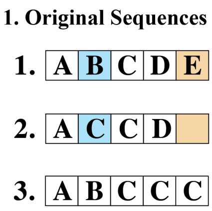
This edge list is then used to compute the probability of each event type transitioning to the next, and these proportions are then used to populate the final transition matrix.

We generated transition matrices (P and Q) for all 177 courses included in the study and stored them in a list of matrices.

For each institution, we generate transition matrices for each course. We then calculate the L_1 distance (also referred to as Manhattan distance or taxicab metric) (d_1) between transition matrices ($P - Q$) on a course by course basis and sum the absolute values between them:

$$d_1(P, Q) = ||P - Q||_1 = \sum_{i=1}^{n} |P_i - Q_i| \tag{4.1}$$

where P and Q are transition matrices flattened into one-dimensional vectors.



2. OM Distances

	1	2	3
1	0.0	3.0	4.0
2	--	0.0	5.0
3	--	--	0.0

Figure 4.4: The process of calculating OM distance. (**1**) Original three sequences of elements, read from left to right. (**2**) Matrix showing the OM distance between sequences as the cheapest total editing cost.

For example, the distance between P and Q, $d_1(P, Q)$, in Figure 4.3 amounts to 1.89. The final distance matrix contains each of these calculated differences for all matrix pairings and is then is a suitable format to be clustered—noting that all matrices must contain the same columns and rows to ensure appropriate calculations.

Trajectory Mining

The trajectory mining method first computes a distance matrix using the optimal matching (OM) method. This distance matrix is populated by edit distances (or the minimal editing cost): the minimal cost of all insertions, substitutions, and deletions to transform one sequence into another [161]. In accordance with the method introduced in [161], substitutions (C_S) have an editing cost of 2.0 and insertions & deletions (C_I) have an editing cost of 1.0. The editing costs according to [161] are:

$$C_S = 2 - p(i|j) - p(j|i) \quad \text{and} \quad C_I = 1 - p(i|j) - p(j|i) \tag{4.2}$$

where $p(i \mid j)$ is the transition rate between states i and j.

Figure 4.4 illustrates the process of arriving at the distance matrix between two sequences with a substitution colored in blue and an insertion colored in orange for sequences 1 and 2.

4.3.5 Clustering Similar Courses

In service of **RQ4.3**, we uncover similarities in courses' structures by employing Ward's method of hierarchical agglomerative clustering. This method starts by considering all courses as n independent clusters. The algorithm progresses by forming $n-1$ clusters and computing the error sum of squares and r^2 value at each step. Clusters are then formed by grouping units which yield the lowest sum of squares and highest r^2 values. When all n units are combined into a single large cluster tree (or dendrogram), the algorithm stops.

Once we have completed Ward's hierarchical clustering method, we then determine the optimal number of clusters within that single tree. To do so we employ the Calinski-Harabasz index method [36]. This method evaluates the validity of clusters according to the average within-cluster sum of squares and the average between-cluster sum of squares [168]. The index aims to maximize both the distance between cluster centers as well as the individual cluster compactness. We next verified this result by calculating the silhouette scores of each clustering result, an alternative method for measuring cluster tightness and separation [217].

These approaches are a common and widely accepted way of uncovering trends in large datasets [1, 36, 168] and have been successfully applied to large-scale learning problems in the past [24]. Based on the results of these analyses, we then address **RQ4.3** by drawing semantic meaning through qualitative analyses of the clusters.

4.3.6 Exploring Course Learning Outcomes

After developing an understanding of common course designs, we next explore the extent to which similar course designs are related to learning outcomes (**RQ4.4**). To evaluate in an exploratory fashion whether there are statistically significant differences in completion rates between clusters, we fit a one-way ANOVA model considering course completion rates among verified learners (those who went through a process to verify their identity with edX) by cluster group.

4.4 Results

4.4.1 Abstracting Structure from Content

In service of **RQ4.1**, we find that our abstraction of courses sufficiently enables qualitative insights into course design decisions. For example, the authors on this paper from HarvardX can confirm an abundance of video in Figure 4.2, as reflected in the bar graph. Additionally, HarvardX pivoted toward smaller, modular courses. In some cases, taking long 16 week courses and breaking them up into multiple course—reflected in the average course

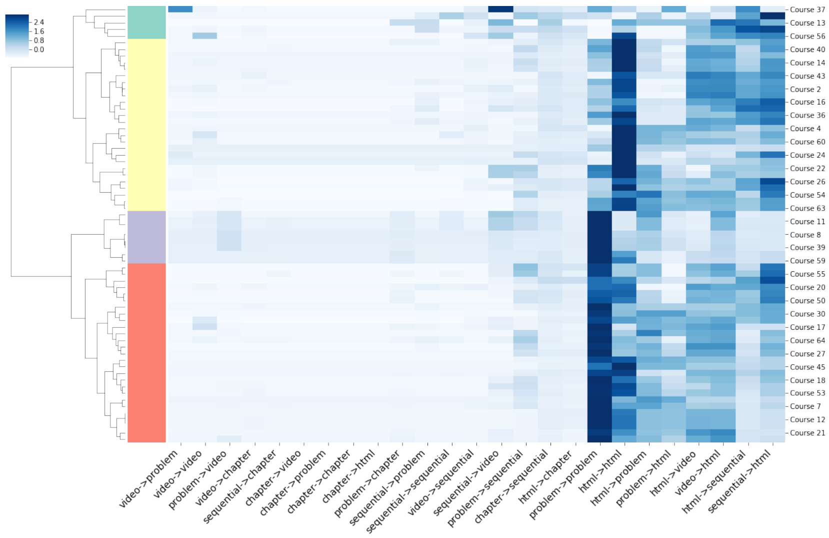


Figure 4.5: Hierarchically clustered heatmaps using transition probabilities as input for DelftX. The color of each cell shows the prominence of a transition (x-axis; darker colors are higher values). To the left is a dendrogram showing the agglomerative clustering process, and the leftmost column shows which cluster each course (y-axis) belongs to. Course titles withheld for blind review. Best viewed in color.

length. For DelftX, which offers predominantly STEM courses, we confirm a trend towards longer courses containing more assessment activities.

In the following analyses, we draw the following connections between the syntactic form of the OLX format and the semantics of learning design: *chapter* and *sequential* elements indicate a section break in the course continuity. Sequentials house subtopics of chapters and are used to break up material into manageable chunks for learners. *Video* components are indicative of video lecture activities and are the primary method for introducing learners to new content or concepts. *Problem* elements are used as graded assessment events where learners are given the opportunity to test their newly gained knowledge. Lastly, *html* elements are used to help guide the learner between video lectures and assessments and provide navigational guidance/context.

From this method, we find evidence that despite the limited number of elements available in an online learning platform like edX, substantial variation does indeed occur in the learning and structural design of various courses.

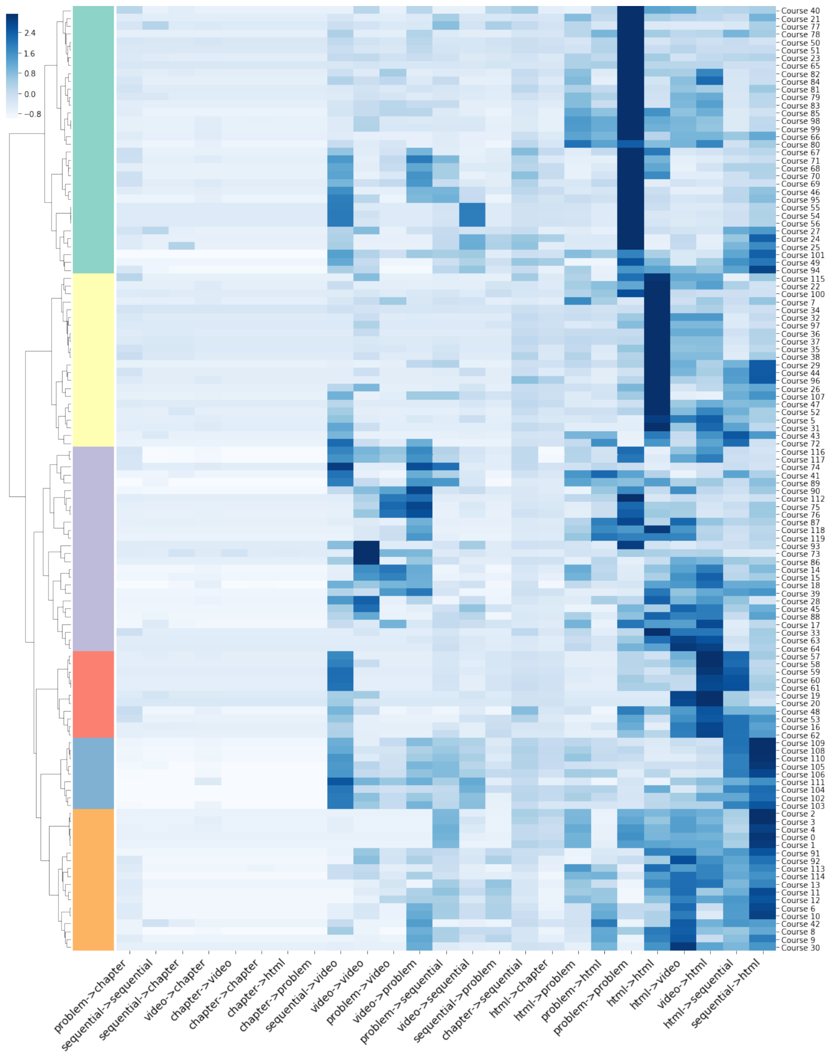


Figure 4.6: Hierarchically clustered heatmaps using transition probabilities as input for HarvardX.

4.4.2 Clustering Similar Course Structures

The following results address the quantitative comparison of course structures toward **RQ4.2**. Figures 4.5 and 4.6 visualize the transition probability features (color-map, where darker cells are larger values) and the dendrogram based on our agglomerative clustering approach. Clusters are indicated by color in the leftmost column of each figure, namely, 4 clusters for DelftX in Fig. 4.5 and 6 clusters for HarvardX in Fig. 4.6.

Table 4.1: The percentage of transition types for all courses within clusters for both institutions. The bottom row indicates the total number of courses included in each cluster. Only the most prominent transition types /factors are shown.

	DelftX				HarvardX					
	Cluster 1	2	3	4	1	2	3	4	5	6
% All Transitions										
html→html	10.5	18.9	8.7	8.5	5.8	31.7	8.8	2.9	3.7	10.2
html→problem	0.0	4.8	1.7	2.4	6.0	4.9	4.9	0.2	4.8	5.9
html→video	10.5	15.4	11.2	14.9	5.2	8.1	9.6	17.4	3.7	13.9
problem→html	2.6	9.7	6.9	10.7	5.1	4.4	5.3	0.4	3.8	5.4
problem→problem	1.3	12.8	61.1	34.6	32.8	7.7	12.2	7.0	3.0	4.2
sequential→html	31.6	14.1	2.9	7.6	7.3	7.7	5.4	9.2	15.0	15.5
sequential→problem	7.9	4.0	2.5	4.1	1.5	0.8	0.9	2.6	2.5	2.9
sequential→video	5.3	0.4	0.0	0.2	6.0	4.0	5.0	9.2	10.8	1.3
video→html	10.5	12.8	0.4	12.2	5.1	10.2	9.1	26.1	4.8	9.7
video→video	1.3	0.4	0.0	0.2	2.6	2.4	7.8	1.4	4.5	2.0
# Courses	4	23	6	24	34	21	25	11	9	18

In identifying the ideal cluster number for the transition probability method we relied on the Calinski-Harabasz index [36] and silhouette [217] method, along with viewing our dendrograms (y-axis of Figures 4.5 and 4.6) for sensible cutoffs [154].

To determine the optimal number of clusters to use with the trajectory mining approach, we again computed clustering quality measures using the Calinski-Harabasz index [36] and silhouette [217] method. We determined the optimal number of clusters for DelftX to be four and for HarvardX to be three.

4.4.3 Key Structural Components

With regard to **RQ4.3** which is concerned with identifying the key structural components that define each cluster of similar courses based on quantitative analyses of their syntactic structure, we highlight the qualitative insights offered by each method into the semantic trends which define each cluster. By contextualizing each element into its place in the course relative to other elements, we identify learning design patterns that distinguish each category.

Transition Probability: DelftX

With regard to DelftX, Figure 4.5 shows two key transition types with prominent transition rates correlated to clusters, namely, *problem-problem* and *html-html*, both indicated by darker color. These are in contrast to the less-prominent transitions found in the left portion of the graph (such as *video-problem* transitions). The cluster map indicates that some transition rates have larger effects than others.

The most dominant feature in **Cluster 1** (green) is the *sequential-html* transition type which accounts for 31.6% of all transitions in the cluster, indicating frequent use of text/reading activities to introduce new sequences. Another prominent feature of is the proportion of *video-video* events, which account for 1.3% of all transitions. And even though this indicates a low prominence of consecutive video lectures, it is the highest among clusters from University A (but lower than any HarvardX cluster; to be discussed).

In **Cluster 2** (yellow), the *html-html* transition type accounts for 18.9% of all transitions, indicating a substantial amount of consecutive reading activities. And with *html-video* also being a dominant feature in this cluster, we see that those sequences of consecutive reading activities are often followed by a video lecture activity. Also worth noting is the trend that any transition involving html elements/reading activities is high in this cluster, indicating that, regardless of context, courses here are comprised mainly of reading activities.

The *problem-problem* transition type is the most prominent feature of **Cluster 3** (purple) in accounting for 61.1% of all transitions in the courses. That is nearly twice as prominent as any other transition frequency from either institution. We may assume long assessment activities to be the main function of the courses in this cluster. There are no chains of consecutive video lectures, and sections never begin with videos.

The *problem-problem* transition type is also a dominant feature in **Cluster 4** (red), but this cluster is distinguished from Cluster 3 with its relatively high frequency of *video-html* transitions. While Cluster 3 contained very few video lecture activities, Cluster 4 strikes a closer balance of being assessment heavy while still offering more video lecture activities. From this transition we further note that reading activities typically follow video lectures, likely providing a summary or preparing learners for the next assessment activity.

Transition Probability: HarvardX

With regard to HarvardX, Figure 4.6 shows more clusters and more variation among clusters. While containing a largely even distribution of most transition types, **Cluster 1** (green) is dominated by the *problem-problem* feature, which accounts for 32.8% of all transitions in this cluster. This indicates that, similar to Clusters 3 and 4 in DelftX, these courses contain numerous long assessment activities. Another trait of courses in this cluster is the relatively high prominence of the *html-problem* feature. This indicates that courses in this cluster most often preface their assessment activities with a reading activity.

In **Cluster 2** (yellow), 31.7% of all transitions among courses are *html-html*, meaning that courses in this cluster have frequent, extended strings of consecutive reading activities. Also prominent at 10.2% of transitions is the *video-html* type. From this we infer that those long strings of reading activities are often preceded by lecture video activities.

The most distinguishing feature of **Cluster 3** (purple) is *video-video* transition types, accounting for 7.8% of all transitions in this cluster. Across institutions, this cluster has the highest frequency of consecutive video activities. Video lectures are the primary method of instruction here, and we also see these lectures are typically followed by strings of consecutive assessment activities (accounting for 12.2% of all transitions in the cluster).

Cluster 4 (red) is primarily characterized by prominent *video-html* transition types (26.1%), which is more than twice the frequency of any other cluster for this transition type. This indicates that video lecture activities for courses in this cluster are most frequently followed by reading activities. And given that the most common transition from reading activities is to video lecture activities (*html-video* 17.4%), we can see that courses in this cluster often adopt the pattern of alternating video lecture and reading activities.

The most unique trait of **Cluster 5** (blue) is a relatively even distribution of all event transition types. The two most prominent are *sequential-html* and *sequential-video*, at 15.0% and 10.8% respectively. This may indicate that new sections in these courses are typically introduced with either reading or video lecture activities, noting the frequency of new sections beginning with video activities in this course is the highest among clusters from HarvardX.

The *sequential-html* (15.5%) transition type is also the most prominent in **Cluster 6** (orange), but this cluster is differentiated from Cluster 5 by its low *sequential-video* transition type (1.3%), which this is the lowest of any cluster from this institution. Also prominent in this cluster is the high frequency of the *html-video* transition type, which shows a mixing of video lectures and reading activities.

Trajectory Mining: DelftX

In Figures 4.7 and 4.8 we observe that the clustering results from the trajectory mining approach are primarily influenced by (i) the length of a given sequence, (ii) the frequency of activity types within a sequence, and (iii) the temporal order/placement of activities within the entire course trajectory, whereas the clustering based on transition probabilities (above) was illustrative of the sequence and order of activity types.

With regard to DelftX (Figure 4.7), in **Cluster 1** we observe mostly reading (representing 42% of all activities) and assessment activities (34%). One interesting characteristic of this cluster's learning design is the equal frequency of section breaks and video lecture activities (each at 11%). This indicates that each section in the course consists of a single video lecture activity. Reading activities are most prominent at the beginning of courses in this cluster.

The most distinguishing trait of **Cluster 2** is the prominence of assessment activities (48%) found throughout the sequence, with some notable large spikes in frequency throughout the courses—indicative of long exams and problem sets. Reading activities are also prominent at the beginning of this cluster of courses. With the ratio of video lectures (12%) to section breaks (6%) strongly favoring the former, we observe that, unlike Cluster 1, each section is typically made up of two video lecture activities.

Cluster 3 is clearly characterized primarily by its short length, being on average half the length of others. The cluster is also quite noisy—lacking any discernable patterns. While the state distribution plot may not be the most illustrative due to its length, we do observe activity frequencies largely comprised of reading activities (42%).

The state distribution for **Cluster 4** indicates reading activities to be the prominent activity (54%). There are three times as many reading activities as there are assessments (18%) and very few video lecture activities (11%). We also observe high frequency of reading activities at the beginning of these courses, which indicates design patterns where introductory texts are used to prime learners.

Trajectory Mining: HarvardX

With regard to HarvardX (Figure 4.8), in **Cluster 1** we observe mostly assessment activities (40%) followed by a relatively high frequency of video lecture activities (20%). The general trajectory of these courses can be understood as designs of short introductory reading activities at the beginning of the course followed by long sequences heavy with assessment activities with the sporadic video lecture mixed in.

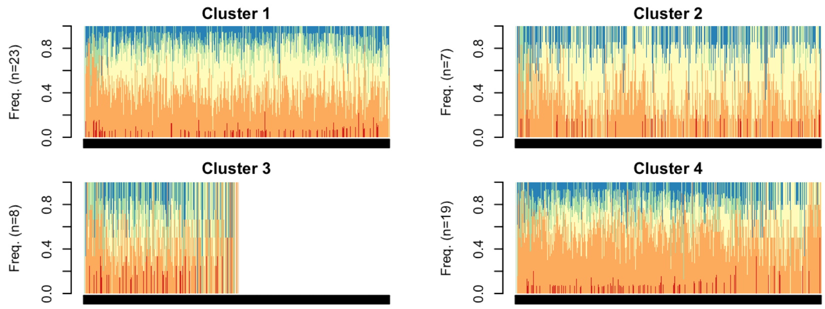


Figure 4.7: State distribution plots from the trajectory mining method per cluster for DelftX showing, at each point along the x-axis, the proportion of each component type.

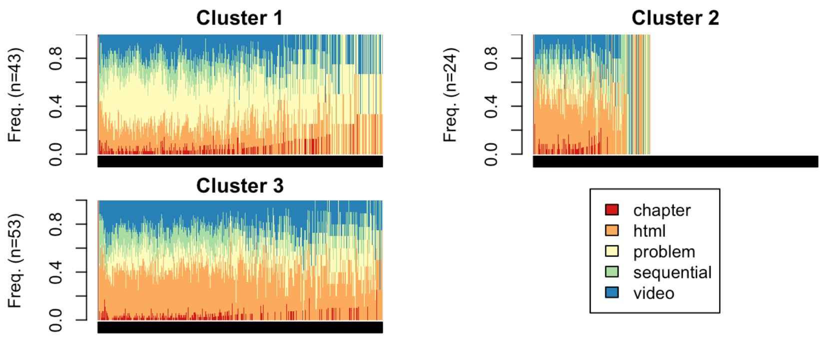


Figure 4.8: State distribution plots per cluster for HarvardX.

Similar to DelftX's Cluster 3, **Cluster 2** consists of courses with short length (on average 250 components). In addition, it appears noisy and without clear patterns from the visualization. However we observe that it is largely comprised of reading activities (39%) with very few video lectures (22%).

Cluster 3 contains courses with a high frequency of video lectures (23%) and reading activities (45%). As is the case with Cluster 2 from DelftX, these are the only clusters with more videos than problems, indicating that courses in these clusters focus primarily on content delivery.

An interesting trend across all three clusters for HarvardX is that each cluster's courses start with a spike in reading activities. This is most likely introductory or motivational material aimed at helping students persist through the course. A similar trend can be discerned in clusters from DelftX.

While the trajectory mining approach provides insights along three structural components (length, frequency of activity types, and temporal location of activities), the transition probability approach, even though it is limited to considering a single primary structural component (transitions and the order of activities), offers concrete insights into the order of a course's activities, which makes it directly applicable to principles from the learning design literature about designing activity sequences. However, the results in Figures 4.7 and 4.8 show that the trajectory mining approach enables insightful temporal analyses in that they show how evolution of patterns and sequences over the various stages of the courses and reveal key similarities and differences not only between clusters but institutions as well.

4.4.4 Learning Outcomes

With service to our exploratory **RQ4.4**, we examine the extent to which clusters are correlated with different learning outcomes as measured by course completion rates (the proportion of verified learners earning a passing grade). To see if any of the observed differences in completion rates between clusters are statistically significant (at the $\alpha = 0.05$ level), we conducted an exploratory analysis by fitting a one-way ANOVA model. For DelftX (Figure 4.9 containing means and standard errors), we find the differences in neither model (transition probability and trajectory mining) to be statistically significant ($p = 0.74$ and $p = 0.31$ respectively).

For HarvardX (Figure 4.10), a one-way ANOVA shows that for the transition probability approach, there is a statistically significant relationship between clusters and completion rates ($p = 0.002$). We therefore conducted a Tukey post-hoc test to identify which pairs of clusters were significantly different. We observe significant differences between Clusters 1 and 5 ($p = 0.002$) and Clusters 5 and 6 ($p = 0.004$). The ANOVA model for the trajectory mining approach was not statistically significant ($p = 0.39$). We present any differences strictly as correlation (not causal) and a sign that more work should be done in the future to explore any causality in this relationship.

4.5 Discussion

The selection of the two institutions for the current study was a product of both of them having offered a large number of MOOCs and a mutual interest and willingness to collaborate. While these institutions combined offer a large collection of courses, they represent less than 2% of all institutions (and less than 10% of courses) on the edX platform. More generalizeable findings are likely found by including more courses and institutions in future analyses.

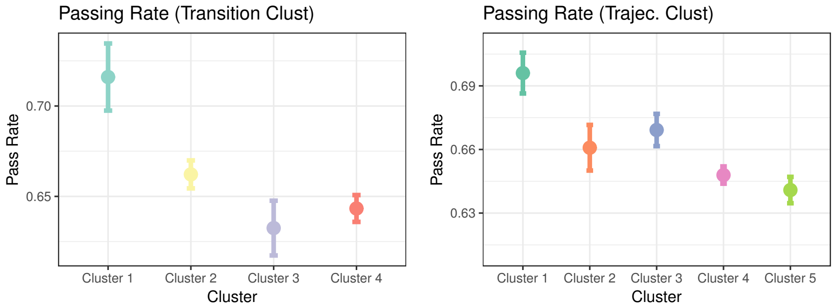


Figure 4.9: The mean and SEM (error bars) of passing rates of each cluster from DelftX courses.

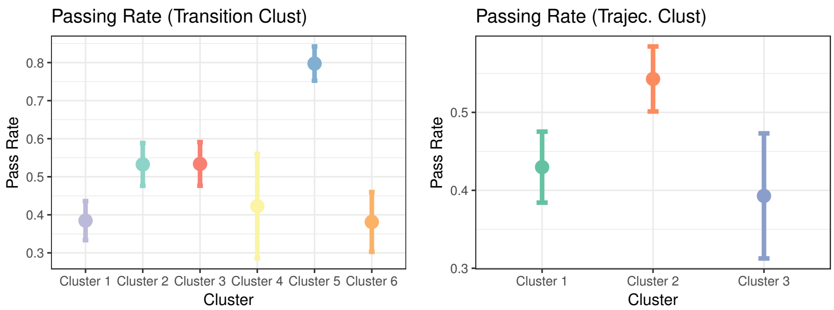


Figure 4.10: The mean and SEM (error bars) of passing rates of each cluster from HarvardX courses.

Regarding the findings from **RQ4.4** from HarvardX, while we are not yet equipped with enough evidence to present this as a causal relationship, we note that HarvardX not only has more courses than DelftX, but also more variation in the learning design and structure of courses. We are encouraged that our methods show differences in learning outcomes based on our course-design abstraction, and this further indicates that this research would benefit greatly from the involvement of more institutions so that we can consider the full spectrum of learning designs and continue to dig deeper into their relationship with learning outcomes.

Future work should explore to what extent increasing the number of grams (sequences longer than two pairs of activities/elements) for the transition probabilities can impact the (i) insights afforded by the results and visualization and (ii) learning outcomes from each cluster. It should also be insightful if in future research, instead of taking the entire course to encode as input, one conducted a similar method using only course chapters or weeks.

Additionally, we explored the predictive power of our course design data on a course by course basis for HarvardX. In the linear model, predictor variables included total number of activities and transitions, and the outcome variables included certification rates and the percentage of chapters visited (a proxy for learner engagement). Each variable was transformed $x \to \log(1 + x)$ prior to regression and normalized to unit variance. For each outcome variable we performed a step-wise regression to identify the optimal subset of predictor variables.

We found virtually no predictive power for certification outcomes using multi-regression (R^2 nearly zero), but did find significance for grade and the percentage of chapters explored ($p \leq 0.05$) within our regression coefficients for 15 of our 25 predictors. For activity frequencies, we found the number of reading activities and section breaks significant, with a negative effect on both the grade and on the percentage of chapters explored (the R^2 of the regression was 0.26 for the grade and 0.63 for the percentage of chapters explored).

We discuss this modeling simply to indicate our abstraction of course design may have predictive power for aspects of learner behavior, i.e., not just outcomes such as grades or certification. Our future work plans to address this more deeply by taking advantage of broader categories of learner metrics.

4.6 Conclusion

In this research we present a successful method of abstracting the design of a MOOC according to principles from the learning design literature (**RQ4.1**). Using this method we then quantitatively compare and contrast the design of the courses using both transition probability clustering and trajectory mining (**RQ4.2**). This then enabled us to draw qualitative insights about the commonalities among courses in each cluster—revealing latent themes in learning design patterns by MOOC instructors and designers (**RQ4.3**). To explore the validity of these findings, we evaluate the extent to which these identified trends in the learning design are associated with learning outcomes in the courses examined (**RQ4.4**). This new avenue of documenting and understanding pedagogy at scale enables novel lines of inquiry in online learning research by directly connecting teaching/ learning design trends to measurable trends in learner engagement.

We are inspired by our ability to automate the process of categorizing course designs and propose that future work needs to continue to refine and test our abstraction method and how it impacts categorization. We also hope to expand our outcome metrics in order to further explore the relationships with course design. Above all, we hope that our work will be a first step in showing the value of addressing digital learning environments from a course structure perspective and finding new challenges as digitization takes an even firmer hold in the learning sciences.

Part III

Study Planning

This part serves **RQ2** (*How can MOOC environments be improved to advance the possibilities of experimentation?*) by focusing on the learning strategy of study planning. Given that myriad research in the past—which had been carried out in in-person, mostly lab-based contexts—provides evidence for study planning as an effective learning strategy which can cause improved learning outcomes, we here evaluate the extent to which these findings transfer and apply to large-scale learning environments.

This is in integral line of inquiry due to the the growing popularity and dependence on online education today. This line of research challenges the notion that all of the study strategies which have been deemed effective in traditional learning environments can be directly used and applied with success in the online context. With this iterative series of experiments, each building upon the last, online learning researchers can begin to develop new principles for learning sciences which are specific to the online context and highlight areas where online and traditional learners behave differently.

In **Chapter 5** we present an evaluation of the effectiveness of a new type of intervention (again evaluated in a randomized-controlled trial) that was designed to improve learners' study planning behavior. We designed this next intervention to require minimal effort and engagement from the learner while still providing them with valuable information that can positively affect their learning experience. To this end, we created the `Learning Tracker`, a dashboard-type visualization that shows a learner's own behavior (across six behavioral metrics) compared to a previously successful learner in the same course. We ran this intervention in four MOOCs and found that it significantly increased passing rates across all four courses.

In **Chapter 6** we present a pilot study in the form of a randomized-controlled trial in a MOOC setting which evaluated the effectiveness of a simple study planning interface. With this interface, learners were given the opportunity to explicitly state their study plans for each week (at the beginning), and at the end of each week they were prompted to think back and reflect on how well they did in achieving their goals and sticking to their plans. Despite the fact that similar interventions had strong, significant effects in traditional learning contexts, we did not observe any significant change in learner outcomes based on this intervention. This served as an early indicator that interventions that have been effective in traditional learning environments cannot be readily applied to MOOC learners. The main study in Chapter 6 introduces `SRLx`, a personalized study planning tool which borrows the key components from the `Learning Tracker` (visualized feedback of behavior, this time delivered and updated in real-time), along with new features which offer the learner more ways to engage. For example, with this new tool, learners can set their own qualitative (text-based descriptions of their goals, motivations, and intentions for the course) and quantitative (number of quizzes, videos, etc. to engage with) goals with regular feedback moments built-in. So that the most possible learners could benefit, we offered `SRLx` to an entire MOOC and evaluated the manner with which learners engaged with it.

In summary, this part begins by testing the applicability of study planning to the MOOC context, and upon identifying promising results, contributes a fully interactive, personalized study planning system that can be implemented in any of the more than 2,000 MOOCs on the edX platform.

Chapter 5

Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale

Social comparison theory asserts that we establish our social and personal worth by comparing ourselves to others. In in-person learning environments, social comparison offers students critical feedback on how to behave and be successful. By contrast, online learning environments afford fewer social cues to facilitate social comparison. Can increased availability of such cues promote effective self-regulatory behavior and achievement in Massive Open Online Courses? We developed a personalized feedback system that facilitates social comparison with previously successful learners based on an interactive visualization of multiple behavioral indicators. Across four randomized controlled trials in MOOCs (overall $N = 33,726$), we find: (1) the availability of social comparison cues significantly increases completion rates, (2) this type of feedback benefits highly educated learners, and (3) learners' cultural context plays a significant role in their course engagement and achievement.

This chapter is published as "Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale" [66], by D. Davis, I. Jivet, R. Kizilcec, G. Chen, C. Hauff, and G.J. Houben in Proceedings of the of 7th International Conference on Learning Analytics and Knowledge, 2017.

5.1 Introduction

A mechanism for increasing access to higher education content, Massive Open Online Courses have afforded millions of people worldwide the opportunity to learn for little or no cost. To achieve this unprecedented scale, MOOCs provide the same material to all learners, no matter what background, motivation, and skill set they possess. Yet this approach falls short of leveraging the technical possibilities of contemporary educational resources to offer learners personalized support, such as giving guidance to learners who are less adept at regulating their learning process over several weeks to achieve mastery. Low course completion rates (typically between 5-10%) highlight the need for additional support in MOOCs. While many learners have no intention to complete MOOCs and instead use them to fulfill alternative needs (e.g., to refresh their memory of a specific topic or to meet new people), the majority of learners who are motivated and committed to complete the course still fail to achieve their goal [131, 130]. Most learners report that they could not find the time to keep up with the course, a challenge that is related to insufficient self-regulatory abilities [274, 129]. Self-regulated learning (SRL; i.e., the ability to plan, monitor, and actively control one's learning process) is associated with a higher likelihood of achieving personal course goals in MOOCs, including course completion [134, 167]. However, the current design of MOOCs does not support learners to engage in SRL [176]. In particular, most MOOC platforms do not provide learners with personalized feedback beyond grades [64], and thus, learners may not know if their engagement in the course is conducive to achieving their learning goals.

We propose a technological solution that facilitates social comparison to help learners regulate their learning behavior to support course completion. According to social comparison theory [82], people establish their social and personal worth by comparing themselves to others. Offering learners the opportunity to compare their behavior with that of their peers promotes increased student achievement in formal learning environments [19, 104, 200]. Students in in-person classrooms can easily identify role models and regularly monitor these role models' behavior and compare it to their own. However, this affordance of social comparison is missing in most online "classrooms." Instead, online learners need to be self-directed and regulate their learning process independently with sparse social and normative signals.

In addition to evaluating the impact of providing learners with personalized feedback, we further examined the potential of adjusting the framing of the feedback to match learners' cultural context. Framing feedback in a way that is consistent with the norms and achievement-based motivation of learners' cultural context is expected to support internalization and behavior change. Prior work has observed differences in the way learners from different countries and cultures interact with MOOCs [94, 129, 169]. We define cultural context based on two established country-level cultural dimensions: *individualism* by Hofstede et al. [99] and *tightness* by Gelfand et al. [90].

We explore the extent to which insights from the social comparison and cultural psychology literature can be translated to support learners in MOOCs. We evaluate how to offer feedback based on social comparison in an online learning environment. To this end, we design, develop, and empirically evaluate a personalized and scalable feedback system that presents MOOC learners with a visual comparison of their behavior to that of their "successful" peers, that is, those who completed the course in the past. We deployed the system in four edX MOOCs offered by the Delft University of Technology with a total of $N = 33,726$ learners. In each deployment we drew on research findings across multiple domains including learning analytics, educational psychology, and social & cultural psychology to inform the design on both the feedback we provide (i.e. the behavioural metrics shown to the learners) and how the feedback is framed (e.g., individualistic- or collectivist- oriented framing).

Our work extends prior research by testing a theory-informed technological solution in a large and diverse population (i.e., MOOC learners) for a prolonged period of time. These are our main findings:

- Personalized social-comparison feedback increases course completion rates.
- Only highly educated learners benefit from this kind of feedback.
- Course engagement and achievement varies by cultural context: learners in countries with a "loose" culture outperform those in countries with a "tight" culture.

5.2 Background

In this section we provide the theoretical and empirical underpinnings to our work which facilitates social comparisons with personalized feedback. We discuss (i) previous studies on incorporating feedback in online learning, (ii) the theory of social comparison and its application to learning, and (iii) past research on the impact of learners' cultural context on learning behavior.

Feedback Providing feedback is one of the most effective teaching strategies to improve student achievement [96]. Given the scale of MOOCs it is impossible for a teacher or teaching assistant to personally monitor and attend to each learner's unique needs. Therefore, up to this point, the majority of feedback solutions developed for MOOCs and other online learning environments have been for the course instructor, typically in the form of a dashboard representing aggregated learner data [160, 174, 257].

While teacher-facing feedback systems can provide key insights for improving the course experience, they are unlikely to address the issue that many learners feel lost and isolated in MOOCs [122]. Personalized feedback promises to promote effective SRL behavior by facilitating self-monitoring of learning processes [119]. One of the most important lines of research which aims to provide learners with personalized feedback is that of Open Learner Models (OLM), an educational interface that gives learners insight into their current knowledge state and activity patterns, which are typically unavailable to them [32]. By allowing learners to visualize and reflect on their own learning and achievements, OLMs have been proven to work as powerful metacognitive feedback tools that impact learners' use of SRL strategies [33, 93]. We designed the `Feedback System` informed by prior work on the design of accessible, understandable, and scrutable [121] learner models [53, 120].

There has been little progress in developing and deploying personalized feedback for large-scale MOOC environments, and most work focuses on supporting teachers [223]. In the present research, instead of presenting aggregate data for all learners in a course, we addresses the challenge of delivering individualized, targeted feedback to each learner based on her behavior in the course relative to her peers' behavior to facilitate social comparison. The present research contributes an empirically evaluated scalable and personalized feedback intervention to the literature on learning analytics. Recent studies have begun to run controlled experiments [204], but most feedback system evaluations thus far explain the design, development, and implementation considerations without rigorously testing whether the added support contributes to behavior change or learning gains [257].

Social Comparison The feedback learners receive through the `Feedback System` is grounded in social comparison theory, initially proposed by Festinger [82]. The theory posits that, guided by a drive to continuously improve, people evaluate their abilities through comparison to others when they are lacking objective means of comparison. It has received empirical validation and found application in various domains, including marketing, health psychology, interpersonal relationships, and also in education [56, 162]. In one study, social comparison was used to improve the Web search behavior of novice users [15]. The authors found that showing non-expert searchers visual indicators of the search behaviors of expert searchers resulted in closer alignment with effective behavior and, therefore, more successful search task completion among novices.

Social comparison is an inherent phenomenon in traditional classroom environments because of both the visibility and accessibility of similar peers [162]. Multiple studies have demonstrated that comparing oneself to self-selected peers who perform slightly better has a beneficial effect on middle school students' grades [19, 104]. Forced comparisons also have a beneficial effect on performance when the target of comparison is performing slightly better than the learner, although no effects were found when there was a big performance gap between two sides [103].

In the context of a small online learning platform ($N = 55$), Papanikolaou [200] investigated students' attitudes towards viewing the learner model of others. Her results showed that when learners compare their behavior to that of a "desired" one, they are then motivated to recognize and adapt their learning strategies. She suggests that the "desired" state should be generated based on real data coming from peers who are "worth following." We build on this insight by considering MOOC graduates of previous editions as the basis for creating a role model.

Guerra et al. [93] integrated social comparison features in the form of peer and class progress in the design of an intelligent interface for a learning management system to provide additional motivation and navigation support. This approach showed a positive effect on engagement and efficiency in two studies ($N = 89$), but no significant effects on learner performance in terms of final grades or learning gains. On the other hand, Rogers et al. [213] investigated "discouragement by peer excellence" in a MOOC setting and concluded that learners who are exposed to examples of excellent peer achievements risked feeling less capable of performing at the level of those peers. The `Feedback System` is different in that it shows the *behavior* patterns of the average completing learner, so as not to risk discouragement.

The present research adds to the literature on social comparison in the online learning environments by investigating the effects of forced comparison of learners' performance and engagement in a MOOC setting. With the `Feedback System`, MOOC learners can visualize their behavior compared to that of successful learners, offering them a model against which they can evaluate their own study habits.

Culture MOOC learners come from all over the world and cover a profoundly wide range of cultural contexts. Prior MOOC research has observed a learner's culture as affecting behavior within the course. For example, Liu et al. [169] explored patterns in MOOC learner behavior in relation to Hofstede's cultural dimensions [99]. The authors clustered countries based on similarity across four cultural dimensions and found significant variation in learner behavior between the clusters. Moreover, Kizilcec et al. [128] found in two randomized experiments in MOOCs that the effect of a self-regulation intervention depended on learners' cultural context ($N = 17,963$): only learners in individualist countries benefited from the brief writing activity. Thus, prior work supports the hypothesis that cultural factors shape learner behavior in MOOCs. We examine two country-level cultural dimensions: individualism [99] and tightness [90].

Hofstede's dimension of individualism-collectivism characterizes cultural variation around the world. Cultures high in individualism are those which emphasize the individual as an independent actor with loose social relations. Cultures high in collectivism are characterized by tightly-knit social relations and shared responsibility for the collective well-being [99]. Gelfand et al. conceived an index that ranks countries by their cultural tightness: tight cultures are those with "strong norms and a low tolerance of deviant behavior," and conversely, cultures of low tightness (or loose cultures) are those with "weak social norms and a high tolerance of deviant behavior" [90]. The present study attempts to adapt feedback to learners' cultural context so that it resonates with the learner, facilitates internalization of the feedback, and promotes positive behavior change.

Prior work suggests that cultural differences shape people's regulatory focus, whether they are motivated by pushing for success (promotion) or by avoiding failure (prevention) [98, 170]. Members of individualist and tight cultures focus more on promotion, while members of collectivist and loose cultures focus more on prevention [170, 178]. We apply this insight in the design of our feedback framing messages to appeal to learners in different cultural contexts.

5.3 MOOC Overview

For our experiments, we employed our personalized `Feedback System` to learners across four MOOCs—all of them re-runs (i.e. not in their first edition)—provided by the Delft University of Technology on the edX platform:

`WaterX` The *Drinking Water Treatment* MOOC teaches technologies for drinking water treatment. Its second edition ran between 12 January and 29 March 2016. It is a seven-week course with 63 instructional videos and 42 summative quiz questions. A total of 10,943 learners registered for the course. To complete the course, learners had to gain at least 60% of all scores (i.e. passing threshold $\tau_{pass} = 60\%$).

`UrbanX` The *Urban Sewage Treatment* MOOC learners are taught how to design and manage solutions for urban sewage. The second edition of the seven-week course ran between 12 April and 20 June 2016 with 8,137 learners. There are 272 summative quiz questions ($\tau_{pass} = 60\%$) and 71 videos.

`BusinessX` *Responsible Innovation: Ethics, Safety and Technology* teaches learners how to deal with risks and ethical questions arising from new technologies. 2,352 learners registered to the second edition which ran between 11 April and 14 June 2016. The course has 79 summative quiz questions ($\tau_{pass} = 59\%$) and 54 videos.

`CalcX` *Pre-university Calculus* is the only MOOC in our list that targets beginning Bachelor students and was designed as a refreshment course before entering higher education. The third iteration of this course ran from 28 June 2016 through 27 September 2016 with 12,294 learners, 85 videos, and 327 summative quiz questions ($\tau_{pass} = 60\%$).

We found the `WaterX`, `UrbanX`, and `BusinessX` MOOCs to attract a similar population of learners: two thirds of the enrolled learners were male, the median age was 28, and the majority of learners held a BSc or MSc degree. The learner population in the `CalcX` course was instead targeted at high-school students who were about to enter university. While the gender balance was consistent with other MOOCs (30% female), the median age was only 25, and the most common education level was a high school diploma (45%).

For each learner, we collected all available edX log traces such as the learners' clicks, views, dwell time on the edX platform, and their provided answers to the quiz questions.

5.4 Approach

In Section 5.4.1 we first introduce the research questions driving our work before detailing the design of our `Feedback System` which was deployed in different instantiations across the four MOOCs just described.

5.4.1 Research Questions

The first **R**esearch **Q**uestion and **H**ypotheses are based primarily on the social comparison literature in the context of education and learning environments:

RQ5.1 Does providing personalized social comparison feedback increase learner achievement and self-regulatory behavior in MOOCs?

H5.1.1 In line with previous findings [15, 200], we expect that providing learners a comparison of their own behavior to that of previously successful peers will increase learner *achievement* (measured in terms of completing/passing the course) and *engagement* (activity levels within the course environment).

H5.1.2 Learners will change the aspects of their behavior that the `Feedback System` makes them aware of.

H5.1.3 Certain feedback metrics (and combinations of metrics) will be more effective than others in leading to desirable changes in student behavior.

Based on prior work which has shown that learners from different cultural contexts learn and behave differently in MOOCs [94, 129, 134, 169], we explore:

RQ5.2 Which learners benefit most from the `Feedback System`?

We also examine the differences in learning behavior according to learners' cultural context. We expected the effects of the feedback to depend on learners' cultural context in terms of individualism and tightness, and moreover, that matching the framing of feedback to learners' culture to be beneficial:

RQ5.3 Does feedback framed in line with a learner's cultural context lead to increased achievement and self-regulatory behavior compared to a culturally mismatched framing?

H5.3.1 Learners from individualist cultures will show more engagement than those from collectivist cultures with the individual- promotional framing, while learners from collectivist cultures will show more engagement with the collectivist- prevention framing.

H5.3.2 Learners from tight cultures will show more engagement than those from loose cultures with the collectivist-prevention framing, while learners from loose cultures will show more engagement with the individual-promotional framing.

5.4.2 Feedback System Design

Recall, that our design rationale of the `Feedback System` (presented as the “Learning Tracker" to learners in the courses, cf. Figure 5.1) is to provide learners feedback about their own behavior that enables them to make well-informed decisions about their learning strategies going forward [279] as a result of increased self-awareness. The `Feedback System` can be thought of as a mirror with which learners can view and react to their own, previously-invisible behavior. Since SRL skills are generalizable, the design should be agnostic to the content of each specific MOOC the feedback system is deployed in. We identified three key criteria for our system design:

- *Traceable*: we can only provide feedback on behavior we can extract and derive from edX’s log traces[2];
- *Scrutable* [121]: afford learners the ability to intuitively understand and explore the information presented;
- *Actionable*: learners should be able to take action and change their behavior based on what they learn from the presented feedback.

After surveying the literature on learner model visualizations, we settled on employing a single spider chart to visualize six metrics of learners’ behavior in relation to that of their successful peers, as shown in Figure 5.1. The spider chart’s key benefits include: (i) a single, embodied representation of multiple metrics, (ii) numerous indicators displayed in a small space, (iii) a simple representation of metrics—data is shown as single points along radial straight lines, and (iv) easily comparable—information is represented as differently colored areas that can be layered [219].

In all four courses, the experimental conditions were not made explicitly known to the learners; the `Feedback System` appeared seamlessly integrated with the rest of the course materials.

We operationalized previously successful students, or “role models", as learners who earned a passing grade in the *previous edition* of the MOOC

[2]edX provides fine-grained log traces of each learner’s clicks & views, provided answers to assignments, forum interactions, etc.

(note that this setup requires that subsequent editions of the same MOOC have few changes). We updated the `Feedback System` every week (based on the learners' activities on the platform in *all* weeks leading up to the current one) so that the learners could see an up-to-date representation of their activities as compared to that of the role models. The learners' behaviors in the courses were tracked by the standard edX tracking log system.

In each MOOC, the `Feedback System` was placed in the *Weekly Introduction* unit of each course week so that it would be readily available and immediately visible to learners upon entering the new course week, enabling them to reflect on their SRL behavior so far. With the exception of the `Feedback System`, all learners received the same course materials, independent of the experimental condition.

Feedback metrics Table 5.1 shows an overview of the feedback metrics given to learners in each MOOC. At the end of each week in the course, the metrics were computed based on the log traces of *all* weeks prior. These metrics were chosen based on the following criteria: relevance to self-regulated learning, clarity/intuitiveness to the learner, and availability in the log data. For each metric, all values of previously successful learners were sorted and the top 5% and bottom 5% of values were discarded to remove outliers. The mean of the remaining values was computed, yielding a single value per metric — we consider this mean to be indicative of the tendency of the whole successful group of learners. We operationalize "sessions" as strings of activity with less than an hour gap between two events. As shown in Table 5.1, we used different feedback metrics in different MOOCs to explore the impact of the choice of metrics (**H5.1.2** and **H5.1.3**).

Feedback System Alterations Apart from the different metrics, we also explored three refinements of the `Feedback System`:

1. Planning ahead: in `WaterX` the learners only received feedback about their behavior up to now and how it compares to that of successful learners. In this alteration (in `UrbanX` and `BusinessX`), we also provide the learner with a visualization of the role models' behavior (labelled as "Average graduate this week" in Figure 5.1) in the upcoming week, enabling learners to plan ahead instead of only reflect.
2. *Interactive visualization*: instead of a static feedback image (as provided in `WaterX`), in this alteration, we provide learners with an interactive visualization they can explore, i.e. mouse over the metrics to reveal exact

Table 5.1: Overview of feedback metrics and alterations presented to learners in each MOOC. A • indicates the presence of the metric/alteration.

	WaterX	UrbanX	BusinessX	CalcX
Feedback metrics				
Quiz submission timeliness (days)	•	•	•	•
Time on the platform (in hours)	•			
Time watching videos (in hours)	•			
Number of videos accessed	•			
Number of quiz questions attempted	•	•	•	•
Proportion of time spent on videos while on the platform (in %)	•			
Average time on the platform per week (in hours)				•
Number of revisited video lectures				•
Number of forum visits		•		
Number of forum contributions				•
% of time spent on quizzes				•
Number of sessions per week		•	•	
Mean session length (in minutes)		•	•	
Mean time between sessions (in hours)		•	•	
% of time-on-task - time spent on video-lecture, quiz or forum pages			•	
Alterations				
Interactive visualization		•	•	•
Planning ahead		•	•	•
Feedback framing				•

numbers and comparisons (cf. Figure 5.1), and toggle on/off the metrics of the average successful learner for the upcoming week.

3. *Cultural framing*: in the first three MOOCs, the `Feedback System` provides no written interpretation of the visualization; instead learners are left

to draw their own conclusions. In `CalcX` we additionally provide an explanatory text (as shown in Figure 5.1) that offers a clear interpretation of the learner's "on-trackness".

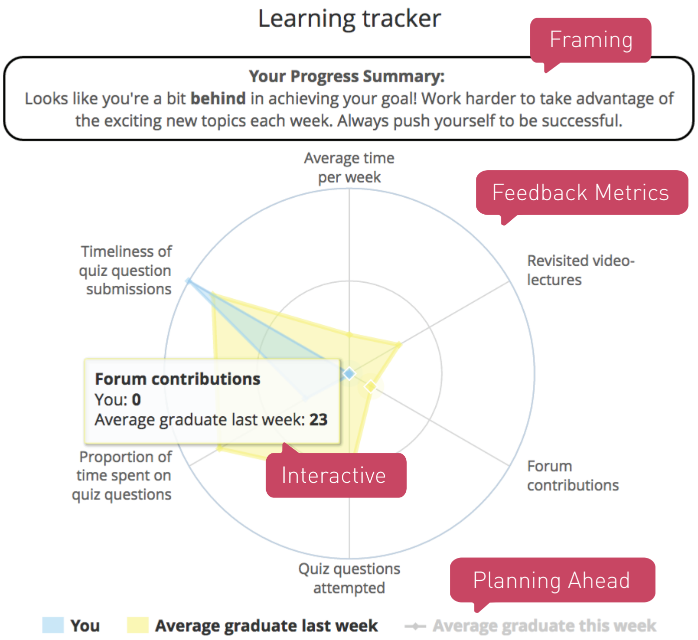


Figure 5.1: The `Feedback System` as shown in the individualistic-promotional condition in `CalcX`, annotated for clarity.

5.4.3 Studies

In each MOOC, we deployed a variation of the `Feedback System`. Table 5.1 summarizes the feedback metrics and variations deployed. For random assignment, we used a between-subjects design, where learners were assigned to either the control or a treatment condition and remained in this condition throughout the study. Table 5.2 shows a breakdown of the number of learners assigned to each condition for each MOOC. To gather baseline data from the first two weeks of each course, we released the `Feedback System` in the treatment conditions in the third week in each experiment. As noted before, the `Feedback System` is then updated on a weekly basis to reflect the updated learner activity data.

In the control condition across all experiments, learners did not receive the `Feedback System`. However, the edX platform offers a very basic form of learner feedback: a learner can visit her "progress" page and view the number of points scored so far in the course. This progress page is available to all learners, independent of their condition assignment. In the treatment condition, learners received the `Feedback System` in addition to edX's progress page.

Table 5.2: Overview of the number of learners enrolled and assigned to the control and treatment groups respectively. The number of active learners (having spent at least 5 minutes in the course platform) is in the parentheses beneath.

	`WaterX`	`UrbanX`	`BusinessX`	`CalcX`
Enrolled	10,943 (2,519)	8,137 (1,517)	2,352 (324)	12,294 (3,415)
Control Group	5,460 (1,268)	4,038 (771)	1,184 (164)	4,142 (1,150)
Treatment Group 1	5,483 (1,251)	4,099 (746)	1,168 (160)	4,087 (1,147)
Treatment Group 2	–	–	–	4,065 (1,118)

In all but one study there is only one treatment condition. In `CalcX`, we had two treatment conditions, one for each culture-specific framing of the explanatory feedback text:

- `CalcX` *treatment 1* received text with an individualistic promotion-focused framing;
- `CalcX` *treatment 2* received text with a collectivist prevention-focused framing.

We determined each learner's cultural context based on the IP address used to access the course relying on Maxmind's GeoIP lookup database[3], as not all learners self-report their nationality. For learners with more than one IP address used, we consider the first one they used to access the course as their country.

We developed a strong manipulation of the culture-specific framing by drawing on the cultural difference in (1) individualistic vs. collectivist appeals (collectivist cultures see the self embedded in a relational network, while the self-concept is more independent in individualist cultures), and (2)

[3]http://www.maxmind.com

prevention- vs. promotion-focus (the prevention of negative outcomes is emphasized over the promotion of positive outcomes in collectivist cultures, and vice versa for individualist cultures) [98, 171, 178]. We designed two texts for each treatment group: one for learners who are "on track" (characterized by exhibiting similar behavior to that of the role model learners) and one for learners who are "behind" (characterized by exhibiting less course engagement compared to the role model learners). The texts (four overall) and how those texts align with a particular framing are shown in Table 5.3. The learners were evaluated as "on-track" or "behind" based on their *on-trackness score*, OT. The on-trackness score quantifies the similarity between a learner's behavior and that of the previously successful learners: we normalize each metric to a value in the range $[0, 10]$ (chosen for convenience to work well in the spider chart setup) and then compute the difference, d_i, between the learner's score on metric m_i and the previously successful learners' average score on m_i. If $d_i \leq -1 \; \forall m_i, i = \{1, .., 6\}$ the learner is classified as behind, otherwise she is on-track — this is a very conservative classification, the learner has to have a lower engagement level on every single metric before she is considered as being behind.

Table 5.3: Overview of the supplementary texts the treatment groups received in `CalcX`, depending on their performance in the course so far (either on track or behind). The alignment of the words and phrases with the intended framing is highlighted. Sentences prefixed by ↱ are directly addressed at the individual (individualistic framing). Best viewed in color.

	Treatment Group 1 (individualistic promotional framing)	**Treatment Group 2** (collectivist prevention framing)
On track	*Looks like you're right on track to achieve your goal! ↱Keep taking advantage of the exciting new topics each week. Always push yourself to be successful .*	*Looks like you're keeping up with the course for now ! We're doing our best to introduce you to exciting new topics each week. Please don't let us down now .*
Behind	*Looks like youre a bit behind in achieving your goal! ↱ Work harder to take advantage of the exciting new topics each week. Always push yourself to be successful .*	*Looks like you're a bit behind in the course right now! We're doing our best to introduce you to exciting new topics each week. Please don't let us down now .*

The study design and all analyses conducted as part of the `CalcX` experiment[4] were pre-registered through the Open Science Framework, vetted, and approved to meet the requirements of the Center for Open Science Preregistration Challenge[5]. All manuscripts, data, and scripts used for analysis are available at: http://osf.io/ys6au.

5.4.4 Measures & Method of Analysis

The primary outcome variable that we targeted with the design of our `Feedback System` is course completion, which indicates that a learner achieved the required minimum passing score on all summative quiz questions and thus earned a certificate. Course completion demonstrates sustained commitment to the course and mastery over the course material. The `Feedback System` is designed to support this type of sustained commitment and learning, even if individual student intentions may vary. The secondary outcome is to promote SRL and meta-cognitive awareness. While many SRL processes are meta-cognitive and remain unobserved, it is possible to infer some of them based on learner's logged actions with the course materials [112, 134, 230, 231]; for example, goal-setting & planning, time management, self-monitoring, and social comparison.

For non-binary measures, to test if differences between experimental conditions are statistically significant, we used the non-parametric Kruskal-Wallis test, because these measures were not normally distributed and exhibited unequal variances across conditions. For binary measures, we tested differences in proportion using a χ^2 test. We present the results of each test by each group's mean and median along with the χ^2 value, degrees of freedom, and level of statistical significance. Due to the commonly high levels of attrition in MOOCs (65%-74% of learners never returned to the course after enrolling in one of our four MOOCs), the subsequent analyses only consider data generated by *active learners*. We define active learners as those having spent *at least five minutes* on the course platform. See Table 5.2 for the breakdown of registered vs. active learners per MOOC.

[4]We pre-registered this experiment because it was the fourth and final study of the present research and included an added manipulated variable in the cultural framing.

[5]https://cos.io/prereg/

5.5 Results

We present our findings for the **R**esearch **Q**uestions outlined in Section 5.4.1. We discuss the impact of the `Feedback System` on course completion and engagement in Sections 5.5.1 & 5.5.2, heterogeneous treatment effects of the `Feedback System` in Section 5.5.3, and lastly in Section 5.5.4, we compare the effects for different cultural framings of the feedback.

5.5.1 Course Completion

We hypothesized that the `Feedback System` will increase learner achievement in terms of course completion (**H5.1.1**). Table 5.4 shows the completion rates in all conditions for the first three experiments. The completion rate is consistently higher in the treatment condition than in the control condition in all experiments. Pooling across experiments, we observed an increase in the completion rate from 15.5% to 18.9% ($\chi^2 = 5.87, p = 0.008$). Thus, regarding hypothesis **H5.1.1**, we conclude:

> The `Feedback System` significantly increases course completion rates in MOOCs.

Table 5.4: Course completion rates across the first three studies among the *active learners*. Overall, the difference in completion rate between the groups is statistically significant ($p = 0.008$).

	Condition	N	# Pass	Pass Rate
`WaterX`	**Control**	1,268	160	12.6%
	Treatment	1,251	188	15.0%
`UrbanX`	**Control**	771	136	17.6%
	Treatment	746	165	22.1%
`BusinessX`	**Control**	164	46	28.0%
	Treatment	160	54	33.8%
Overall	**Control**	2,203	342	15.5%
	Treatment	2,157	407	18.9%

In the fourth experiment, which tested two treatment conditions with different cultural framings against the control of not providing the `Feedback System`, we also observed higher completion rates in the treatment conditions (Table 5.5). However, this difference was not statistically significant ($ps > 0.25$). However, the overall completion rate in the `CalcX` course was

Table 5.5: Course completion rates in the `CalcX` course among *active learners*. A binomial test of independent proportions revealed no statistically significant differences between the three conditions.

Condition	N	# Pass	Pass Rate
Control	1,150	45	3.91%
Indiv.-Promotion	1,147	62	5.41%
Collect.-Prevention	1,118	51	4.56%

extremely low (1.7%). This suggests that the sample is drawn from a population of less committed learners and that potential effects could be obfuscated by high levels of unexplained variance in completion outcomes. Another contributing factor to this rift between `CalcX` and the other three courses is the fact that `CalcX` was self-paced (content released all-at-once), whereas the others were instructor-paced (content released weekly), thus providing less structure/support to the learners.

Moreover, we hypothesized that showing certain combinations of feedback metrics will better promote positive changes in behavior than others (**H5.1.3**). We explored this by changing the (combination of) metrics in each of the four iterations of the `Feedback System` (see Table 5.1). Given that the course completion rates increased across all four iterations each with a different combination of feedback metrics (with two of the six metrics—quiz submission timeliness (how far ahead of the deadline responses were submitted) and quiz questions attempted—were present in all four) we conclude:

> Each combination of metrics shown to the learners produced increases in completion.

5.5.2 Engagement

In light of the positive effect of the `Feedback System` on course completion, we next evaluated specific changes in learner behavior corresponding to the behavioral metrics that were visualized in the `Feedback System` (**H5.1.1**). These metrics, which varied across experiments, were most likely to be directly affected through social comparison. Table 5.6 shows the results of Kruskal-Wallis tests[6] comparing the various feedback metrics between the treatment and control groups in study to test **H5.1.2**. A common thread across the three experiments was that of the `Feedback System` increased the

[6] While the Kruskal-Wallis test measures the difference between rank orders, the median values are often zero, so in the table we show the mean for better context.

number of summative quiz questions that learners submitted, which directly promotes course completion.

Table 5.6: Results of the Kruskal-Wallis tests for the behavior metrics (feedback metrics) provided in the `Feedback System` for `WaterX`, `UrbanX`, and `BusinessX`. Statistically significant differences are in bold.

	Metric	**Ctrl** $\bar{x}$	**Treat.** $\bar{x}$	χ^2	p
`WaterX`	quiz questions attempted	4.2	4.7	4.46	**0.04**
	videos accessed	7.0	7.0	0.01	0.94
	time on platform (hours)	4.5	4.6	0.17	0.68
	time watching videos (hours)	0.8	0.8	0.04	0.85
	ratio video/total time (%)	25.0	25.0	0.11	0.75
	submission timeliness (days)	27.9	31.3	4.20	**0.04**
`UrbanX`	quiz questions attempted	5.7	6.6	3.16	0.08
	sessions per week	3.8	4.0	2.11	0.15
	avg. session length (minutes)	8.1	8.2	0.18	0.67
	time between sessions (hours)	117.0	120.0	0.29	0.59
	forum visits	2.7	3.0	2.88	0.09
	submission timeliness (days)	28.3	32.0	3.27	0.07
`BusinessX`	quiz questions attempted	21.4	25.3	3.97	**0.05**
	sessions per week	0.5	0.7	4.89	**0.02**
	avg. session length (minutes)	32.8	46.7	8.42	**0.00**
	time between sessions (hours)	95.9	92.2	1.17	0.28
	time-on-task (%)	64.3	67.5	0.32	0.57
	submission timeliness (days)	19.9	21.4	1.12	0.29

Looking at each feedback metric individually in Table 5.6, we observe 15 out of 18 times an improvement from control to treatment condition[7]; three times no change is observed. The treatment condition does not lead to a worse effect in any feedback metric. While only a handful of these differences are statistically significant, this consistency lends itself to some explanatory power over the statistically significant increases in course completion rates: while on an individual level, only some metrics show significant increases as a result of the `Feedback System`, on a macro level—that which accounts for a learner's *overall* activity in the course—we infer that these small increases in engagement all effectively coalesce into a boost in desirable behavior that leads to increased completion rates. We draw the following conclusion:

> The `Feedback System` causes desirable changes in learner *engagement*.

[7] A high "time between sessions" score is not better per se, but it indicates a desirable high-spacing learning routine

Table 5.7: Results of the Kruskal-Wallis tests for `CalcX`. Statistically significant differences indicated in bold. "T" indicates Treatment Groups.

Metric	Ctrl $\bar{x}$	T1 $\bar{x}$	p	Ctrl $\bar{x}$	T2 $\bar{x}$	p	T1 $\bar{x}$	T2 $\bar{x}$	p
mins/week	31.9	33.4	0.24	31.9	33.8	0.74	33.4	33.8	0.14
revisited lectures	3.44	3.62	0.44	3.44	3.42	0.55	3.62	3.42	0.17
forum posts	0.34	0.53	0.86	0.34	0.36	**0.05**	0.53	0.36	0.06
quiz attempts	31.3	32.4	0.22	31.3	33.5	0.77	32.4	33.5	0.13
% time on quizzes	37.0	34.0	**0.02**	37.0	36.4	0.70	34.0	36.4	0.07
timeliness (days)	47.48	45.70	0.25	47.48	46.71	0.36	45.70	46.71	0.83

Table 5.7 shows the results of the same analysis on the engagement metrics across the three conditions in `CalcX`; the results are less consistent.

In **H5.1.2**, we hypothesize that learners change aspects of their behavior that are reflected back to them in the `Feedback System`. Since there is no consistency among significant increases in the provided behavior metrics, we conclude:

> Learners do not change specific behaviors based on what metrics are shown in the `Feedback System`.

5.5.3 Who benefited from the feedback?

Going beyond average treatment effects of the `Feedback System`, we now evaluate heterogeneous treatment effects, that is, how the feedback affects different groups of learners (**RQ5.2**). Specifically, we focus on heterogeneity by prior education level, as this might determine learners' ability to use the information provided in the `Feedback System`. We gather learners prior education levels from their edX user profile; learners who do not report their education level are omitted from this analysis. We define *high prior education* learners as those with a Bachelors, Masters, or PhD degree, and *low prior education* learners as those with any degree below Bachelors. Table 5.8 compares the average final grades in the control and treatment conditions of the first three courses separately for high vs. low prior education learners.

In `WaterX`, `UrbanX` and `BusinessX` we observed a consistent increase in final grades for highly educated learners, but not for less educated learners. However, this pattern did not replicate in the `CalcX` course, as education level did moderate the effect on grades ($p = 0.82$)[8]. Nevertheless, the results for `CalcX` are harder to interpret due to the relatively low completion rate in this course. Moreover, `CalcX` stands out in that a majority of low prior

[8] Once more we report `CalcX` separately due to the overall difference in completion rate compared to `WaterX`, `BusinessX` and `UrbanX` as shown in Tables 5.4 & 5.5.

Table 5.8: Mean final grades (out of a possible 100 points) grouped by prior education levels. The "Prior Education" column indicates the highest degree the learner has earned; "N" is the sample size; and "p" shows the result of a Kruskall-Wallis test. Significant values are in bold.

Course	**Prior Education**	**N**	**Ctrl** $\bar{x}$	**Treat.** $\bar{x}$	p
`WaterX`	High	2,006	13.2	15.7	0.15
	Low	788	11.8	11.5	0.16
`UrbanX`	High	1,337	17.4	21.3	**0.04**
	Low	438	16.4	14.0	0.66
`BusinessX`	High	299	23.7	29.1	**0.04**
	Low	92	21.4	22.8	0.78
OVR	High	3,642	16.3	19.5	$<$**0.01**
	Low	1,318	14.4	13.6	0.36

education learners were enrolled in this course, while the `WaterX`, `UrbanX` and `BusinessX` courses had a majority of high prior education learners. Based on these analyses, we conclude that:

> The `Feedback System` only helps to improve the *achievement* (final grade) of learners who are already highly educated.

This finding suggests three possibilities: (i) the `Feedback System` is too complex for people falling in the low prior education category to understand, (ii) highly educated learners are better able to synthesize the information offered by the `Feedback System` and translate it into positive behavior as they are already experienced learners (with at least some SRL skills), and/or (iii) less educated learners are not concerned with obtaining a certificate, but rather focus on knowledge acquisition.

5.5.4 Framing Feedback to Cultural Contexts

In the `CalcX` course, we tested **H5.3.1** and **H5.3.2** about supplementing the `Feedback System` with culture-specific feedback. As before, we evaluated each hypothesis both in terms of learner achievement and engagement. All pre-registered analyses for this experiment are reported in Section 5.5.4. Additional exploratory analyses are reported in Sections 5.5.4 and 5.5.4.

Pre-registered: Completion & Engagement

We compared the completion rate and six behavioral measures (the ones shown in the `Feedback System`) between the treatment and control conditions separately by learners' cultural context. To address **H5.3.1**, we segmented learners into three groups of individualism—*high*, *balanced*, and *low individualism*—and compared completion rates of learners in high vs. low individualism cultures in each condition. There was no significant increase in completion rates for either feedback framing, neither for learners in low individualism cultures nor for those in high individualism cultures (all $p > 0.12$). Likewise, we tested for treatment effects in contexts defined by cultural tightness (**H5.3.2**) and also found no significant increase in completion rates (all $p > 0.29$). Results for learner engagement were also not significant (cf Table 5.7). Finally, we tested the moderating role of education level, as in the prior experiments (**RQ5.2**), but found no evidence in support of moderation ($\chi^2 = 0.40, p = 0.82$). We thus conclude that:

> Supplementing the `Feedback System` with feedback framing tailored to cultural tendencies of individualism and tightness does not increase learners' course *achievement* or *engagement*.

Increased "Active" Threshold

From the exceptionally low completion rate of `CalcX`, we gathered that a high proportion of uncommitted learners rendered the data set noisy. Whereas the `WaterX`, `UrbanX`, and `BusinessX` experiments yielded a consistent main effect on course completion, this effect was not detectable in the `CalcX` experiment. To focus our analysis in `CalcX` on more committed learners, we imposed a stricter threshold for "active" learners. Considering only learners who accessed the course platform for at least an average of 1hr/week, we proceeded by analyzing data for *highly active* learners ($n = 658$). This threshold is reasonable given the amount of course content per week (between 6–8 hours). Moreover, the overall completion rate in this sample was 15.65%, a similar rate as in the other experiments.

Among highly active learners, we find that the individualist framing increased completion rates regardless of a learner's own cultural context from 12.8% in the control condition to 19.9%, a 7.1 percentage point increase ($t = 2.02$, $p = 0.04$). Moreover, we find that the effect of the individualistic framing was especially large for learners in tight cultures, effectively tripling the completion rate from 12.1% in the control condition to 36.3% ($t = 2.07$, $p = 0.04$). We therefore conclude that:

> The individualist framing was most effective in increasing course completion rates overall, and especially for learners in tight cultures.

The effect of the individualist framing is surprising in terms of its large magnitude and cultural heterogeneity. We expected the individualist framing to resonate in loose rather than tight cultures. Perhaps the individualist framing is more congruent in an environment where learners tend to be anonymous and socially isolated. Learners in tight cultures were also more likely to benefit as there course performance was generally lower, as discussed next.

Lower Achievement in Tight Cultures

In the preceding analyses, we observed a notable cultural difference along the tightness dimension. Pooling across experimental conditions in the `CalcX` course, we found for every metric (cf. Table 5.1) with the exception of *number of forum posts* that learners in tight cultures exhibit significantly higher levels of achievement and engagement than those in loose cultures (*ps* $\leq$0.02). We repeated the analysis for the other three courses and found the same cultural differences. We therefore conclude:

> Learners from countries with low cultural tightness significantly outperform their peers from countries of high cultural tightness in terms of both *engagement* (all p-values $p \leq 0.1$) and *achievement* ($p \leq 0.02$).

This cultural difference in performance could arise from the nature of the MOOC learning experience. MOOCs provide significant latitude for different levels of commitment and engagement; in fact, learners can come and go as they please at no cost. This may especially appeal to loose cultures, where there are few strongly-enforced rules and high tolerance for deviation. In contrast, traditional classroom environments with strict attendance and performance policies would align more with the ideals of tight cultures. Alternatively, the current finding may reflect structural differences that are associated with both tightness and performance, such as infrastructure and education levels.

5.6 Conclusion

This research tested the effect of providing online learners with personalized feedback in four large-scale randomized controlled experiments in MOOCs. The `Feedback System` was designed to promote learners' awareness of both their own SRL behavior and that of their successful peers through social comparison. It significantly increased course completion rates across different courses. The combination of behavior metrics that was shown to learners in the `Feedback System` did not determine the significance of the effect on course completion, highlighting a need for further research on the optimal set of metrics to show. Moreover, we discovered that the `Feedback System` primarily benefited highly educated learners, although the system was envisioned to support those who struggle with self-regulation. This suggests a new challenge for MOOC researchers and designers to make targeted interventions that support learners who are less educated and need more support.

As online courses can be culturally diverse learning environments, we investigated how the `Feedback System` could be adapted to resonate with learners from different backgrounds. Our pre-registered analyses yielded no significant effects of changing the cultural framing of the feedback. In exploratory analyses, however, we found strong benefits of framing feedback with an individualistic and promotion focus. This insight warrants further research to establish its generalizability. Aside from our intervention, we found that learners from loose cultures consistently outperformed learners tight cultures in terms of course engagement and final grades. In light of the two sources of heterogeneity we identified, future MOOC interventions may be strengthened by personalization based on learners' prior education level and cultural context.

In future work, we plan test a different feedback interface design that presents a set of different *personas* that learners can identify with, such as person who works a bit every day and one who works a lot over the weekend.

We will also evaluate new approaches for feedback messages to better support learners with different cultural and educational backgrounds.

Chapter 6

SRLx: A Personalized Learner Interface for MOOCs

Past research in large-scale learning environments has found one of the most inhibiting factors to learners' success to be their inability to effectively self-regulate their learning efforts. In traditional small-scale learning environments, personalized feedback (on progress, content, behavior, etc.) has been found to be an effective solution to this issue, but it has not yet widely been evaluated at scale. In this chapter we present the `Adaptive Retrieval Practice System` (`ARPS`), an interactive widget that we designed and open-sourced to improve learners' self-regulated learning behavior in the Massive Open Online Course platform edX. `ARPS` enables learners to plan their learning on a weekly basis and view real-time feedback on the realization of those plans. We deployed `ARPS` in a renewable energies MOOC to more than 2,900 active learners and performed an exploratory analysis on our learners' SRL behavior.

This chapter is based on two full conference papers, published as "Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-Regulated Learning Strategies at Scale" [65], by D. Davis, G. Chen, van der Zee, Tim, C. Hauff, and G.J. Houben in Proceedings of the 11th European Conference on Technology-Enhanced Learning, 2016. and "SRLx: A Personalized Learner Interface for MOOCs" [70], by D. Davis, V. Triglianos, C. Hauff, and G.J. Houben in Proceedings of the 13th European Conference on Technology-Enhanced Learning, 2018.

6.1 Introduction

Large-scale learning environments open up world-class educational resources to the masses. With this unprecedented scale and reach, however, come new challenges in enabling learners of diverse backgrounds to excel given the unfamiliar context of the massive online classroom. Low course completion rates—dropout rates of 95% are not uncommon [144]—highlight the need for additional support in MOOCs. Past research in this space, e.g. [129, 131, 130, 274] has explored the problems learners face when trying to succeed in these self-directed learning environments. Learners are often unable to find the time to keep up with a course, an issue related to insufficient self-regulatory abilities [129, 274]. *Self-regulated learning* (SRL) is the ability to plan, monitor, and actively control one's learning process. The discipline to plan and follow a self-imposed studying regime is a skill that is learned over time and associated with a higher likelihood of achieving self-set course goals in MOOCs [134, 167]. Learners who were exposed to such training during their studies tend to be more successful in MOOCs than learners without a tertiary education background. The latter though is a target population that is vital to keep the original vision of MOOCs alive: making higher education accessible to those that do not enter the traditional tertiary education system. Learners need tools that enable them to learn *how* to learn.

Today's MOOC platforms (such as Coursera and edX) are not designed in a way that encourages learners to explicitly plan or monitor (with the help of feedback) their learning activities [92]. In general, learners are exposed to very few feedback moments to support their SRL processes.

Yeomans and Reich [272] found that a single planning prompt at the start of a MOOC can positively influence learning outcomes. We have expanded upon this concept first by conducting a pilot study to replicate their simple planning prompt and then by designing and developing the `Personalized SRL Support System`[2] (`SRLx`), an interactive widget for the edX platform that allows learners to explicitly express their motivation, *plan* their learning, *monitor* their progress towards their set goals at any point in time, and *reflect* on them. `SRLx`'s design was based on educational theories and findings in the SRL literature.

We deployed `SRLx` in a MOOC on renewable energies offered by the Delft University of Technology in 2017 with more than 2,900 active learners and empirically evaluate the following research questions:

[2]Open-sourced at https://github.com/dan7davis/Lambda.

RQ6.1 To what extent do MOOC learners adopt and take advantage of a personalized SRL support tool?

RQ6.2 Does `SRLx` support MOOC learners in promoting effective self- regulated learning behavior?

Along with the contribution of an open-sourced system architecture that provides SRL support at scale, we present the following key findings from our analysis of learners' SRL behaviors:

- As the course progresses, learners are able to plan their time commitment more effectively.
- Learners are more conservative with the way they plan to commit time to the course compared to video and quiz activity planning.

6.2 Related Work

Zimmerman et al.'s model of self-regulated learning [278] comprises three cyclical phases: forethought, performance, and self-reflection. Learners first formulate a plan for their learning activities, they then carry out and act according to their plan, and finally they look back at their behavior and examine their strengths and areas for improvement. In this section we first examine self-regulated learning research in the classroom and then delve into SRL studies conducted within MOOCs.

Self-regulated learning in the classroom

Goal setting has been shown to be an important factor across all levels of education. Past research has investigated to what extent aspects such as *who* sets the goals, *when* are they set, *what* goals are set and *why* are those set influence the effectiveness of goal setting. While these studies have been conducted across a range of education levels, they have all taken place in the traditional classroom or lab setting.

Schippers et al. [221] showed that engaging and teaching undergraduate students about goal setting at the beginning of their studies has a positive impact across a prolonged period of time—after one year, a 98% reduction in the gender achievement gap and a 38% reduction in the ethnicity achievement gap was observed compared to the previous year's cohort of students.

At the secondary education level, Zimmerman et al. [280] found that social-studies class students perform better (as measured by their final grade) when they set their own goals and benchmarks, than when having those imposed on them by teachers. Regularly reviewing and reflecting upon one's study goals and behaviors was found by Sagotsky et al. [220] to be significantly more effective in terms of grades and study behavior than just setting goals in a user study with primary and middle school students. A similar result was found by Mahoney et al. [173] among 27 undergraduate students who were assigned to one of three experimental conditions while preparing for an exam: (i) continuous self-monitoring, (ii) intermittent self-monitoring, and (iii) receiving instructor feedback. In line with [220], students who performed self-monitoring exhibited higher levels of engagement and achievement than students who did not.

Self-regulated learning in MOOCs

Due to the massive nature of MOOC platforms (supporting millions of learners), a large part of the platform development effort has to be spent on continued scalability. This leaves little time and attention for advances in platforms' instructional designs. Prior research in the MOOC setting has so far focused on learner surveys (to elicit their SRL needs), pre-course SRL interventions, MOOC forum interventions, and the notion of learner feedback [67].

Nawrot and Doucet [189] and Hood et al. [100] surveyed MOOC learners about their experiences taking MOOCs. Proper time management was found to be a major hindrance for many MOOC learners [189]. The ability to self-regulate one's learning was found to vary depending on learners' professional backgrounds: higher-educated learners are better able to regulate their learning (including time management) than lower-educated learners [100].

Providing learners with visualizations of their progress enables them to *reflect* upon their learning, and an emerging body of research has begun to empirically evaluate the effectiveness of such feedback [21, 22, 66, 111]. Over time, this reflection should improve learners' use of SRL strategies [33, 93].

One interesting finding by Kulkarni et al. [148] pertains to the timeliness of feedback and its impact on MOOC learners' final grades: feedback (in this case on in-progress assignments) received within 24 hours after assignment submission improves learning outcomes; if the feedback is delayed beyond this point, learners do not benefit from it. According to Davis et al. [66], enabling learners to reflect weekly on their learning behavior in comparison to that of their successful peers (i.e. feedback through social comparison) led to a significant increase in passing rates among learners with high levels of prior education (Bachelor degree or higher). A drawback of this work is the need for a successful cohort to compare against and the fact that learners cannot establish their own plans and goals.

Goal setting and feedback are important techniques to improve learning outcomes in the traditional classroom. In the MOOC setting, SRL interventions have so far either been restricted to pre-course interventions or feedback. We here investigate the effect of regular planning and goal setting in the MOOC setting.

6.3 Study Planning Pilot Study

To evaluate the efficacy of interactivity in study planning interfaces, we first conducted a randomized controlled trial pilot study in one MOOC where, in each course week's introduction section, we prompted learners to enter their plans for the week in a plain text box with the following prompt:

> *In the space below, please describe, in detail, your study plan and desired learning objectives for the week regarding your progress: e.g.*
>
> - *I plan to watch all of the lecture videos.*
> - *I will write down questions I have about the videos or assignments and discuss them in the forum.*

The initial prompts were bookended by a reflection prompt at the end of each week in which learners were instructed to reflect on their planning and execution:

> *How closely did you follow your study plan from the beginning of the week? Did you successfully meet all of your learning objectives? In the space below, explain how you can improve upon your study habits in the following weeks in order to meet your goals.*

The effectiveness of this study planning prompt treatment was evaluated in `Industrial Biotechnology`, a 7-week MOOC that introduced learners to basic biotechnology concepts. `Industrial Biotechnology` had 11,042 total enrolled learners and a 4.08% passing rate. In this experiment, 1,963 learners active in the course and were randomly assigned to either the control (received no prompts) or treatment condition (received prompts).

6.3.1 Findings

In this study planning pilot study, we found high levels of noncompliance—learners engage less with the study planning interface than course content items. Of the 998 learners exposed to the study planning modules in `Industrial Biotechnology`, 759 (76.1%) logged at least one video-watching event. Among these same learners, only 147 (14.7%) clicked on any of the study planning modules.

Study Planning

We analyzed the differences between the two experimental groups in `Industrial Biotechnology`—those who were exposed to a study planning module intervention (condition) and those who were not (control)—and found no significant differences in their final grades, course persistence, and many engagement metrics. However, we do find the following statistically significant results when narrowing the sample to compare highly engaged learners (characterized by having spent more time watching Week 1 videos than the average learner, $\approx$ 33 minutes) in the control group and the learners in the condition group who engaged with a study planning module at least once (referred to as “Study Planners").

Comparing Engagement Between Groups To determine whether there is a significant difference in the engagement levels between the highly engaged learners in the control group (N=329) and the conditioned group (those who clicked on at least one study planning module, N=146). In Table 6.1 we employ two Mann-Whitney U tests, as the data is not normally distributed, showing that the study planners have a higher session count than the highly engaged learns (U=20,070, p=0.003), as well as a higher total amount of time spent in the course in hours (U=19,983, p=0.002).

The results suggest that students who engaged with the study planning intervention are significantly more engaged with other aspects of the course as well. An alternative interpretation, however, could be that students who are highly engaged with the course also tend to engage more with the planning intervention.

Table 6.1: Results of a Mann-Whitney U test comparing the two conditions (study planners vs. highly engaged learners in the control group) in terms of two learner engagement metrics: total amount of logged sessions in the course and total amount of time spent in the course in hours.

Variable	Study Planning Median	Control Median
Session Count	25.0	19.0
Time in course (hours)	18.6	13.1

Comparing Course Persistence Between Groups We operationalize learners' persistence as the corresponding week of a learner's latest quiz submitted or video watched (slightly different from that used in [132], where persistence measured the overall amount of course materials accessed). Whereas the analyses in Table 6.1 included activity throughout the entire course, irrespective of the course week, one symptom of SRL is a learners' persistence through the course, or how many weeks the learner makes it through. We define a learner's "Final Week Reached" as the latest week in the course in which the learner either watched a video or submitted a quiz question. We ran an ANOVA to compare how far into the course learners in each group reached.

The ANOVA yielded significant results, $F(2{,}734)=21.66$, $p<0.001$. Post hoc Games-Howell tests show that the group who engaged with the study planning module (N=146, M=4.60) persisted deeper into the course than highly engaged learners in the control group (N=329, M=3.84, $p<0.001$) and highly engaged learners who were exposed to, but did not engage with, the study planning module (N=262, M=3.28, $p<0.001$).

Figure 6.1 presents a kernel density estimation plot in order to visualize the differences between groups.

Comparing Final Grades Between Groups We conducted an ANOVA to determine whether there was a significant difference in final grade between the three groups of highly engaged learners listed above. The univariate test was significant, $F(2{,}735)=17.147$, $p<0.001$. The results are presented in Table 6.2.

The follow-up Games-Howell test revealed that learners who engaged with the study planning module (M=46.42) earned higher grades than the highly engaged learners in the control group (M=36.44, $p=0.003$) and highly engaged learners who did not engage with the intervention (Non-Planners,

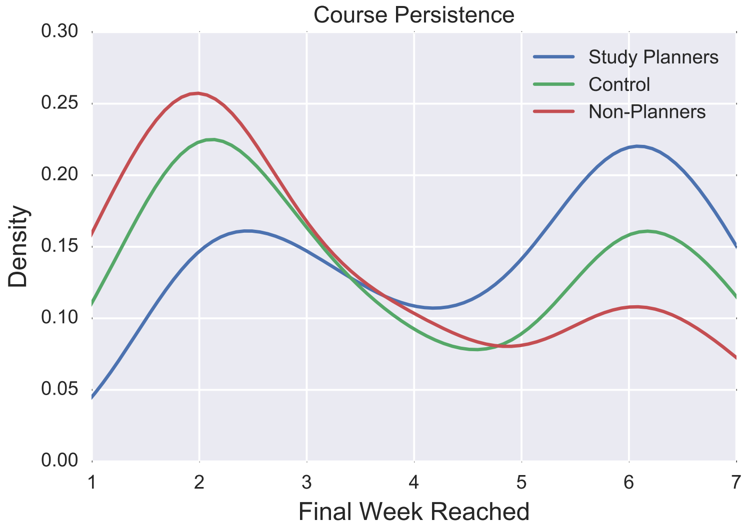


Figure 6.1: KDE plot showing the course persistence of the three groups of learners. All lines were fit using a Gaussian kernel function.

Table 6.2: Results of the ANOVA comparing final course grades among learners who engaged with the study planning module (Mean = 46.42) against those of the two other groups. A final score of 100 would indicate a perfect score.

	Group	Mean	Mean Difference
	Study Planners	46.42	—
i	**Control**	36.44	9.98
ii	**Non-Planners**	29.10	17.32

M=29.10, p<0.001). These results are visualized in Figure 6.2 and illustrate how Study Planners' final grades are higher than the others'.

Study Planners Engagement Correlations Focusing specifically on the learners who interacted with the study planning module intervention, we analyze the relationship between the extent to which they engaged with the intervention and their behavior elsewhere in the course. To do so, we computed a Pearson correlation coefficient to assess the relationship between a learner's average planning module response length (in text characters) and engagement-related variables such as: (i) total amount of time spent in the course, (ii) number of unique sessions logged, (iii) average length (in seconds) of learners' sessions, (iv) total amount of time spent watching videos, and (v)

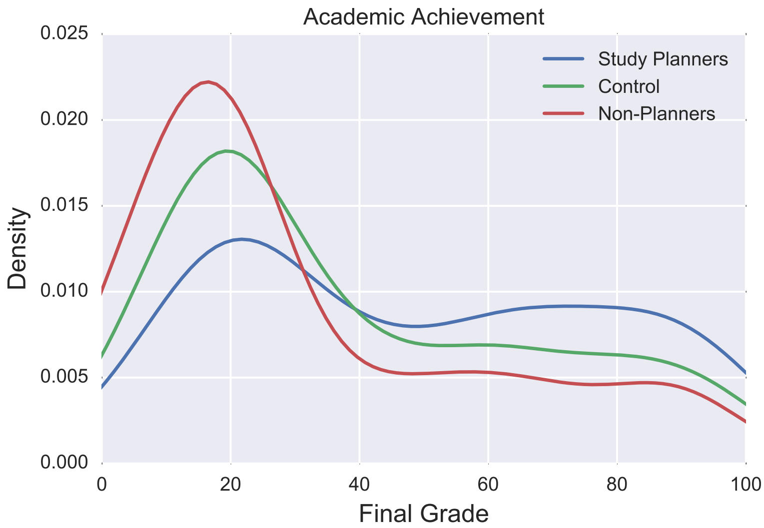


Figure 6.2: KDE plot showing the distribution of final grades earned between the three groups of highly engaged learners. All lines were fit using a Gaussian kernel function.

number of discussion forum sessions. The results are shown in Table 6.3. Two example correlations (unique sessions logged and time watching videos) are illustrated in the scatter plots in Figure 6.3 to show the slope and overall fit of the regression line. Consistent with the Pearson correlation coefficients of 0.268 and 0.346, the plots indicate positive, small-to-moderate correlations.

Table 6.3: Pearson correlation coefficient test results reporting the relationship between learners' average planning module response length and five course engagement metrics. All correlations shown are significant at the $\alpha = 0.01$ level.

Variable	Pearson Correlation	N
Total time in course	0.361	176
Session Count	0.268	176
Avg Session Length	0.346	176
Time Spent Watching Videos	0.346	170
Forum Sessions	0.305	154

The results suggest that increases in the amount of text learners write in the study planning module are correlated with small-to-moderate increases in a number of key course engagement metrics.

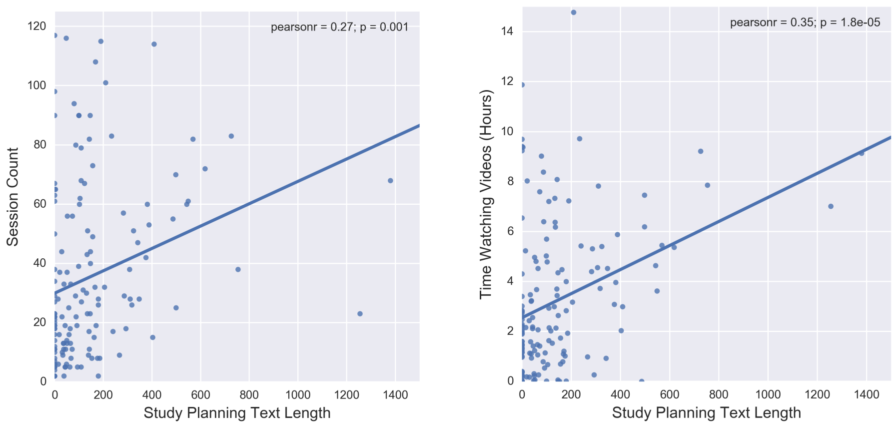


Figure 6.3: Scatterplots illustrating two example results of the five Pearson correlation coefficient tests run in order to characterize the relationship between the amount of text characters entered in the study planning module and two key course engagement metrics: session count (left) and time spent watching video lectures (right).

Overall, we find that mere *exposure* to study planning and retrieval practice interventions is not sufficient to significantly increase learner engagement or final grades. Only when narrowing the samples to learners who actually engaged with the study planning intervention do we see significant results. However, the same does not apply for learners who engaged with the retrieval practice cues, where even learners who engaged with the retrieval cues show no significant difference in any measure of performance. These findings findings highlight the need for a more interactive and engaging study planning experience—simple prompts and text boxed are insufficient for improving learning outcomes and engagement. From the insights gained in this pilot study, we designed and developed `SRLx`, a fully interactive study planning interface with real-time feedback.

6.4 SRLx System Overview

We now first describe the client-side and server-side components of `SRLx` which allow for *real-time event tracking* and then turn to the design rationales behind the four front-end interfaces we developed (cf. Figure 6.4).

Client-side

The edX platform—on which we deployed `SRLx`—allows course designers to embed and execute custom HTML, CSS, and JavaScript code in edX pages, thus enabling the creation of customized interfaces and programming logic. We take advantage of this affordance and embed our client-side code in edX's RAW HTML input elements.

We implemented two functionalities on the client-side: (i) the tracking and persisting of learners' activities to the back-end such as quiz question submissions and video watch events (cf. Section 6.5 for an exhaustive list) via AJAX and (ii) the displaying of our front-ends for goal setting, planning & feedback and the persisting of learners' interactions with them. We describe the activity tracking below and describe the interfaces in more detail at the end of this section.

Activity tracking As `SRLx` provides real-time feedback based on learners' actions on the edX platform, we had to track events such as quiz submissions and video watching events in real-time. The real-time constraint meant that we could not make use of edX's default log data setup which distributes a MOOC's daily logs in 24 hour intervals. We therefore had to track these events ourselves as follows.

edX course components, such as videos or quizzes, are implemented via XBlocks, a component architecture based on Python, HTML, JavaScript and CSS. This allows anyone to create standalone hierarchical components that may include other XBlocks. To capture user interactions, Xblocks emit and subscribe to events using an event tracking library[3]. We enable real-time event tracking by using edX's `Logger` object to subscribe to emitted events using the `listen(eventType, element, callback)` method: all Xblock fragments make use of the `Logger` object to emit events which are subsequently sent to the edX back-end via an `XMLHttpRequest`. We listen to all events of interest and forward those to our back-end.

Back-end

To store and retrieve learner data in real-time, we implemented an HTTPS server in Node.js and persisted the tracked events in a MongoDB database. The server uses a RESTful API to store and retrieve learner events. It

[3]https://github.com/edx/event-tracking

supports the JSON format for both requests and responses. Along with logging edX's learner behavior data, the `SRLx` server also logs all learner interactions with the `SRLx` interfaces.

Front-end

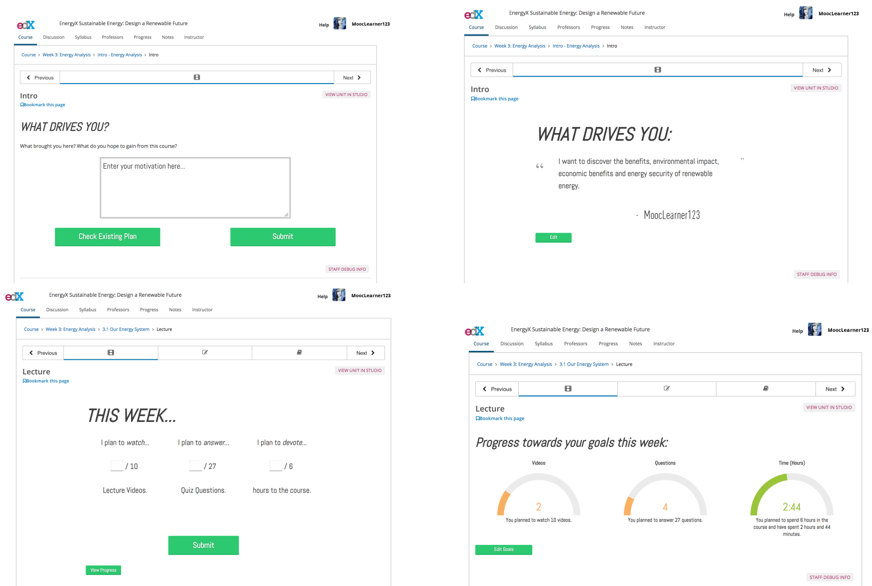


Figure 6.4: The four `SRLx` interfaces as they appear to learners on the edX platform: **motivation expression** (top-left), **motivation feedback** (top-right), **plan formulation** (bottom-left), and **plan feedback** (bottom-right).

The three phases of Zimmerman's model of self-regulated learning [278] (forethought, performance, and self-reflection) are integral to the design of `SRLx`'s four learner-facing interfaces shown in Figure 6.4: motivation expression (forethought), motivation feedback (self-reflection), plan formulation (forethought), and plan feedback (performance and self-reflection). We now discuss them in turn.

Motivation expression This interface allows us to gain an understanding of learners' motivations and overall forethought for their attitude towards the course. Modeled after the study planning system evaluated in [221], it is shown on the top-left of Figure 6.4 and prompts learners to write about their motivation and what brought them to the course in the first place. The key question asked to learners is *What drives you?* followed by other prompting questions to help learners express themselves: *What brought you* *here?* and *What do you hope to gain from this course?* Once learners have submitted their motivation it is persisted to our back-end. Learners can view and change their response any time.

Motivation expression feedback In order to provide feedback and encourage a habit of self-reflection, we regularly make learners aware of their latest motivation response by displaying it back to them (top-right of Figure 6.4) throughout each course week/unit. The response is shown as a quotation by the learner underneath the *What drives you:* text together with the learner's edX username (to emphasize once more the source of the quotation).

Plan formulation This interface (Figure 6.4 bottom-left) promotes forethought in prompting learners to formulate and state their plan for the coming course week in terms of engagement with course resources. Specifically, learners are prompted to enter the number of videos they intend to watch, quiz questions they intend to answer, and hours they intend to devote to the course this week. To aid learners in their planning, we provide the total number of videos and quizzes of the week (automatically extracted from the edX course pages) as well as the recommended time to spend in the course that week (as estimated by the course instructors).

Plan feedback To promote awareness learners' performance and encourage self-reflection, the planning feedback interface (Figure 6.4 bottom-right) consists of three gauges showing learners how well they have progressed towards the goals *they set for themselves*, removing all instructor influence. We designed the plan feedback as a data visualization dashboard that allows learners to easily draw their own insights about their progress. Previous research in data visualization for MOOC learners found that more abstract feedback (such as the "timeliness" of the quiz submissions) only benefited learners with a higher education background [66]. Since highly educated learners already have SRL abilities, we aimed to engage those learners that lack self-regulation skills and designed the interface to be clear and straightforward to interpret.

6.5 Study Setup

Participants

We deployed `SRLx` in an edX MOOC on renewable energies offered by the Delft University of Technology. The course consists of 75 individual lecture videos and 295 graded quiz questions. A total of 8,057 learners enrolled in the course. The course started on August 29, 2017 and concluded on November 8, 2017. We made `SRLx` available to all learners but did not provide any additional incentive for using it.

Before the course, the learners were asked to self-report their basic demographic information. 5,349 learners at least partially complied. Of these learners, 25.3% are female; the learners' median age is 26. We also collected information about their prior education level, as this has shown to have a significant impact on learning outcomes and engagement with MOOCs [66]. As is common in MOOCs, we observe a great variety in this respect with learners running the gamut from high school to PhD levels of prior education: 1% had no prior formal education, 20% held at least a high school diploma, 5% an Associate's degree, 45% a Bachelor's degree, 26% a Master's degree, and 3% a PhD. We consider learners' prior education level to be *high* when they have earned at least a Bachelor's degree, and *low* when they have not.

Given that many learners who enroll in a MOOC never enter the platform and log a session (a common occurrence in MOOCs), we narrow down the sample for analysis accordingly. Among all learners enrolled, 2,961 entered the course at least once and are therefore considered as *active learners* in our analyses.

Measures

To evaluate the role that `SRLx` plays in learners' achievement and course engagement, we measure a number of in-course learning behaviors that are commonly used in MOOC studies as well as a number of novel measures enabled by `SRLx`:

- Average quiz score $\in [0, 1]$ (proportion of attempted quiz questions answered correctly);
- Course activities:
    - Number of video interactions (play, pause, fast-forward, rewind, scrub);

  - Number of quiz submissions (submissions, correctness);
  - Number of discussion forum posts;
  - Time spent in the course;
- SRLx interactions:
  - Plan formulation (number of videos & quizzes and hours planned to spend in the course that week);
  - Motivation expression (submission text);
  - Editing (changing an established motivation or plan).

6.6 Results

In this section we analyze the deployment of SRLx along four lines: (i) course-level learning behaviors, (ii) study plan formulation tendencies, (iii) plan achievement rates, and (iv) motivations expressed over time.

6.6.1 Course-level Learning Behaviours

In Table 6.4 we present summary statistics for overall course behavior among all active learners, characterized by having logged at least one session in the course. Table 6.5 shows the number of submissions made via SRLx.

Table 6.4: Overview of the average behavior of active learners. In rows 2 & 3 we partition the set of active learners into *Comply* (learners who formulated at least one plan *and* submitted at least one motivation expression) and *Non-Comply* (the remainder) learners.

Subset	N	Quiz Score	Session Count	SRLx Interact.	Feedback Checks	Quiz Submits	Videos Watched
Active	2,961	0.41	32.57	152.72	3.63	43.11	8.33
Comply	303	0.72	66.48	348.93	7.31	91.56	16.31
Non-Comply	2,658	0.37	28.71	130.35	3.21	37.58	7.42

Of the 2,961 active learners in the course, 872 (32%) engaged with SRLx at least one time (answering **RQ6.1**)—here characterized by having formulated at least one plan *or* submitting at least one motivation expression. While this rate of minimal engagement is substantially higher than past studies, e.g. [65], the true rate of compliance (submitting *both* a plan and a motivation) is still very low, at 10% (303 out of 2,961 active learners).

While the top row in Table 6.4 represents all active learners in the course, the bottom two rows show the impact of self-selection in highlighting the difference in behavior between learners who did and did not engage with `SRLx`: on average, learners using `SRLx` (i.e. our Comply group) log more than twice as many sessions, answer nearly three times as many quizzes, answer more questions correctly and watch more than twice as many videos compared to learners in the Non-Comply group. We cannot claim that this difference is caused by the use of `SRLx`; rather it is at least partially a result of the self-selection of learners who would have been highly engaged and more successful in the course regardless.

However, this trend could also be partially explained by prior research on the *doer effect*, or the "...association between the number of online interactive practice activities students do and their learning outcomes" [143]. This theory states that engagement with interactive course components (such as `SRLx`, discussion fora, or quiz questions) has a stronger learning effect than passive activities such as reading or watching lecture videos. So while `SRLx` is unlikely to be the sole cause of the increase in activity between compliers and non-compliers, theory states that it likely contributed, at least in part, to the more positive learning outcomes of those who engaged with it.

When we split the engagement between the different types of interfaces (Table 6.5), we find that the plan formulation interface was considerably more engaging, with more than twice as many learners formulating plans (on average two plan formulations per learner) than writing up their motivation.

Table 6.5: Number of submissions of motivation expressions, plan formulations, and plan/expression edits. The bottom row shows the number of unique learners to have completed each action type.

	Motivation Expression	**Plan Formulation**	**Edited**
#Submissions	679	1,997	748
#Learners	396	971	338

6.6.2 Study Plan Formulation

In this analysis we focus on the plans the learners made using `SRLx` and thus address **RQ6.2**. We explore the following questions: are the learners overly ambitious with their plan formulation? Are learners able to consistently stick to their plans? Do their planning tendencies/strategies change over time? Figure 6.5 shows an aggregate view of all 1,997 plans submitted in the course.

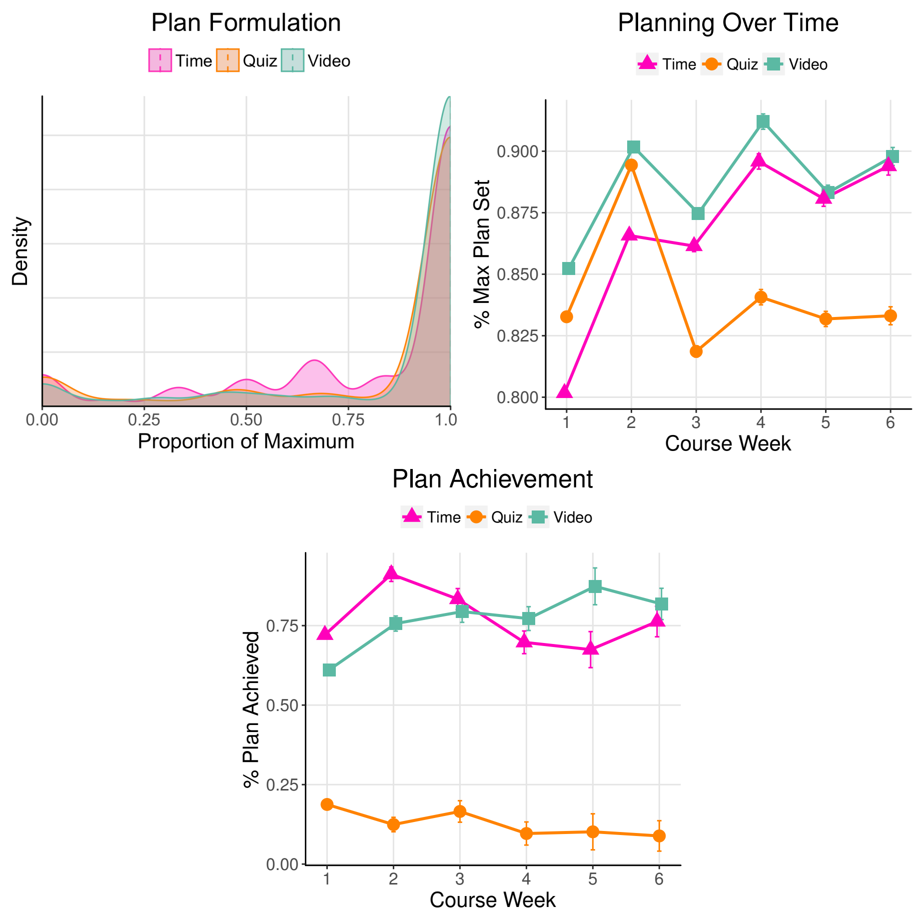


Figure 6.5: In clockwise order: (i) the proportion of learners' formulated plans set for the maximum possible value in the respective course week; (ii) the proportion of the maximum plan set by learners of each activity type over the span of all course weeks; (iii) plan achievement rates for each activity type by course week. Error bars show the standard error.

Figure 6.5 (top left) shows the study planning behavior (in terms of time commitment, quiz submissions, and videos watched) of all learners who formulated and submitted at least one plan in `SRLx`. We find that the majority of plans set were for the maximum given the week's content, i.e. most learners who submitted plans aimed at completing all quizzes, watching all videos and spending the instructor-suggested time on the course platform.

At the same time in Figure 6.5 (top left) we observe that the goals set pertaining to the proportion of time (from the recommended six hours per week) learners plan to commit to the course is lower than that of quiz submissions and videos. A Wilcoxon rank sum test with continuity correction ($W = 2,210,200$, $p < 0.0001$) indicates a significant difference between time plans ($\bar{x} = 0.838$, $\sigma = 0.34$) and video plans ($\bar{x} = 0.88$, $\sigma = 0.29$). From this analysis we conclude that learners are more conservative with the way they plan their time commitment to the course than the way they plan to engage with course materials.

To examine planning behavior at a more detailed level, in Figure 6.5 (top right) we segment planning behavior by course week and illustrate the change over time. Compared to the rather steady rate of ambition (proportion of maximum plan set) with quiz plans (overall mean of 84.7% of the maximum), learners exhibited an overall trend of increasing their ambition each week for time- and video-related plans—a 9 percentage point increase from Week 1 to Week 6 for time plans (mean of 80% to 89%) and a 5 percentage point increase for video plans (mean of 85% to 90%). While these two increases can be attributed to less-ambitious learners dropping out of the course, the lower rate for quiz-related plans still holds throughout the entire course.

6.6.3 Plan Achievement

Figure 6.5 (bottom) shows the rate at which learners achieve each aspect of their plans each course week (**RQ6.2**). Whereas in the previous section we discussed how learners are conservative with their plan formulations as it pertains to time, we see in Figure 6.5 (bottom) that learners are strong at achieving their plans for time commitment and video lecture viewing with high consistency across course weeks—an important insight given that poor time management has been identified by prior research [134, 274, 189, 129] as one of the primary causes of attrition in MOOCs.

It is also worth noting that the consistency and success of learners' time planning achievement is not a product of less ambitious goals being set. Refer back to Figure 6.5 (top right) to see that the opposite is actually true; learners become *more* ambitious with their time plans as the course progresses, and learners are still able to achieve their plans with high consistency.

For the learners' video watching plan achievement, we observe a slight increase across the weeks with an overall mean of 63% completion. For learners' achievement of their quiz question-related plans, we observe substantially lower completion rates than those regarding time—falling from 19% in Week 1 to a mere 9% in Week 6.

We hypothesize that these results on plan achievement are a product of the difficulty of each activity type. Though not trivial, spending time in the platform requires little more than a learner's presence. Slightly more demanding is the activity type of watching lecture videos; and most challenging of all three is answering quiz questions, which is not only dependent on the previous two activities but also requires the application of newly-acquired knowledge. In other words, the rate by which learners complete their plans is commensurate with the exigency of the respective activity type.

As previous research on MOOC learners has identified achievement gaps among learners [133], we next conducted an exploratory analysis on plan completion per activity type as a function of a learner's prior education level (with *high* education learners having earned at least a Bachelors degree, accounting for 75% of learners in the course). We observe no significant difference in plan completion rates in any of the three activity types according to a Wilcoxon rank sum test with continuity correction, thus indicating that learners are able to effectively use `SRLx` across a wide range of ability levels. This suggests that `SRLx` is equally usable and effective for learners of all prior education levels.

6.6.4 Motivation Expression

Finally, we also conducted a preliminary analysis of the motivation texts our learners submitted. Among the 2,961 learners exposed to the `SRLx` interface, 396 submitted at least one motivation expression. These motivations range from learners working towards having better career opportunities to changing the world—the latter theme became markedly more prominent as the course progressed. The average word count is 23.9 (median 15, minimum 1, maximum 329). In Table 6.6 we randomly picked examples of *short* (at most ten words), *medium* length (up to 25 words) and *long* (26 words or more) submissions.

Replicating the methods in [272] applied to MOOC learner texts on course intentions, we evaluated the predictive value of the length of a learner's text submission on their (i) current grade, (ii) average quiz question score, and (iii) total time spent in the course platform and were not able to find a significant effect in any of the metrics.

Table 6.6: Random sample of short, medium, and long submissions through the Motivation Expression interface.

S1	*Build up on sustainable energy knowledge*
S2	*I expect to get to know the future of energy*
M1	*I hope to learn more about sustainable ways of using and obtaining energy.*
M2	*I want a clean planet I want to be responsible for that*
L1	*As a junior architect I am interested in learning more about the relationship between energy use and building design and how intelligent design can have positive impacts on building energy use as well as occupant health and happiness.*

The ten most frequent terms occurring among all motivations are (in descending order): *energy*, *renewable*, *sustainable*, *knowledge*, *learn*, *future*, *course*, *hope*, *better* and *sources*. These terms speak to the motivation of many learners to use the knowledge to improve the world; interestingly, no job related term appears in this list (the term *career* occurs at rank 20), indicating that many of our learners have an intrinsic, rather than an extrinsic motivation. They are brought to the course and engage with the materials not out of need for career change or certification (as was commonly observed among MOOC learners in previous work [130]), but rather out of a desire to be able to spark positive change in the world. Given the topic of the course and its relevance to the issues facing society today, this certainly affects learner motivation in some sense, but this also demonstrates that MOOCs can be instrumental to shaping the next generation of emerging technologies in making the subject matter accessible to the masses.

6.7 Discussion

Based on the existing literature and theory on self-regulated learning, we designed `SRLx` to encourage and support learners in adopting effective self-regulated learning habits in MOOCs. `SRLx` enables learners express their (changing) motivation and to set their own goals and track their progress towards them in real-time instead of following instructor-prescribed goals.

To evaluate the efficacy of `SRLx` we deployed it in a MOOC with more than 2,900 active learners to observe to what extent and how learners engage with it.

Despite the inconsistencies we observed based on previous related work, learner interactions with `SRLx` offer novel insights about the role of motivation expression and plan formulation for MOOC learners. We find (i) that as the course progresses, learners are able to plan their time commitment more effectively, (ii) a strong trend of intrinsic motivation shared by learners with the motivation expression interface, and (iii) learners are most conservative with the way they plan to commit time to the course compared to video and quiz activity planning.

Given our findings on the progression of learner's planning strategies over time with `SRLx`, we are able to offer an explanation of the findings by Yeomans and Reich [272] who found that plans that were formulated about time were less likely to succeed: that intervention took place at the beginning of a course, where learners formulated time plans over the long-term—requiring the foresight of many weeks in the future; `SRLx`, on the other hand, allows learners to set a new plan at the beginning of each course week (short- to medium-term). Combined with our evidence that learners become more effective at plan formulation over the span of the course, we conclude that time-specific plans are likely only to be ineffective when on a long-term scale; and when used on a short- to medium-term scale, they can be effective and attainable.

Future research should implement `SRLx` as a randomized controlled trial, or A/B test, in MOOCs to explore questions of causality—does `SRLx` directly cause learners to learn and engage more?

Finally, `SRLx`, as presented here, is completely individualistic—learners only receive feedback on their own plan formulations and motivation expressions. By making `SRLx` social, or showing learners the planning behavior and performance of their peers as well as their own, this could present a promising way to leverage the scale of MOOCs and improve learner performance through increased social presence.

Part IV

Retrieval Practice

This part serves **RQ2** (*How can MOOC environments be improved to advance the possibilities of experimentation?*) by focusing on the learning strategy known as retrieval practice (also referred to as the testing effect). Retrieval practice is the learning science principle which states that the active recall of information is more effective at encoding it into long term memory than passive revisiting. For example, taking a practice test is a much more effective learning strategy than re-watching a video lecture, because in the practice test one is actively challenged/prompted to produce the information, whereas in the video lecture, the information is delivered to the watcher. Retrieval practice is touted as one of the most effective learning strategies, and scores of past research studies have supported its effectiveness in the traditional learning setting. For that reason, we identify this as a key learning strategy to evaluate in terms of its transferability to MOOCs.

The first study we ran to evaluate this question is presented in a pilot study. Here we inserted retrieval cues after each video lecture in a MOOC, where we A/B tested an intervention which asked the learners to write out and explain the key concepts (in 3–5 sentences) of the previous video—a common retrieval practice tactic. We found that this intervention yielded high levels of non-compliance—again due to the fact that we ask the learners to produce and type out their response in a text box. Due to the high levels of non-compliance, we did not observe any significant effects of the retrieval practice intervention on eventual learner outcomes.

There are a number of ways to engage with retrieval practice, and the text box approach above is just one. In the main study, we evaluated a new approach to promoting retrieval practice which did not require learners to produce their own text/writing, but instead prompted them to reactivate prior knowledge in the form of a formative assessment question from an earlier week in the course. To accomplish this we developed `ARPS`, an adaptive retrieval practice system, which tracked which content a learner had previously visited and intermittently asked the learner pop-quiz questions from previous weeks to keep their memory of that knowledge active. Even though this A/B test did not indicate significant effects of the intervention on learning outcomes, the data generated from `ARPS` (due to its frequent assessment of learners' knowledge levels of various topics), enabled a novel analysis of how MOOC learners' knowledge evolves (or decays) over time.

In summary, this part contributes a deeper understanding of the extent to which the learning strategy of retrieval practice can be effectively facilitated in a MOOC setting. We find inconclusive evidence of this due to the systematically high levels of non-compliance with interventions designed to facilitate this practice and highlight the need for future research to explore techniques for better engaging MOOC learners in promoting effective study habits.

Chapter 7

Knowledge Retention and Retrieval Practice in MOOCs

Massive Open Online Courses are successful in *delivering* educational resources to the masses, however, the current retention rates — well below 10% — indicate that they fall short in helping their audience become effective MOOC learners. In this chapter, we first share the results of a pilot study conducted in order to test the effectiveness of *retrieval practice* (i.e. strengthening course knowledge through actively recalling information), which has been found to be highly effective in traditional learning environments. In contrast to the classroom-based results, we do not confirm our hypothesis, that small changes to the standard MOOC design can teach MOOC learners valuable self-regulated learning strategies.

Retrieval practice has been established in the learning sciences as one of the most effective strategies to facilitate robust learning in traditional classroom contexts. The cognitive theory underpinning the "testing effect" states that actively recalling information is more effective than passively revisiting materials for storing information in long-term memory. In the main study of this chapter, we document the design, deployment, and evaluation of an `Adaptive Retrieval Practice System` (`ARPS`) in a MOOC. This push-based system leverages the testing effect to promote learner engagement and achievement by intelligently delivering quiz questions from prior course units to learners throughout the course. We conducted an experiment in which learners were randomized to receive `ARPS` in a MOOC to track their performance and behavior compared to a control group. In contrast to prior literature, we find no significant effect of retrieval practice in this MOOC environment. In the treatment condition, passing learners engaged more with ARPS but exhibited similar levels of knowledge retention as non-passing learners.

This chapter is based on published peer-reviewed work as “Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-Regulated Learning Strategies at Scale” [65], by D. Davis, G. Chen, van der Zee, Tim, C. Hauff, and G.J. Houben in Proceedings of the 11th European Conference on Technology-Enhanced Learning, 2016. and “The Half-Life of MOOC Knowledge: A Randomized Trial Evaluating Knowledge Retention and Retrieval Practice in MOOCs” [68], by D. Davis, R. Kizilcec, C. Hauff, and G.J. Houben in Proceedings of 8th International Conference on Learning Analytics and Knowledge, 2018.

7.1 Introduction

Retrieval practice is one of the most effective and well-established strategies to facilitate robust learning. Also known as the testing effect, retrieval practice is the process of reinforcing prior knowledge by actively and repeatedly recalling relevant information. This strategy is more effective in facilitating robust learning (the committing of information to long-term memory [142]) than passively revisiting the same information, for example by going over notes or book chapters [4, 49, 212, 97, 166, 118, 117].

Given the wealth of scientific evidence on the benefits of retrieval practice (cf. Section 7.2) and the adaptability of digital learning platforms, in this paper we explore to what extent the testing effect holds in one of today's most popular digital learning settings: Massive Open Online Courses. Research into both MOOC platforms and MOOC learners' behavior has found learners to take a distinctly linear trajectory [63, 260, 89] through course content. Many learners take the path of least resistance towards earning a passing grade [273] which does not involve any back-tracking or revisiting of previous course units—counter to a regularly-spaced retrieval practice routine.

Although MOOC platforms are not designed to encourage retrieval practice, prior work suggests that MOOC learners with high Self-Regulated Learning (SRL) skills tend to engage in retrieval practice of their own volition [134]. These learners strategically seek out previous course materials to hone and maintain their new skills and knowledge. However, these learners are the exception, not the norm. The vast majority of MOOC learners are not disciplined, self-directed autodidacts who engage in such effective learning behavior without additional support. This motivated us to create the `Adaptive Retrieval Practice System` (`ARPS`), a tool that encourages retrieval practice by *automatically* and *intelligently* delivering quiz questions from previously studied course units to learners. The system is automatic in that the questions appear without any required action from the learner and intelligent in that questions are adaptively selected based on a learner's current progress in the course. We deployed `ARPS` in an edX MOOC (`GeoscienceX`) in a randomized controlled trial with more than 500 learners assigned to either a treatment (`ARPS`) or a control group (no `ARPS`).

Based on the data we collect in this randomized trial, we investigate the benefits of retrieval practice in MOOCs guided by the following research questions:

RQ7.1 How does an adaptive retrieval practice intervention affect learners' academic achievement, course engagement, and self-regulation compared to generic recommendations of effective study strategies?

RQ7.2 How does a push-based retrieval practice intervention (requiring learners to act) change learners' retrieval practice behavior?

In addition to collecting behavioral and performance data inside of the course, we invited learners to complete a survey two weeks after the course had ended. This self-report data enabled us to address the following research question:

RQ7.3 To what extent is robust learning facilitated in a MOOC?

The primary contributions of our study show that (i) retrieval practice, in contrast to substantial prior work, may not benefit learners in a MOOC (**RQ7.1**); (ii) passing and non-passing learners who receive `ARPS` do not differ in their knowledge levels (as measured by `ARPS`) but rather in their course engagement levels (**RQ7.2**); and (iii) passing and non-passing learners do not differ in long-term knowledge retention (**RQ7.3**).

7.2 Related Work

We now review prior research in the areas of retrieval practice, spaced vs. massed practice, and long-term knowledge retention to inform the study design.

7.2.1 Retrieval Practice

Adesope et al. [4] conducted the most recent meta-analysis of retrieval practice. They evaluated the efficacy of retrieval practice compared to other learning strategies such as re-reading or re-watching, the impact of different problem types in retrieval practice, the role of feedback, context, and students' education level.

The effect of retrieval practice is strong enough overall for the authors to recommend that frequent, low-stakes quizzes be integrated into learning environments so that learners can assess knowledge gaps and seek improvement [4]. They also found that multiple choice problems not only require low levels of cognitive effort, they were the most effective type of retrieval practice problem in terms of learning outcomes compared to short answer questions. And while certainly a boon to learners (the majority of studies in the review endorse its effectiveness), feedback is actually not required or integral to effective retrieval practice. From studies that did incorporate feedback, the authors found that delayed feedback is more effective in lab studies, whereas immediate feedback is best in classroom settings. Of the 217 experiments (from the 118 articles included in the meta-analysis), 11% took place in traditional classroom settings as part of the curriculum, with the vast majority taking place in laboratory settings.

Roediger and Butler [212] also offer a synthesis of published findings on retrieval practice. From the studies reviewed, the authors offer five key points on retrieval practice for promoting long-term knowledge: (i) retrieval practice is superior to reading for long-term retention, (ii) repeated testing is more effective than a single test, (iii) providing feedback is ideal but not required, (iv) benefits are greatest when there is lag time between learning and practicing/retrieving, and (v) retrieval practice increases the likelihood of learning transfer—the application of learned knowledge in a new context [212].

Consistent with the findings from [97, 212, 4], Johnson and Mayer [113] evaluated the effectiveness of retrieval practice in a digital learning environment focused on lecture videos. In the study, learners who answered test questions after lecture videos—pertaining to topics covered in the videos—outperformed learners who merely re-watched the video lectures in terms of both long-term knowledge retention and learning transfer [113].

7.2.2 Spaced vs. Massed Practice

The literature on spaced versus massed practice has shown that a higher quantity of short, regularly-spaced study sessions is more effective than a few long, massed sessions [49]. There is considerable overlap in the research on retrieval practice and that on spaced versus massed practice. As outlined in the studies above, an optimal study strategy is one of a regularly spaced retrieval practice routine [182, 39, 49].

Spaced versus massed practice has been evaluated in the MOOC setting by Miyamoto et al. [182], who analyzed learners' log data and found that learners who tend to practice effective spacing without guidance or intervention are more likely to pass the course relative to those learners who do not engage in spacing. We leveraged these insights from the learning sciences in the design of `ARPS`.

7.2.3 Expected Knowledge Retention

Scientific evaluation of the human long-term memory began at the end of the 19$^{\text{th}}$ century, leading to the earliest model of human memory loss/maintenance: the Ebbinghaus curve of forgetting [79]. The curve begins at time 0 with 100% knowledge uptake with a steep drop-off in the first 60 minutes to nine hours, followed by a small drop from nine hours to 31 days.

Custers [61] conducted a review of long-term retention research and found considerable evidence in support of the Ebbinghaus curve in terms of shape—large losses in short-term retention (from days to weeks) which level off for longer intervals (months to years)—but not always in terms of scale. The result of their meta-analysis shows that university students typically lose one third of their knowledge after one year, even among the highest-achieving students.

Considering the effect on retrieval practice on long-term retention, Lindsey et al. [166] conducted a similar study to the present research in a traditional classroom setting and found that their personalized, regularly spaced retrieval practice routine led to higher scores on a cumulative exam immediately after the course as well as a cumulative exam administered one month after the course. In their control condition (massed study practice), learners scored just over 50% on the exam, whereas those exposed to the retrieval practice system scored 60% on average. For the control group, this marked an 18.1% forgetting rate, compared to 15.7% for those with retrieval practice. They also found that the positive effect of retrieval practice was amplified with the passing of time.

Duolingo, a popular language learning platform with hundreds of thousands of daily users, has developed their own forgetting curve to model the "half-life" of knowledge—theirs operates on a much smaller time scale, with a 0% probability of remembering after seven days. Based on the retrieval practice and spacing effect literature, they also developed a support system to improve learners' memory. Findings show that their support system, tuned to the "half-life regression model" of a learner's knowledge, significantly improves learners' memory [242].

It is worth noting, however, that forgetting is viewed as an adaptive behavior: forgetting liberates the memory of outdated, unused information to free up space for new, immediately relevant memories and knowledge [211]. Retrieval works adjacently to this in that by regularly reactivating and revisiting knowledge, the brain does not tag it as unused and forgettable, but rather recognizes its relevance and, accordingly, stores it in long-term memory.

7.3 Retrieval Practice Pilot Study

To evaluate the efficacy of interactivity in retrieval practice interfaces, we first conducted a randomized controlled trial pilot study in one MOOC where we prompted learners to write 3–5 sentence summaries of each lecture video after watching it in order to stimulate the active retrieval of their recently learned knowledge.

We deployed the retrieval practice intervention in the `Functional Programming` course, a 13-week MOOC which introduces basic functional programming language constructs. Nearly 28,000 learners enrolled, and 5% eventually passed the course. $9,836$ learners were active in the course and were randomly assigned to one of three experimental conditions.

In the original course design (i.e. no intervention) of `Functional Programming`, each week's video lecture is broken up into two or three segments. And although the students must navigate themselves from one segment to the next, there are no other learning materials or activities between. In order to activate the learning process, we inserted retrieval practice cues designed to make learners stop and process the information presented in the video lecture.

In each course week, we inserted a retrieval cue directly after the final lecture video, thus prompting the learners to stop and think before moving on to the weekly quiz. The only exception to this design was one particular course week[2] where we inserted retrieval practice cues after each of the three segments of the weekly lecture, as in the previous edition of the course learners had perceived that week's material as the most challenging.

This experiment had three groups (or conditions): (1) the control group without an intervention, (2) the "cued" group, and (3) the "given" group which was provided a 3–5 sentence summary after each lecture. The "cued" group was shown the following prompt along with a blank text input box:

> *Please respond in 3-5 sentences to the following question: "In your opinion, what are the most important points from the previous video?"*

[2]"Week 7: Functional Parsers and Monads"

Note that these responses were not seen, graded, or given any feedback from the instructor — serving strictly as an activity for learners to exercise and improve memory recall. The “given" group, instead of being asked to create a summary themselves, was provided with a 3–5 sentence summary of the video as generated by one of the authors highly familiar with the functional programming paradigm. We included the “given" group in our study to determine the effect of actively retrieving information from memory versus being provided a summarizing text for passive reading.

7.3.1 Findings

Learners engage less with interventions than course content items. Of the 3,262 learners in the “cued" condition in `Functional Programming`, 2,166 (66.4%) logged at least one video-watching event in the course. Among these same learners only 719 (22%) clicked on any of the retrieval practice interventions.

We first tested whether the learners of the cued, given, and control groups score differently in the weekly quizzes. To this end we performed a MANOVA test with the highly engaged learners (characterized by having spent more than the group's mean time watching videos in Week 1 which is ≈ 22 minutes) in each of the three conditions as a fixed factor and the grades on the weekly quizzes as a dependent variable. The MANOVA test followed by the post hoc Games-Howell (equal variances not assumed) test yielded no significant differences between each group's weekly quiz grade.

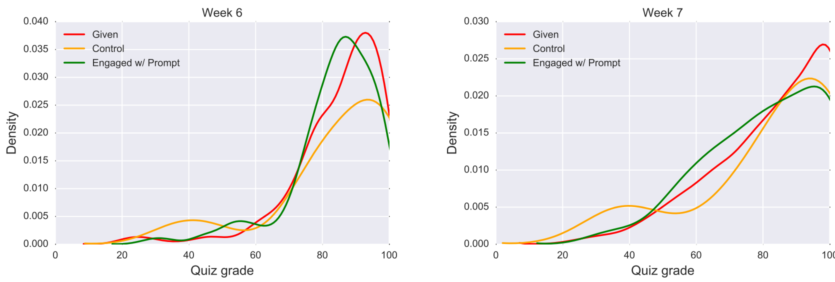


Figure 7.1: KDE plot showing the distribution of weekly quiz grades across the groups of highly engaged learners. All lines were fit using a Gaussian kernel function. None of the differences between groups are statistically significant at the $\alpha = 0.01$ level.

In the previous analysis all highly engaged students from each condition were included. However, as many students did not engage with the intervention, this can give a distorted view of its effects. Therefore, we next isolated those learners who actively engaged (characterized by viewing the intervention for at least 10 seconds) with an intervention prompt at least once.

Using these new group definitions, we still observe no statistically significant differences between the groups as a result of a MANOVA (to test the difference between weekly quiz scores), and a one-way ANOVA (to test the difference between course final scores). Figures 7.1 and 7.2 illustrate these null findings via Kernel Density Estimation (KDE) plots.

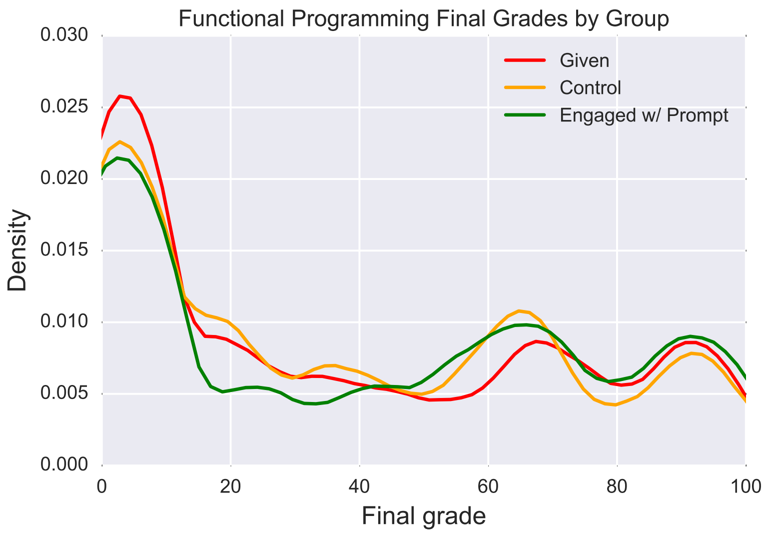


Figure 7.2: KDE plot showing the distribution of final course grades across the groups of highly engaged learners. All lines were fit using a cosine kernel function. None of the the differences between groups are statistically significant at the $\alpha = 0.01$ level.

Overall, we find that exposure to static text retrieval practice prompts is not sufficient to significantly increase learner engagement or final grades. When addressing the issue of noncompliance, we observe that even learners who engaged with the retrieval cues show no significant difference in any measure of performance. To this end, we took the insights gained from this study and applied them to the design of an automated & adaptive retrieval practice system, `ARPS`.

7.4 Adaptive Retrieval Practice System Overview

The `Adaptive Retrieval Practice System` (`ARPS`) is a client-server application (written in `JavaScript/node.js`)[3] that provides *automated*, *scalable* and *personalized* retrieval practice questions to MOOC learners on a continuous basis. We developed `ARPS` specifically for use within the edX platform in taking advantage of the RAW HTML input affordance. This allows course teams/instructors to build custom interfaces within the platform that render along with the standard edX content (such as videos, quizzes, etc.).

The `ARPS` back-end keeps track of the content a MOOC learner has already been exposed to through client-side sensor code that logs a learner's progress through the course and transmits it to the back-end. Once the back-end receives a request from the `ARPS` front-end (a piece of JavaScript running in a learner's edX environment on pages designated to show retrieval practice questions), it determines which question to deliver to a learner at a given time based on that learner's previous behavior in the course by randomly selecting from a personalized pool of questions only pertaining to content the learner has already been exposed to. Each question is pushed to the learner in the form of a `qCard`, an example of which is shown in Figure 7.5. These `qCard`s appear to the learner as a pop-up within the browser window. We log all `qCard` interactions—whether it was ignored or attempted, the correctness of the attempt, and the duration of the interaction.

In contrast to previous interventions in MOOCs [65, 66, 128, 135, 272], we push questions to learners instead of requiring the learner to seek the questions out. We adopted this push-based design in order to allow learners to readily engage with the intervention with minimal interruption to the course experience. This design also addresses the issue of treatment noncompliance that has arisen in past research [65, 66]. `ARPS` is seamlessly integrated in the course, requiring as few additional interactions as possible. In the case of Multiple Choice (MC) questions (example problem text in Figure 7.3), the entire interaction requires just a single click: the learner selects their chosen response and if correct, receives positive feedback (a ✓ mark accompanied by encouraging text), and the `qCard` disappears. Incorrect responses invoke negative feedback (a **x** symbol alongside text encouraging the learner to make another attempt) which disappears after 4 seconds and returns the learner to the original question so they can try the problem again.

[3]The code is available at https://github.com/dan7davis/Lambda.

We also enabled one other question type[4] to appear in `qCards`: Numeric Input (NI) problems (an example is shown in Figure 7.4). These problems require the learner to calculate a solution and enter the answer in a text box. While requiring more effort than a single click response, we included this problem type to allow for a comparison between the two.

> A body with a low density, surrounded by material with a higher density, will move upwards due to buoyancy (negative density difference). We analyze the situation of a basaltic magma generated at a depth of 10 km and surrounded by gabbroic rocks. Will the magma move downward, remain where it is or move upward?

Figure 7.3: Example of an easy (less than 5% of incorrect responses) Multiple Choice question in `GeoscienceX`.

> Suppose an earthquake occurred at a depth of 10 kilometers from the surface that released enough energy for a P-wave to travel through the center of the Earth to the other side. This is for the sake of the exercise, because in reality sound waves tend to travel along the boundaries and not directly through the Earth as depicted. Assume the indicated pathway and the given thicknesses and velocities. How many seconds does it take for the seismic P-wave to reach the observatory on the other side of the Earth?

Figure 7.4: Example of a difficult (5% correct response rate) Numerical Input question in `GeoscienceX`.

7.5 Study Design

We now describe the MOOC we deployed `ARPS` in as well as the design of our empirical study.

7.5.1 Participants

A total of 2,324 learners enrolled in the course titled *Geoscience: the Earth and its Resources* (or `GeoscienceX`), which was offered on the edX.org platform between May 23, 2017 and July 26, 2017. The course consists of 56 lecture videos and 217 graded quiz questions. Of the 132 total problems from the 217 in the course question bank deemed suitable for use with `qCards`

[4]Additional question types that are supported by the edX platform can easily be added to `ARPS`; in this paper we focus exclusively on MC and NI questions as those are the most common question types in the MOOC we deployed `ARPS` in.

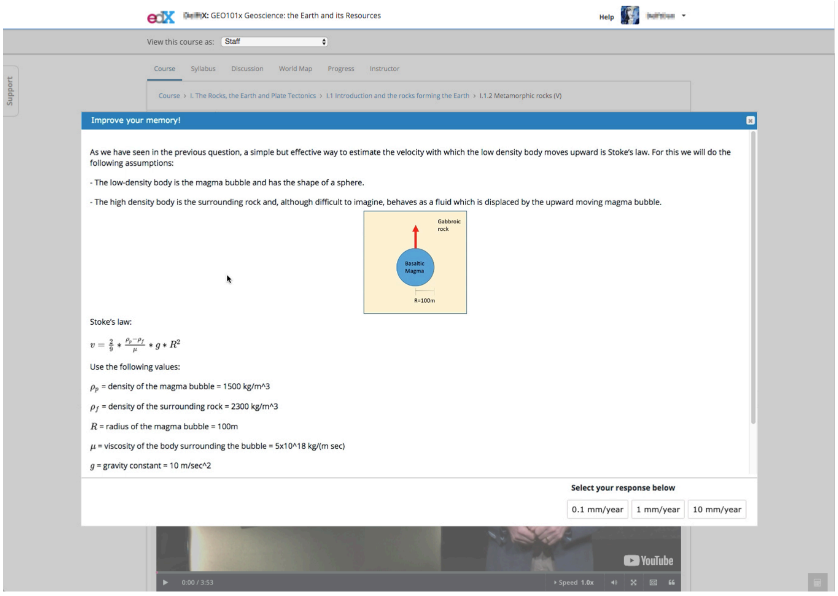


Figure 7.5: Example `qCard` in the `GeoscienceX` course. The main body of the `qCard` contains the question text, and the bar at the bottom contains the MC answer buttons. The grey "x" at the top right corner closes the window and dismisses the problem.

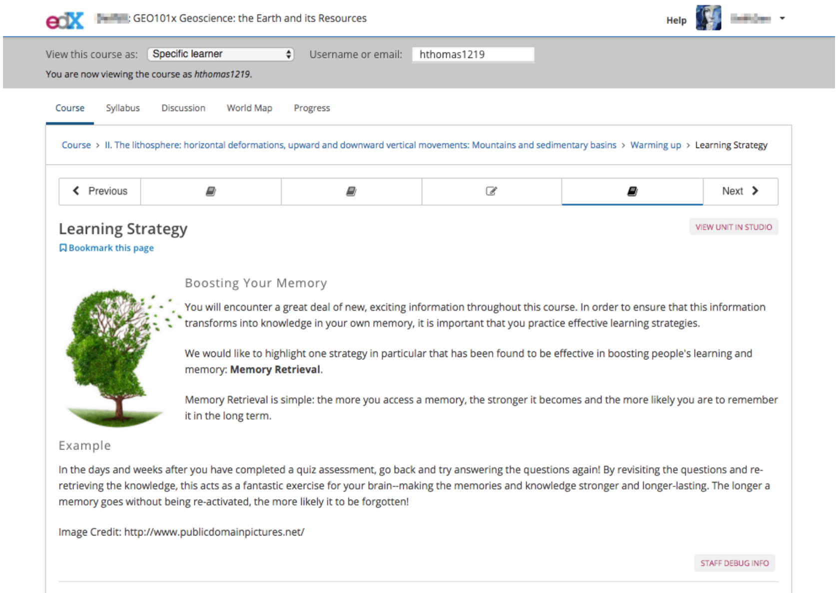


Figure 7.6: Page shown to learners in the control condition at the beginning of each course week describing how to practice an effective memory retrieval routine.

(multi-step problems were excluded so that each `qCard` could be answered independently), 112 were Multiple Choice and 20 were Numerical Input problems.

Based on self-reported demographic information (available for 1,962 learners), 35% of participants were women and the median age was 27. This course drew learners from a wide range of educational backgrounds: 24% held at least a high school diploma, 7% an Associate's degree, 42% a Bachelor's degree, 24% a Master's degree, and 3% a PhD. Learners were not provided any incentive beyond earning a course certificate for participating in the study.

We define the study sample as the 1,047 learners who entered the course at least once (out of the 2,324 who initially enrolled): 524 assigned to the control condition and 523 to the treatment condition.

A post-course survey & quiz (cf. Section 7.5.2) was sent to all 102 learners who engaged with the `ARPS` system (9 complete survey responses—8.8% response rate) and the 150 highest performing learners in the control condition in terms of final grade (11 complete responses—7.3%).

7.5.2 Procedure

This study was designed as a randomized controlled trial in the `GeoscienceX` course. Upon enrolling in the course, learners were randomly assigned to one of two conditions for the duration of the course:

- **Control condition**: A lesson on effective study habits was added to the weekly introduction section. The lesson explained the benefits of retrieval practice and offered an example of how to apply it (Figure 7.6).
- **Treatment condition**: `ARPS` was added to the course to deliver quiz questions (via a `qCard`) from past weeks. The same weekly lesson on study habits as in the control condition was provided to help learners understand the value of the tool. In addition, information on how the adaptive retrieval system works and that responses to the `qCard` do not count towards learners' final grade was provided. The `qCard`s were delivered before each of the 49 course lecture videos (from Weeks 2–6) across the six course weeks. A button at the bottom of each lecture video page enabled learners to receive a new `qCard` on demand after the initial one to keep practicing.

To assess how well learners retained their knowledge from the course, we sent a post-course survey to the most active learners in the course (in terms of time spent in the platform) two weeks after the course had ended. The survey contained a random selection of ten assessment questions from the GeoscienceX course. Learners in the treatment condition additionally received eight questions about their experience with ARPS. We evaluated the results of this post-course assessment with respect to differences between the two cohorts in long-term knowledge retention.

7.5.3 Measures

In order to measure and compare the behavior of learners in both the control and treatment conditions, we consider the following measures of in-course events (tracked and logged on the edX platform):

- **Final grade** (a score between 0 and 100);
- **Course completion** (binary indicator: pass, no-pass);
- Course activities:
  - Video interactions (play, pause, fast-forward, rewind, scrub);
  - Quiz submissions (number of submissions, correctness);
  - Discussion forum posts;
  - Duration of time in course;
- ARPS interactions:
  - Duration of total qCard appearance;
  - Response submissions (with correctness);
  - qCard interactions (respond, close window).

The following data were collected in the post-course survey:

- Course survey data
  - **Post-Exam Quiz Score** (between 0-10);
  - Learner intentions (e.g., to complete or just audit);
  - Prior education level (highest degree achieved).

We have selected the three bolded variables as our *primary outcome variables* for this study for the following reasons: (i) a learner's final grade is the best available indicator of their performance in the course in terms of their short-term mastery of the materials and (ii) the Post-Exam Quiz score measures how well learners retained the knowledge weeks after finishing the course.

Table 7.1: Course outcomes in the control and treatment group.

Condition	Subset	N	Non-Zero Grade	Passing Rate	Grade Quantiles
Control	All	524	31%	8%	[0, 0, 2]
Treatment	All	523	34%	7%	[0, 0, 2]
Treatment	Complier	102	76%	34%	[2, 19, 74]
Treatment	Noncomplier	421	23%	0.2%	[0, 0, 0]

7.6 Results

This section presents the findings from each of the five analyses we conducted: (i) estimating the causal effect of the intervention based on the randomized controlled experiment (**RQ7.1**), (ii) examining how learners interacted with `ARPS` (**RQ7.2**), (iii) modeling how learners' knowledge changed over time (**RQ7.3**), (iv) estimating the rate of learners' long-term knowledge retention (**RQ7.3**), and (v) understanding learners' experience with `ARPS` from a qualitative angle using survey responses. Each subsection concludes with a statement synthesizing its key finding.

Table 7.2: Summary statistics for the mean value of the measures listed in Section 7.5.3 for analyses including all learners in both conditions who logged at least one session in the platform. The vertical line separates standard course behavior measures and those collected by `ARPS`.

Group	N=	Final Grade	Pass Rate	Vid Interact	Quiz Submit	Forum Posts	Time in Course	Time w/ qCards	qCards Seen
Ctrl	524	9	8%	6.52	34.77	0.26	4h47m	–	–
Treat	523	8	7%	5.83	30.88	0.29	3h40m	23m9s	7.71

7.6.1 Effect of Encouraging Retrieval Practice

The goal of the randomized experiment is to estimate the *causal effect* of retrieval practice (**RQ7.1**). By comparing learners in the control and treatment group, we can estimate the effect of the encouragement to engage in retrieval practice with `ARPS`. However, many learners who were encouraged did not engage in retrieval practice, which is a form of treatment noncompliance. Specifically, of the 523 learners assigned to the treatment, only 102 interacted at least once with a `qCard` (i.e. complied with the treatment). For this reason, in order to estimate the effect of retrieval practice itself, we also analyze the experiment as an encouragement design.[5]

The primary outcome measure is the final course grade, which determines certificate eligibility (the passing threshold is 60%). Table 7.1 contains summary statistics for grade and certification outcomes in the control group and the treatment group, overall and separately for treatment compliers and noncompliers. First, we estimate the Intent-to-treat Effect (ITT), which is the difference in average outcomes between the treatment and control groups. We find that the ITT is not significant for certification (log odds ratio$= -0.215, z = -0.920, p = 0.357$), getting a non-zero grade (logOR$= 0.143, z = 1.08, p = 0.280$), and the continuous grade itself (Kruskal-Wallis $\chi^2_{df=1} = 0.592, p = 0.442$).

Next, we use an instrumental variable approach (Two-stage Least Squares) to estimate the effect of retrieval practice for those who used it (i.e. a Local Average Treatment Effect, or LATE) [6]. For a binary instrument Z, outcome Y, and compliance indicator G, we can compute the Wald estimator:

$$\beta^{IV} = \frac{E\,(Y|Z=1) - E\,(Y|Z=0)}{E\,(G|Z=1) - E\,(G|Z=0)} \tag{7.1}$$

The LATE is not significant either for certification ($\beta^{IV} = -0.078, z = -0.893, p = 0.371$), getting a non-zero grade ($\beta^{IV} = 0.160, z = 1.11, p = 0.267$), and the continuous grade itself ($\beta^{IV} = -0.066, z = -0.889, p = 0.374$).

Finally, we estimate the per-protocol effect, which is the difference in average outcomes between treatment compliers and control compliers (i.e. the entire control group). We find large differences in terms of certification (logOR$= 1.74, z = 6.66, p < 0.001$), getting a non-zero grade (logOR$= 2.00, z = 7.94, p < 0.001$), and the continuous grade itself (Kruskal-Wallis $\chi^2_{df=1} = 99, p < 0.001$). However, the per-protocol estimates do not have a causal interpretation because different subpopulations are compared: all learners in the control group versus those highly motivated learners who comply in the treatment group. For instance, note that treatment compliance is strongly correlated with receiving a higher grade (Spearman's $r = 0.56, p < 0.001$).

[5]The study was pre-registered at www.osf.io/4py2h. Due to the small sample size and compliance rate, we adjusted our analytic approach. Specifically, we analyze the experiment as an encouragement design beyond estimating average treatment effects, and we did not apply the specified sample exclusion criteria because they could inadvertently bias the causal inference.

In addition to estimating effects based on the final course grade, the preregistration also specifies a number of process-level analyses (**RQ7.2**). In particular, we hypothesized that learners who receive the treatment would exhibit increased self-regulatory behavior in terms of (i) revisiting previous course content such as lecture videos, (ii) self-monitoring by checking their personal progress page, and (iii) generally persisting longer in the course. No evidence in support of the hypothesized behavior was found, neither in terms of the ITT (Kruskal-Wallis $\chi^2_{df=1}$s$< 0.68, p$s> 0.41) nor in terms of the LATE ($|z|$s$< 0.98, p$s> 0.32). Focusing on learners in the treatment group, we also hypothesized that learners who attempt `qCards` at a higher rate would learn more and score higher on regular course assessments, which is supported by the data (Spearman's $r = 0.42, p < 0.001$). In summary (and in contrast to previous studies on the topic [4, 49, 212, 97, 166, 118, 117]):

> The causal analysis yields no evidence that `ARPS` raised learning, performance, or self-regulatory outcomes in this course.

This may be due to the low sample size or rate of compliance in this study. We also observed a selection effect into using `ARPS` among highly motivated learners in the treatment group. Among those learners, increased engagement with `qCards` was associated with higher grades, though this pattern could be due to self-selection (e.g., more committed learners both attempt more `qCards` and put more effort into assessments). To better understand how different groups of learners used `ARPS` and performed on subsequent learning assessments, we conducted a series of exploratory analyses.

7.6.2 Engaging with Retrieval Cues

Question-by-Question Analysis

Figure 7.7 illustrates learners' responses for every question delivered by `ARPS`, which indicates which questions learners tended to struggle with (or ignore). The figure reveals that the choice to attempt or ignore a `qCard` is strongly associated with a learner's eventual passing or failing of the course. Moreover, it shows a steady decrease in learner engagement over time, not only among non-passing learners, but also among those who earned a certificate. Thus, attrition in MOOCs is not limited to those who do not pass the course; even the highest-achieving learners show a tendency of slowing down after the first week or two (also observed in [273]).

From Figures 7.8 and 7.9, we observe that passing and non-passing learners do not appear to differ in their rate of giving incorrect responses (which

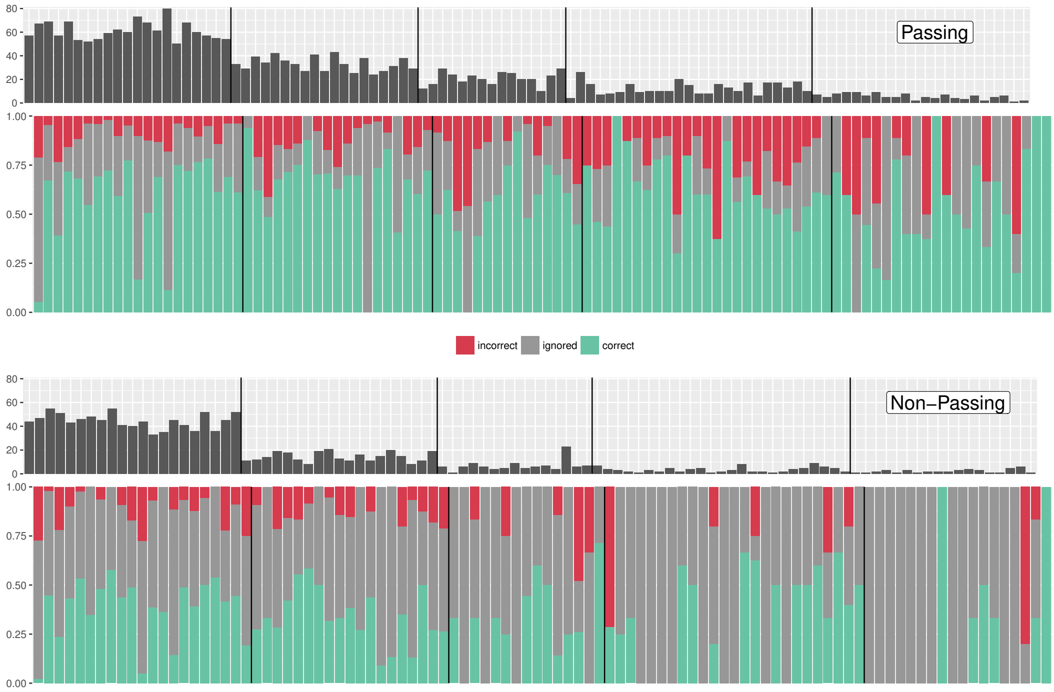


Figure 7.7: A question-by-question breakdown of every learner interaction with the `qCards`. The top two figures represent the behavior of passing learners—the upper image shows the number of learners being served that question, and the lower shows how they interacted with it—and the bottom two show that of non-passing learners. Questions are shown in order of appearance in the course (left to right), and the solid vertical line indicates a change in course week. Best viewed in color.

would indicate misconceptions or a lack of understanding the materials). Instead, they differ in their choice to ignore the problems all together. When removing the instances of ignored `qCards` and focusing only on attempted problems (right-hand side of Table 7.3), we observe a significant albeit small difference (6% difference, $\chi^2 = 9.63$, $p = 0.002$) between the proportion of correct or incorrect responses between passing and non-passing learners (cf. Table 7.3). In other words, passing and non-passing learners both perform about the same on these quiz problems—and yet, with no discernible difference in their assessed knowledge, only some go on to earn a passing grade and course certificate.

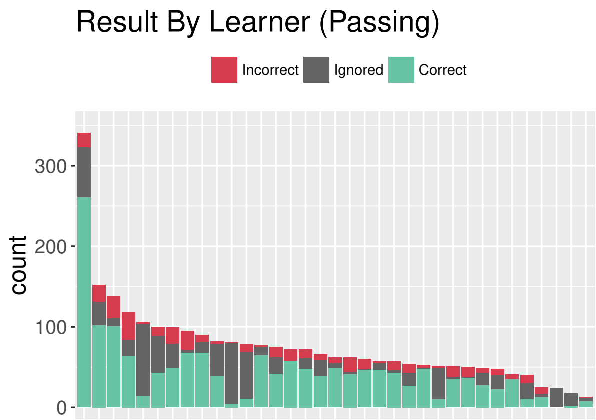


Figure 7.8: Each bar corresponds to one passing learner. Only one learner took advantage of the "infinite quizzing" capability by frequently using the "Generate new `qCard`" button. Best viewed in color.

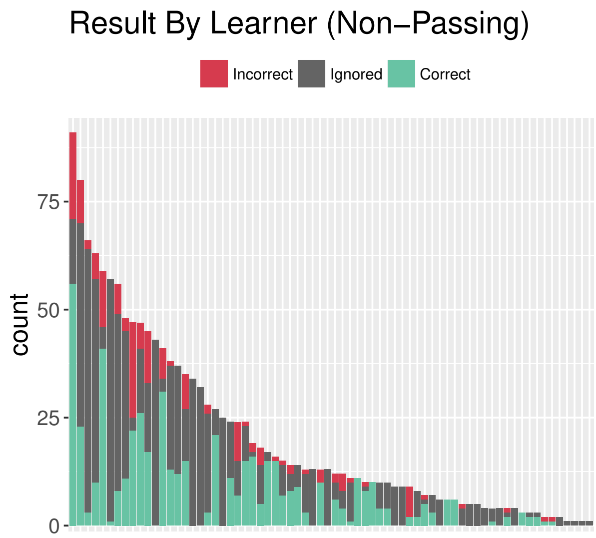


Figure 7.9: Each bar corresponds to one non-passing learner. Best viewed in color.

First Question Response

To further explore the predictive power of a learner's choice to either attempt or ignore the `qCards`, we next analyzed each learner's first interaction with a `qCard`. Figure 7.10 (left) shows the passing rate of learners segmented according to their first interaction with a `qCard`. Learners who attempted rather than ignored the first `qCard` had a 47% chance of passing the course.

Table 7.3: `qCard` problem response (left) and correctness (right). Significant differences at the $p < 0.001$ (between passing and non-passing) are indicated with †.

	Attempted†	Ignored†	Correct	Incorrect
Non-passing	0.47	0.53	0.76	0.24
Passing	0.73	0.27	0.82	0.18

In contrast, learners who ignored the first `qCard` delivered to them only had a 14% chance of passing. Figure 7.10 (right) additionally illustrates the relationship between the result of the first `qCard` attempt and passing the course. There were notably few learners who responded incorrectly, but their chance of passing the course was still relatively high at 33% compared to those who simply ignored the `qCard`.

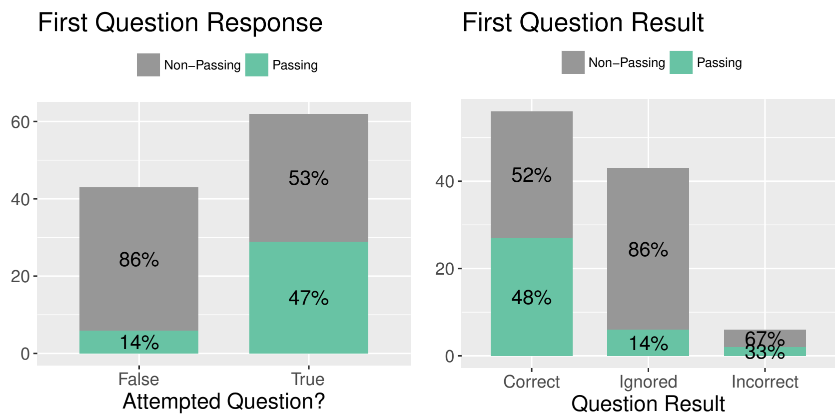


Figure 7.10: The likelihood of course completion based on learners' response (left) and result (right) to the first `qCard` they were shown. "True" indicates both correct or incorrect responses, and "False" indicates the `qCard` was ignored. Best viewed in color.

To evaluate whether the response of a learner's second `qCard` problem adds any predictive value, we replicated the analysis shown in Figure 7.10 for the responses to the first two `qCards` delivered to each learner. No difference in predictive value was observed by considering the second consecutive response—learners who answered their first two consecutive `qCards` correctly had a 53% chance of earning a passing grade.

We conclude that initial adoption of `ARPS` appears to depend partly on learners' motivation to complete the course.

Response Duration

We next explore how much time learners spent interacting with `qCards` and how time spent predicts the outcome of the interaction. Figure 7.11 shows the proportion of correct, incorrect, and ignored responses as a function of time elapsed with a `qCard`. We find that the decision to ignore the `qCard` happened quickly, with a median duration of 7 seconds (from the time the `qCard` appeared to the time the learner clicked the "x" button to close it). When learners did attempt to answer the question, the amount of time they spent did not have any association with the correctness of their response; the median duration for correct and incorrect responses was 18 seconds and 16 seconds, respectively.

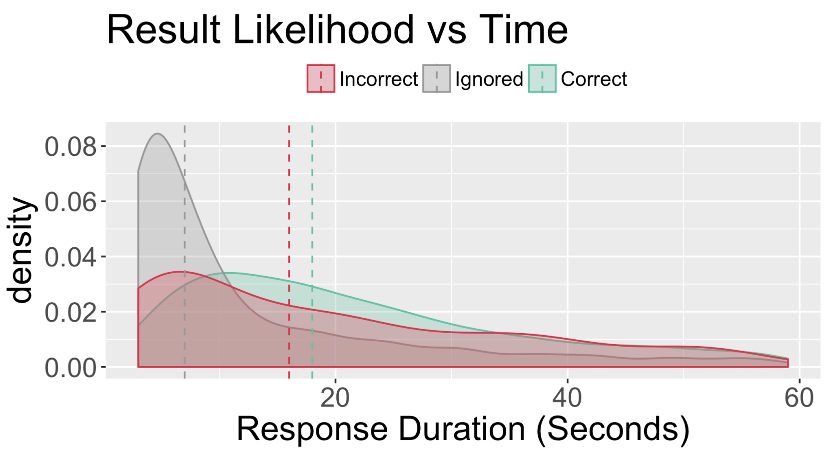


Figure 7.11: Kernel Density Estimation plot showing the relationship between time with the `qCard` visible and the result of a learner's response. The median time for each result is indicated with the dashed vertical line. Best viewed in color.

From the question-by-question, first question response, and response duration analyses, we conclude:

> There is no significant difference in assessed knowledge between passing and non-passing learners; the key difference lies in a learner's willingness to engage with the retrieval practice questions.

7.6.3 Modeling Knowledge Over Time

One of the contributions of `ARPS` is the data set that it generates: by tracking learners' responses to these periodic, formative, and ungraded questions throughout the entire course, we have a longitudinal account of learners' evolving knowledge state throughout the entire process of instruction. In this section we explore how learners' knowledge (as measured by performance with the `qCards`) deteriorates over time (**RQ7.3**).

Figure 7.12 shows the cumulative week-by-week performance of both passing and non-passing learners. As `qCards` could only be delivered with questions coming from *prior* course weeks, the x-axis begins with Week 2, where only questions from Week 1 were delivered. This continues up to Week 6 where questions from Weeks 1–5 could be delivered.

The left (Passing) graph in Figure 7.12 illustrates the forgetting curve of the passing learners in `GeoscienceX`. We observe a statistically significant decrease in performance between Weeks 2 and 6 (correct response rate dropping from 67% to 49% respectively; $\chi^2 = 32.8$, $p < 0.001$). While the proportion of ignored responses remains steadily low, the proportion of correct responses drops by 18% (nearly identical to the forgetting rate found in [166]). The rate of incorrect responses increased from 4% to 25% ($\chi^2 = 87.8$, $p < 0.001$).

On the right (Non-Passing) graph in Figure 7.12 we observe that the choice to ignore `qCards` was common through the entire course duration, with a slight increase in the later weeks. We also observe a significant decrease in correct response rates for non-passing learners ($\chi^2 = 15.7$, $p < 0.001$). However, unlike passing learners who exhibited a significant increase in incorrect responses, there is no significant change for non-passing learners. The change, instead, is in the rate of ignored responses, increases from 47% in Week 2 to 61% in Week 6.

We identify two main contributing factors to this decline in performance over time. First, the amount of assessed content increases each week; in Week 6 there are five course weeks worth of content to be assessed, whereas in Week 2 there is only content from Week 1 being assessed. Second, people simply forget more with the passing of time [211]; each passing week moves the learner temporally farther away from when the content was initially learned.

We next explore the relationship between testing delay and learners' memory and performance on `qCards`. In Figure 7.13, the x-axis represents the difference between a learner's current week and the week from which the `qCard` came. For example, if a learner was currently watching a lecture video in Week 5 and the `qCard` delivered was a question from Week 2, that would be a difference of three. While Figure 7.12 shows how the amount of content covered/assessed is related to performance, Figure 7.13 illustrates how the testing delay is related to performance.

We observe very similar trends as above for both passing and non-passing learners. For passing learners there is a 23% drop in correct response rates from 1 Week Elapsed to 5 Weeks Elapsed (65% to 42%, $\chi^2 = 23.6$, $p < 0.001$). Also significant is the 13% increase in incorrect response rate (8% to 21%, $\chi^2 = 17.5$, $p < 0.001$). The increase in ignored question frequency is not significant for passing learners, though it is large and significant for non-passing learners: between 1 Week Elapsed and 5 Weeks Elapsed, ignored questions increased by 22% from 50% to 72% ($\chi^2 = 4.9$, $p = 0.025$). Overall, for non-passing learners, we observe increased ignoring, decreased correct problem attempt rates, and steady incorrect problem attempt rates.

This pattern shows that non-passing learners are able to recognize, attempt, and correctly answer `qCard` problems that are more proximate to their current stage in the course. This suggests a high level of self-efficacy especially among the non-passing learners; they are able to identify questions that they likely do not know the answer to and choose to ignore them.

Another encouraging finding from this analysis is that of learners' short-term knowledge retention. As partially illustrated by Figure 7.13, considering problems that were attempted with 1 Week Elapsed, passing learners answer 88% of problems correctly. Non-passing learners also show good performance with 79% correct (note that the required passing grade for the course was 60%).

From the above findings on learner knowledge as a function of both time and course advancement, we conclude:

> Learner quiz performance deteriorates with the introduction of more course concepts/materials and the passing of time.

7.6.4 Long-Term Knowledge retention

Long-term knowledge retention is the primary learning outcome affected by highly-spaced retrieval practice, which is typically evaluated in either a final, cumulative exam in a course, or a post-exam with some lag time between learners' exposure to the material and assessment [166, 61]. As the `GeoscienceX` course only featured weekly quizzes, we took a random selection of ten quiz questions from throughout the six end-of-week quizzes and created a post-course knowledge assessment. Delivered to learners in a survey format

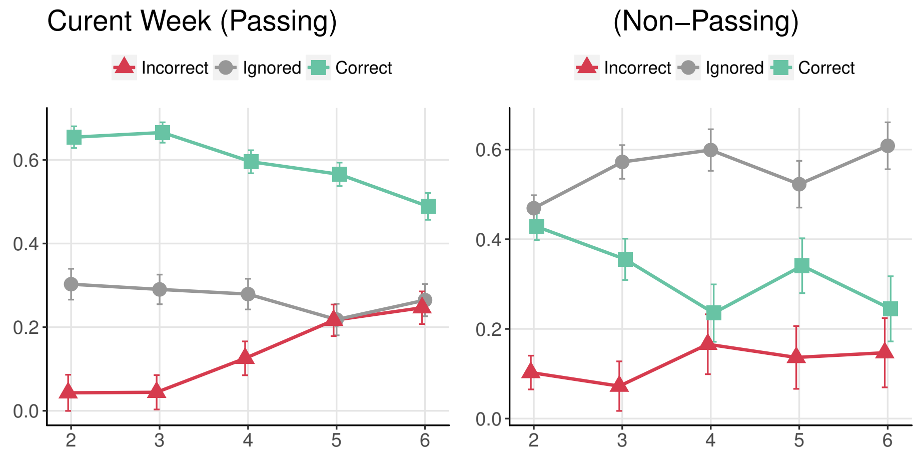


Figure 7.12: Week-by-week results of learners' interaction with the `qCards`. The x-axis represents the course week w, and the y-axis represents the proportion (%) of correct, incorrect, or ignored responses (with `qCards` showing queries from course weeks 1 to $w-1$). Error bars show standard error. Best viewed in color.

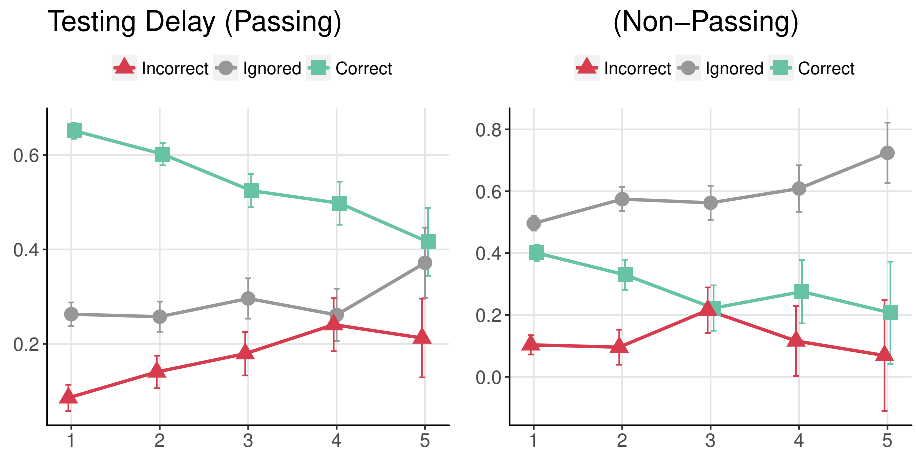


Figure 7.13: The x-axis represents the number of elapsed weeks between the course week where the topic was introduced and the course week in which the `qCard` was delivered (testing delay), and the y-axis represents the proportion of each response. Error bars show standard error. Best viewed in color.

two weeks after the course had ended, we compared the performance between the two experimental conditions.

A χ^2 test revealed no significant difference in long-term knowledge retention between the control condition and learners in the treatment condition who interacted with the intervention at least once (**RQ7.3**). The mean score for learners in the control and treatment conditions was 6.2 ($SD = 1.9$) and 6.6 ($SD = 1.8$), respectively, out of a possible 10 points ($N = 20$, $t(17.6) = -0.45$, $p = 0.66$).

Results from these analyses are consistent with prior literature [166, 61] on long-term knowledge retention in finding that, regardless of experimental condition and whether or not a learner passed the course:

> Approximately two thirds of course knowledge is retained over the long-term.

7.6.5 Learner Experience

To evaluate learners' experience with `ARPS`, we adapted the System Usability Survey [28] for the context of the present research. The scale was included in the post-course survey and learners indicated a cumulative usability score of 73.9 ($SD = 12.2$) on the SUS scale. According to [13], this is categorized as "acceptable usability" corresponding to a "C" grade. This means that the system's usability falls into the third quartile of SUS scores overall [13]—especially positive given that this was deployed not as a production system but as a research prototype.

To gain deeper insight into learners' experience and find out which specific aspects of the system could be improved, we also offered learners the opportunity to describe their experience with `ARPS` in two open response questions. One prompted them to share which aspects of `ARPS` they found to be the most enjoyable and another asked about frustrating aspects of `ARPS`.

One learner explained how the type of problem delivered was a key factor in their use of `ARPS`:

> "It [would] be better if only conceptual questions [were] asked for [the] pop quiz, it's troublesome if calculation is required. If calculation is required, I would prefer that the options are equations so that we can choose the right equation without evaluating them."

Other learners reported similar sentiments and also shared insights that indicate a heightened level of self-awareness induced by the `qCards`. Learners shared their perspectives talking about how the system helped *"...remind me [of] things that I missed in the course"* and how it gave them *"the chance to see what I remembered and what I had learned."* These anecdotes are encouraging as for these learners the system was able to encourage a deliberate activation of previously-learned concepts which may have otherwise been forgotten.

Upon seeing the learner feedback about how the problem type affected the learner's experience, we conducted a follow-up analysis to see if there was any indication that other learners felt the same way (as expressed through their interaction with `ARPS`). Figure 7.14 reveals that, indeed, this learner was not alone in their sentiment; we find that there was a 69% likelihood of learners attempting a MC `qCard` problem type compared to 41% attempt rate for NI problems. A χ^2 test shows this difference to be statistically significant ($p < 0.001$). Given that the question type (mostly evaluations of mathematical equations) is consistent across both problem types (MC and NI), we can conclude that these differences are indeed an effect of the problem type. This finding supports our initial design decision for a hyper-efficient interaction process—learners are far more likely to attempt a problem which only requires a single click selecting from a list of answers than one that requires two extra processes: they must first generate an answer from scratch and then type it out. From the data we are unable to identify which of these two extra processes contributes more to the problems being ignored, so we consider them in tandem.

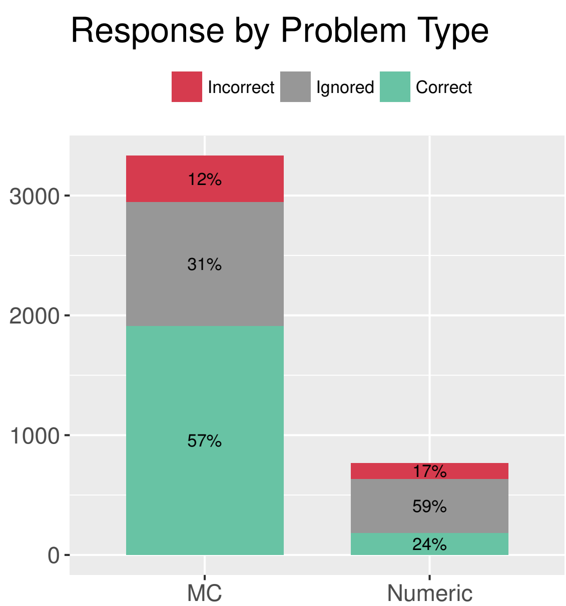


Figure 7.14: Breakdown of `qCard` interaction results across the two problem types. Best viewed in color.

7.7 Conclusion

Decades of prior research on the effectiveness of different learning strategies has found retrieval practice to be effective at supporting long-term knowledge retention [4, 49, 212, 97, 166, 118, 117, 61, 166]. However, how to effectively support retrieval practice in digital learning environments has not yet been thoroughly examined. The vast majority of prior work was conducted in offline learning environments, including university laboratory settings. Critically evaluating the efficacy of retrieval practice in digital learning environments promises to advance theory by developing a deeper understanding of how retrieval practice can be effective in a digital context as well as in a highly heterogeneous population that is embodied by MOOC learners.

We evaluated an `Adaptive Retrieval Practice System` in a MOOC to address the emerging issue of supporting learning strategies at large scale and to bridge retrieval practice theory into the digital learning space. We found noncompliance to be a major limitation in our evaluation of the system and its effectiveness. Many learners did not engage with the intervention, which limits our ability to draw causal inferences about the effect of retrieval practice on learners' achievement and engagement in the course.

We acknowledge the following limitations of the present study: (i) the `qCards` could potentially act as a distraction and make a learner more inclined to disengage, and (ii) despite the course being designed by trained course developers, there is a possibility that the assessment items used may not effectively measure the psychometric properties of learning, which would threaten the validity of our claim that retrieval practice does not improve learning outcomes.

Despite the lack of causal findings, the data collected from `ARPS` allowed us to offer multiple insights into the online learning process as it pertains to the persistence and transience of knowledge gains. By examining learner behavior and engagement with the intervention, we were able to track their performance on the same problem or topic and observe how their performance is affected by both the passage of time and introduction of new course materials.

We observed an encouraging trend of learners showing high levels of short- and medium-term knowledge retention, which is indicative of the early stages of learning. To what extent this newly gained knowledge is integrated into long-term memory warrants further research in the context of large online courses. Despite the null results from our causal analyses (Section 7.6.1), the wealth of evidence showing that retrieval practice is one of the most effective strategies to support knowledge retention makes this approach ripe for further investigation in online learning settings. The key to developing this theory further, however, is to design systems and interfaces that foster high levels of engagement to collect more causal evidence.

Chapter 8

Conclusion

The work presented in this thesis uses and advances *learning analytics* techniques to computationally model learning & teaching behaviors in online education. We also here develop technical solutions to support and improve this modeling. We present a series of research efforts which begins by gaining an understanding of how both instructors and learners alike behave without our intervention in edX MOOCs and then the work culminates in a series of randomized controlled trials evaluating the effectiveness of interventions that have been adopted from traditional learning environments previously found to be highly reliable and effective.

These lines of research presented here are centered on the primary research question of this thesis:

RQ How does the design of Massive Open Online Courses affect learner success and engagement?

We conduct deep inquiry into this question newly afforded by the widespread use of online learning platforms by taking advantage of both the clickstream data scale and granularity offered by MOOC platforms like edX. These new affordances led to a rise of learning analytics research which has offered new insights into the online learning process. We here address the identified shortcomings of online learning behavior (namely self-regulation and active learning) and develop interventions to improve learning behavior with two sub-RQ's:

RQ1 To what extent do teaching and learning strategies that have been found to be effective in traditional learning environments translate to MOOCs?

RQ2 How can MOOC environments be improved to advance the possibilities of experimentation?

In this thesis we have addressed these questions through: (i) modeling the default or natural behavior of instructors (how they design courses) and learners (how they engage with courses), (ii) reviewing the literature in seeking the most promising solutions for approaches to improve large-scale online active learning behavior, (iii) study planning interventions, and (iv) retrieval practice interventions. After presenting the summary of contributions from this thesis, we conclude this chapter with a prospective outlook of recommended avenues of future research to continue the momentum of this body of research.

8.1 Summary of Contributions

8.1.1 Improving Learning Behavior & Course Design

In Chapter 2 we have contributed a synthesis of the published literature in the domain of large-scale learning intervention studies. By limiting the inclusion criteria to studies that include adult learners as participants, employ a randomized controlled trial design, and test the effectiveness of a scalable instructional intervention, this work offers a state of the in the type of active learning strategies most likely to transfer to the MOOC context and learner demographics. In addressing **RQ1** (To what extent do teaching and learning strategies that have been found to be effective in traditional learning environments translate to MOOCs?), we find from our review that studies with more than 500 participants are the least likely to yield significant results and we also identify the three most promising types of interventions most likely to be effective in helping MOOC learners improve their engagement and learning outcomes: simulations & gaming, interactive multimedia, and cooperative learning.

8.1.2 Teaching & Learning Paths

While previous research in the area of online learning pathways found that learners often adopt nonlinear trajectories through online courses [260, 63], Chapters 3 and 4 expands the prior understanding by contextualizing learners' learning paths against the designed path set forth by the instructor. In our research efforts in this domain we developed methodologies to model both the behavior of learners and the design tendencies of course instructors. This work has major implications for both researchers and online course designers. Researchers have already begun to build off of this work in exploring at which moments in a learner's pathway in a course they change their behavior and deviate from an established path [203], thus offering an even more actionable insight for the instructor to intervene if this change is deemed undesirable. With regard to **RQ2** (How can MOOC environments be improved to advance the possibilities of experimentation?), we find that course designers can greatly benefit from this line of research because it begins to formalize the causal effect between course design and learner achievement [85, 86]. By regarding the design of a course as a dataset worthy of analysis, we can continue to work towards a concrete understanding of how certain structural or course design elements directly effect learners and their course learning outcomes.

8.1.3 Study Planning

Due to its established effectiveness in traditional learning settings, we attempted to to adapt study planning interventions to large-scale learning environments to evaluate their effectiveness in a new context in Chapters 5 and 6.

Social Comparison We next developed an intervention for MOOCs which provided learners with a dashboard-type visualization tool which they could use to both measure and track their own learning behavior, but also compare their behavior to a previously successful learner in the same course (to elicit social comparison). With regard to **RQ2**, we ran this intervention in randomized controlled trials in four separate MOOCs and found that the social comparison intervention significantly increased passing rates in every course examined (the increases in passing rates ranged from 3%–6%). In these experiments we successfully leveraged the scale of MOOCs (by presenting the aggregate behavior of all successful learners) in offering a way to objectively improve learner achievement.

Study Planning in MOOCs We began by exposing learners to a simple study planning intervention in a MOOC. We created an interface which prompted learners at the beginning of each week to state their goals and their plans on how they will achieve those goals for the week. And at the end of each course week, the system prompted learners to reflect back on their goals and plans and write down how well they did in achieving those. While this type of intervention had previously been found to be highly effective in traditional learning environments, we observed high levels of noncompliance—learners largely ignored the intervention. This served as an early indication that more care needs to be paid to translating interventions to the online context and that they cannot simply be adopted "out of the box."

We next designed, developed and evaluated an interactive SRL intervention (`SRLx`) for MOOC learners to plan and monitor their activities in the course. We analyzed learners' tendencies to actually use the system (compliance) as well as their actual engagement with the system. In service of **RQ2**, we find that learners were significantly more conservative with the way they planned to allocate time to the course compared to other metrics and we also observed that, in the particular course in which this was evaluated, learners' engagement with the system revealed that they were predominantly intrinsically motivated and not taking the course for direct career advancement.

8.1.4 Retrieval Practice

Retrieval practice has been touted as one of the most effective learning strategies known for improving long-term knowledge retention [4]. Accordingly, in Chapter 7 we conducted two experiments which evaluated the effectiveness of retrieval practice interventions in MOOCs.

Retrieval Practice in MOOCs In this pilot study, we conducted a randomized controlled trial evaluating the effectiveness of providing learners with retrieval prompts after each lecture video. We developed a system which prompted the learner to state, in their own words in 3–5 sentences, the key points from the previous lecture video. The theory behind this type of intervention is that this encourages learners to pause and actively recall the information they learned about in the video instead of quickly moving on to the next activity without taking the time to process the information. With regard to **RQ2**, we again found noncompliance to be a major issue in this research. Learners were not required to engage with the retrieval prompts, and the vast majority indeed chose not to. This finding serves as yet another indication that new considerations and measures need to be taken in translating interventions from traditional to online learning environments.

Modeling Knowledge Over Time To work toward a reliable way to get MOOC learners to engage in retrieval practice so that they can reap the benefits of improved long-term knowledge retention, we developed `ARPS`, an adaptive retrieval practice system. This system automatically delivered quiz questions from previous units in the course (that the learner had previously been exposed to) at random moments throughout the course. By doing so, the system encouraged and prompted learners to reactivate the knowledge they had learned weeks ago but had likely not reactivated since—and a regular schedule of reactivation is the key to encoding knowledge in long-term memory. We A/B tested `ARPS` in a randomized controlled trial and observed high levels of noncompliance. We observed insignificant results when comparing the key learning outcomes of the control vs. treatment groups. Even though the A/B test results were null, the data generated from `ARPS` (intermittent assessment events) enabled us to quantify the evolution (and degradation) of learners' knowledge over time. With regard to **RQ2**, we here model a forgetting curve of MOOC learners' knowledge and also find that MOOC learners rate of forgetting is similar to that of students in traditional learning environments, where approximately 2/3 of knowledge is retained over the long term [79, 61, 166].

8.2 Future Work

This thesis has made numerous contributions in the domain of large scale learning analytics from various perspectives—from the theoretical underpinning of self-regulated learning to the technical development of real-time data capturing (`SRLx`) and feedback in the edX platform. There is, however, still lots of work that needs to be done in this space to improve the quality of online education at scale [114] beyond the interventions presented above. The potential of massive learning environments is so vast, and learning analytics as a field is still in its infancy (with its first focused conference taking place in 2011[1]). While the present body of work has made a number advances to the field, this section outlines the limitations of the current state of the field and highlights the lines of research we see as having the greatest potential for widespread impact. Based on our findings from the preceding chapters, we here look beyond the current state of MOOCs and offer opportunities for the field of large scale learning analytics to help shape the future of higher education in three key domains: (i) Empirically Evaluating Instructional Strategies, (ii) Modeling Teaching & Learning Paths, and (iii) Digital Learning Theory.

Empirically Evaluating Instructional Strategies

As discussed in Chapter 2, the scalable nature of online learning platforms now affords researchers to conduct studies of unprecedented scale and heterogeneity in the learning context. The massive scale allows researchers to zoom in to underrepresented populations [135] in the learning literature with high statistical power to draw causal conclusions not previously common in the literature. This topic has recently begun to gain widespread attention in the learning analytics community, where new workshops [30] and recent experiments [29, 67] have been increasingly present.

To achieve this goal of ubiquitous experimentation and thorough evaluation of optimal teaching and learning strategies, we believe the online education community (practitioners and researchers alike) must adopt a culture where we are constantly conducting experiments and testing new approaches to instruction and pedagogy [141]. While some may already engage in this practice in isolation, the greatest impact can be had by the widespread adoption of this approach, because one experiment's results alone are not enough. To truly create a field of robust science, interventions must be replicated and

[1] https://tekri.athabascau.ca/analytics/

validated by multiple studies in a wide variety of contexts to work towards reproducible scientific findings [241, 52, 205]. For example, in Chapter 5, we introduce the learning tracker and find that it significantly increased passing rates of learners in four different MOOCs. While that is indeed promising preliminary evidence for the effectiveness of this intervention, it needs to be evaluated and reproduced in a wide range of contexts before it becomes a standard best practice that can be expected to work in any situation. And while peer-reviewed publication of such findings would serve as meaningful contributions to the advancement of this science, it is just as important for instructors and practitioners not interested in publishing to engage in a habit of iteratively improving their pedagogical practices by testing new innovations in their teaching strategies to facilitate more effective learning [141]. Inspiration for these experiments can and should be drawn from the substantial body of existing research generated from experiments carried out in traditional learning environments [96] so that we can continue to refine our understanding of how and which teaching & learning strategies are effective at scale.

We also encourage future research to enable new types of learning data to consider when conducting learning analytics research through the interdiction of new technologies. Whereas in the early years of MOOC research, the focus was to process and comprehend the novel, massive data sets generated by these new courses; now that there are a number of solutions to handling MOOC data[2], implementing new technology within the learning context to augment the existing collection of data will offer novel insights into the online learning process. For example, in Chapters 6 and 7, where we introduce `SRLx` and `ARPS` respectively, we created systems as interventions which track and generate new types of data not previously available through the edX environment. In `SRLx`, by enabling learners to create personalized study plans, we thus were able to draw new insights about how learners plan and go about achieving their plans. With `ARPS`, by intermittently delivering retrieval practice questions to learners and recording their responses, we were able to quantitatively track the evolution of learners' knowledge over time. More systems like `SRLx` and `ARPS` should be developed and deployed in the future to (i) improve learning outcomes and (ii) generate new data to improve research and understanding of online learning behavior. Future research should focus on developing such technologies with a strong emphasis on usability, human–computer interaction, and engagement. The systematic problem of noncompliance in the research presented in this thesis is highlights a pressing area for future work, and the online learning community

[2]https://github.com/MOOCdb

must find new ways to consistently and meaningfully engage online learners accordingly.

Teaching & Learning Paths

Learning is a long, highly complex process of knowledge building, and the clickstream data from online learning platforms offers an unprecedented level of granularity in our ability to gauge and track a learner's path through a learning experience. Prior to the widespread adoption of online learning environments, research in this domain would be limited to an analysis of the instructor's syllabus, or resource-intensive manual observing of learners' trajectories through tasks or activities in order to uncover insights in how knowledge is formed in the learning context [156]. But now, in the data researchers have access to every action a learner takes within a platform, from opening a new page, to changing the playback speed of a video, to scrolling through their peers' comments on the course discussion forum.

A formidable extension to this previous work on pathways and navigation patterns would be to dovetail it with the recent ongoing work in learner feedback [201, 38]. For example, research on self-regulated learning is typically framed in one of a few widely regarded models [175, 132, 264]. As those models were developed in the context of traditional learning environments without out the data granularity afforded by large-scale learning environments, we are now faced with the opportunity to understand self-regulated learning behavior in far greater detail than was possible before; this could lead to the formulation of new models and/or a more nuanced understanding of the suitability of existing models as applied in the large-scale online context.

Chapters 3 and 4 in this thesis emphasized the value in this data particularly in regarding it through a lens of patterns or sequences. By making sense of these patterns and sequences of learner behavior through sequence-based modeling techniques, future research can continue to evaluate how a learner's path through the course is related to their future behavior and learning outcomes. Furthermore, and perhaps more importantly, a better understanding of the causal effect of various traits of learner pathways can inform interventions which, for instance, could detect a deviation of a learner towards a potentially undesirable path and delivers a targeted intervention nudging or directing them back on a path more likely to lead to positive outcomes.

Digital Learning Theory

One of the most substantial implications of the research presented in this thesis is that we have highlighted the need for a new area of concentration in learning science that is concerned with the online learning context—we must continue to identify deviations where what works in a traditional learning environment does not effectively work online. For example, the experiments carried out in Chapters 5–7 of this thesis each used best practices in designing the instructional interventions in our effort to adapt them from the small-scale, in-person lab setting in which they were originally evaluated and validated to the novel context of MOOCs. It has become clear from this body of research that interventions and techniques known to work in the traditional classroom cannot simply work "out of the box" in the online setting—rather, they must be adapted accordingly based on the context. This thesis highlights the importance and difficulty of this adaptation process; it shows some examples of successful adaptations along with some unsuccessful adaptations. Using the research efforts on study planning and retrieval practice in Chapters 5–7 as a basis, future research should continue to empirically evaluate innovations to instructional strategies at scale in creating a culture of iteratively building the understanding of how teaching and learning works at scale.

The goal of this line of research is to arrive at a set of best practices where we can declare with certainty that if a course is designed using a certain set of design principles (both in terms of interventions to improve SRL and instructional sequencing strategies), then it will give the most learners the greatest chance of succeeding, with the ultimate goal of personalized learning—identifying each learner's individual needs at a given moment and acting upon that accordingly.

Bibliography

[1] Andrew Abbott and Angela Tsay. Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological methods & research*, 29(1):3–33, 2000.

[2] Ali Farhan AbuSeileek. The effect of computer-assisted cooperative learning methods and group size on the EFL learners' achievement in communication skills. *Computers & Education*, 58(1):231–239, January 2012.

[3] Ali Farhan AbuSeileek and Khaleel Qatawneh. Effects of synchronous and asynchronous computer-mediated communication (CMC) oral conversations on English language learners' discourse functions. *Computers & Education*, 62:181–190, March 2013.

[4] Olusola O Adesope, Dominic A Trevisan, and Narayankripa Sundararajan. Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, February 2017.

[5] Abdullah S Aldera and Mohammed Ali Mohsen. Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education*, 68:60–75, October 2013.

[6] Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

[7] Ian Arawjo, Cheng-Yao Wang, Andrew C Myers, Erik Andersen, and François Guimbretière. *Teaching Programming with Gamified Semantics.* ACM, 2017.

[8] Yigal Attali. Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems. *Computers & Education*, 86:260–267, August 2015.

[9] Yigal Attali and Meirav Arieli-Attali. Gamification in assessment: Do points affect test performance? *Computers & Education*, 83:57–63, April 2015.

[10] Yigal Attali and Fabienne van der Kleij. Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education*, 110:154–169, July 2017.

[11] Katherine A Austin. Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers & Education*, 53(4):1339–1354, 2009.

[12] Ryan S.J. Baker, Albert T. Corbett, Kenneth R. Koedinger, and Shelley Evenson *et al.* Adapting to when students game an intelligent tutoring system. In *Intelligent tutoring systems*, pages 392–401, 2006.

[13] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[14] Matthew Barr. Video games can develop graduate skills in higher education students: A randomised trial. *Computers & Education*, 113: 86–97, October 2017.

[15] Scott Bateman, Jaime Teevan, and Ryen W White. The search dashboard: how reflection and comparison impact search behavior. In *CHI '12*, pages 1785–1794, 2012.

[16] Sanam Shirazi Beheshitha, Marek Hatala, Dragan Gašević, and Srećko Joksimović. The role of achievement goal orientations when studying effect of learning analytics visualizations. In *LAK '16*, pages 54–63, 2016.

[17] Sameer Bhatnagar, Michel Desmarais, Chris Whittaker, Nathaniel Lasry, Michael Dugdale, and Elizabeth S. Charles. An analysis of peer-submitted and peer-reviewed answer rationales, in an asynchronous Peer Instruction based learning environment. In *EDM '15*, pages 456–459, 2015.

[18] Yulong Bian, Chenglei Yang, Dongdong Guan, Sa Xiao, Fengqiang Gao, Chia Shen, and Xiangxu Meng. Effects of pedagogical agent's personality and emotional feedback strategy on chinese students' learning experiences and performance. In *CHI '16*, pages 433–444. ACM, 2016.

[19] Hart Blanton, Bram P Buunk, Frederick X Gibbons, and Hans Kuyper. When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of personality and social psychology*, 76(3):420, 1999.

[20] Benjamin S Bloom. Learning for mastery. *Instruction and Curriculum*, 1(2):1–12, 1968.

[21] Robert Bodily and Katrien Verbert. Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 2017.

[22] Robert Bodily and Katrien Verbert. Trends and issues in student-facing learning analytics reporting systems research. In *LAK '17*, pages 309–318, 2017.

[23] Alexander Borek, Bruce M. McLaren, Michael Karabinos, and David Yaron. How Much Assistance Is Helpful to Students in Discovery Learning? In *EC-TEL '09*, pages 391–404. Springer, 2009.

[24] Mina Boroujeni and Pierre Dillenbourg. Discovery and temporal analysis of latent study patterns from mooc interaction sequences. In *LAK '18*. ACM, 2018.

[25] Lori EA Bradford and Norman McIntyre. Off the beaten track: Messages as a means of reducing social trail use at St. Lawrence Islands National Park. *Journal of Park and Recreation Administration*, 25(1): 1–21, 2007.

[26] Christopher G. Brinton, Ruediger Rill, Sangtae Ha, Mung Chiang, Robert Smith, and William Ju. Individualization for education at scale: Miic design and preliminary evaluation. In *IEEE TLT*, volume 8, pages 136–148, 2015.

[27] Cyril Brom, Edita Bromová, Filip Děchtěrenko, Michaela Buchtová, and Martin Pergel. Personalized messages in a brewery educational simulation: Is the personalization principle less robust than previously thought? *Computers & Education*, 72:339–366, March 2014.

[28] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[29] Christopher Brooks, Jim Greer, and Carl Gutwin. The data-assisted approach to building intelligent technology-enhanced learning environments. In *Learning Analytics*, pages 123–156, 2014.

[30] Christopher Brooks, Dan Davis, Nia Dowell, Joshua Gardner, Timothy Necamp, Oleksandra Poquet, Rene Kizilcec, and Joseph Williams. Diving in to educational experiments: Process, evaluation, and reasoning in support of learning (deeper support of learning). In *LAK '19*. ACM, 2019.

[31] Elizabeth J. Brown, Timothy J. Brailsford, Tony Fisher, and Adam Moore. Evaluating learning style personalization in adaptive systems: Quantitative methods and approaches. *IEEE TLT*, 2(1):10–22, 2009.

[32] Susan Bull and Judy Kay. Open learner models. In *Advances in intelligent tutoring systems*, pages 301–322. Springer, 2010.

[33] Susan Bull and Judy Kay. Open learner models as drivers for metacognitive processes. In *International handbook of metacognition and learning technologies*, pages 349–365. Springer, 2013.

[34] Engin Bumbacher, Shima Salehi, Miriam Wierzchula, and Paulo Blikstein. Learning Environments and Inquiry Behaviors in Science Inquiry Learning: How their Interplay Affects the Development of Conceptual Understanding in Physics. In *EDM '15*, pages 61–68, 2015.

[35] Nergiz Ercil Cagiltay, Erol Ozcelik, and Nese Sahin Ozcelik. The effect of competition on learning in games. *Computers & Education*, 87:35–41, September 2015.

[36] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[37] Julia Cambre, Chinmay Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. Talkabout: small-group discussions in massive global classes. In *L@S '14*, pages 161–162, 2014.

[38] David Carless and David Boud. The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, pages 1315–1325, 2018.

[39] Nicholas J Cepeda, Edward Vul, Doug Rohrer, John T Wixted, and Harold Pashler. Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention. *Psychological Science*, 19(11):1095–1102, November 2008.

[40] Mei-Mei Chang, Mei-Chen Lin, and Ming-Jen Tsai. A study of enhanced structured web-based discussion in a foreign language learning class. *Computers & Education*, 61:232–241, February 2013.

[41] Chih-Ming Chen and Chung-Hsin Wu. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, 80:108–121, 2015.

[42] Ching-Huei Chen. Promoting college students knowledge acquisition and ill-structured problem solving: Web-based integration and procedure prompts. *Computers & Education*, 55(1):292–303, 2010.

[43] Ming-Puu Chen, Li-Chun Wang, Hsiu-Ju Chen, and Yu-Chu Chen. Effects of type of multimedia strategy on learning of Chinese characters for non-native novices. *Computers & Education*, 70:41–52, January 2014.

[44] Nian-Shing Chen, Kinshuk, Chun-Wang Wei, and Chia-Chi Liu. Effects of matching teaching strategy to thinking style on learner's quality of reflection in an online learning environment. *Computers & Education*, 56(1):53–64, January 2011.

[45] Alan Cheng, Lei Yang, and Erik Andersen. *Teaching Language and Culture with a Virtual Reality Game*. ACM, New York, New York, USA, May 2017.

[46] Young Hoan Cho and Sang Eun Lee. The role of co-explanation and self-explanation in learning from design examples of PowerPoint presentation slides. *Computers & Education*, 69:400–407, November 2013.

[47] Katheryn R Christy and Jesse Fox. Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, 78:66–77, September 2014.

[48] Yeu-Hui Chuang and Chiung-Wen Tsao. Enhancing nursing students' medication knowledge: The effect of learning materials delivered by short message service. *Computers & Education*, 61:168–175, February 2013.

[49] Ruth C Clark and Richard E Mayer. *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons, 2016.

[50] Derrick Coetzee, Armando Fox, Marti A. Hearst, and Björn Hartmann. Should your MOOC forum use a reputation system? In *CSCW '14*, pages 1176–1187, 2014.

[51] Derrick Coetzee, Seongtaek Lim, Armando Fox, Bjorn Hartmann, and Marti A. Hearst. Structuring interactions for large-scale synchronous peer learning. In *CSCW '15*, pages 1139–1152, 2015.

[52] Open Science Collaboration et al. Estimating the reproducibility of psychological science. *Science*, 349(6251), 2015.

[53] R Cook and J Kay. The Justified User Model: A Viewable, Explained User Model. In *UM '94*, pages 145–150, 1994.

[54] Gemma Corbalan, Fred Paas, and Hans Cuypers. Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education*, 55(2):692–703, September 2010.

[55] Anna Cox, Paul Cairns, Pari Shah, and Michael Carroll. Not doing but thinking: the role of challenge in the gaming experience. In *CHI '12*, pages 79–88, 2012.

[56] Jason Crabtree and Adam Rutland. Self-evaluation and social comparison amongst adolescents with learning difficulties. *Journal of Community & Applied Social Psychology*, 11(5):347–359, 2001.

[57] Christian Crumlish and Erin Malone. *Designing social interfaces: Principles, patterns, and practices for improving the user experience*. O'Reilly Media, Inc., 2009.

[58] Gabriel Culbertson, Erik Andersen, Walker White, Daniel Zhang, and Malte Jung. Crystallize: An immersive, collaborative game for second language learning. In *CSCW '16*, pages 636–647. ACM, 2016.

[59] Gabriel Culbertson, Solace Shen, Malte Jung, and Erik Andersen. *Facilitating Development of Pragmatic Competence through a Voice-driven Video Learning Interface*. ACM, New York, New York, USA, May 2017.

[60] D Cunningham and T Duffy. Constructivism: Implications for the design and delivery of instruction. *Handbook of research for educational communications and technology*, pages 170–198, 1996.

[61] Eugène JFM Custers. Long-term retention of basic science knowledge: a review study. *Advances in Health Sciences Education*, 15(1):109–128, 2010.

[62] Maria Cutumisu and Daniel L Schwartz. Choosing versus receiving feedback: The impact of feedback valence on learning in an assessment game. In *EDM '16*, pages 341–346, 2016.

[63] Dan Davis, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Gauging mooc learners' adherence to the designed learning path. In *EDM '16*, pages 54–61, 2016.

[64] Dan Davis, Guanliang Chen, Ioana Jivet, Claudia Hauff, and Geert-Jan Houben. Encouraging metacognition & self-regulation in moocs through increased learner feedback. In *Proceedings of the LAK 2016 Workshop on Learning Analytics for Learners (LAL)*, 2016.

[65] Dan Davis, Guanliang Chen, Tim van der Zee, Claudia Hauff, and Geert-Jan Houben. Retrieval practice and study planning in moocs: Exploring classroom-based self-regulated learning strategies at scale. In *EC-TEL '16*, pages 57–71. Springer, 2016.

[66] Dan Davis, Ioana Jivet, René F Kizilcec, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Follow the successful crowd: raising mooc completion rates through social comparison at scale. In *LAK '17*, pages 454–463, 2017.

[67] Dan Davis, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Activating learning at scale: A review of innovations in online learning strategies. *Computers & Education*, 125:327 – 344, 2018. ISSN 0360-1315.

[68] Dan Davis, René F. Kizilcec, Claudia Hauff, and Geert-Jan Houben. The half-life of mooc knowledge: A randomized trial evaluating knowledge retention and retrieval practice in moocs. In *LAK '18*, LAK '18, pages 1–10, 2018.

[69] Dan Davis, Daniel Seaton, Claudia Hauff, and Geert-Jan Houben. Toward large-scale learning design: categorizing course designs in service of supporting learning outcomes. In *L@S '18*, pages 1–10. ACM, 2018.

[70] Dan Davis, Vasileios Triglianos, Claudia Hauff, and Geert-Jan Houben. Srlx: A personalized learner interface for moocs. In *EC-TEL 2018*, pages 1–14. Springer, 2018.

[71] Luc De Grez, Martin Valcke, and Irene Roozen. The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers & Education*, 53(1):112–120, August 2009.

[72] Björn B de Koning, Huib K Tabbers, Remy M J P Rikers, and Fred Paas. Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education*, 55(2):681–691, September 2010.

[73] David Dearman and Khai Truong. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *CHI '12*, pages 1391–1400, 2012.

[74] Michael Derntl, Susanne Neumann, and Petra Oberhuemer. Constructing and evaluating a description template for teaching methods. In *EC-TEL*, pages 447–461. Springer, 2009.

[75] Michael Derntl, Susanne Neumann, Petra Oberhuemer, Colin Tattersall, Dominique Verpoorten, and Roland Klemke. Report on the standardized description of instructional models. Technical report, University of Vienna, 2009.

[76] Michel C Desmarais and Ryan SJ d Baker. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1-2):9–38, 2012.

[77] Bernie Dodge, John Whitmer, and James P. Frazee. Improving undergraduate student achievement in large blended courses through data-driven interventions. In *LAK '15*, pages 412–413, 2015.

[78] Michael Eagle and Tiffany Barnes. Evaluation of Automatically Generated Hint Feedback. In *EDM '13*, pages 372–374, 2013.

[79] Hermann Ebbinghaus. *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Duncker & Humblot, 1885.

[80] S Erhel and E Jamet. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, 67:156–167, September 2013.

[81] Rebecca Ferguson and Doug Clow. Consistent Commitment: Patterns of Engagement across Time in Massive Open Online Courses (MOOCs). *Journal of Learning Analytics*, 2(3):55–80, 2016.

[82] Leon Festinger. A theory of social comparison processes. *Human relations*, 7(2):117–140, 1954.

[83] Sally Fincher and Josh Tenenberg. Making sense of card sorting data. *Expert Systems*, 22(3):89–93, 2005.

[84] Davide Fossati, Barbara Di Eugenio, Christopher W. Brown, Stellan Ohlsson, David G. Cosejo, and Lin Chen. Supporting computer science curriculum: Exploring and learning linked lists with iList. *IEEE TLT*, 2(2):107–120, 2009.

[85] Scott Freeman, David Haak, and Mary Pat Wenderoth. Increased course structure improves performance in introductory biology. *CBE-Life Sciences Education*, 10(2):175–186, 2011.

[86] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111 (23):8410–8415, 2014.

[87] Ujwal Gadiraju and Stefan Dietze. Improving learning through achievement priming in crowdsourced information finding microtasks. In *LAK '17*, pages 105–114. ACM, 2017.

[88] Dilrukshi Gamage, Mark E Whiting, Thejan Rajapakshe, Haritha Thilakarathne, Indika Perera, and Shantha Fernando. Improving assessment on moocs through peer identification and aligned incentives. In *L@S '17*, pages 315–318. ACM, 2017.

[89] Chase Geigle and ChengXiang Zhai. Modeling student behavior with two-layer hidden markov models. *Journal of Educational Data Mining*, 9(1):1–24, 2017.

[90] Michele J Gelfand, Jana L Raver, Lisa Nishii, Lisa M Leslie, Janetta Lun, Beng Chong Lim, Lili Duan, Assaf Almaliach, Soon Ang, Jakobina Arnadottir, et al. Differences between tight and loose cultures: A 33-nation study. *Science*, 332(6033):1100–1104, 2011.

[91] José A González, Lluís Jover, Erik Cobo, and Pilar Muñoz. A web-based learning tool improves student performance in statistics: A randomized masked trial. *Computers & Education*, 55(2):704–713, September 2010.

[92] Elena Barberà Gregori, Jingjing Zhang, Cristina Galván-Fernández, and Francisco de Asís Fernández-Navarro. Learner support in moocs: Identifying variables linked to completion. *Computers & Education*, 122:153–168, 2018.

[93] Julio Guerra, Roya Hosseini, Sibel Somyurek, and Peter Brusilovsky. An Intelligent Interface for Learning Content: Combining an Open Learner Model and Social Comparison to Support Self-Regulated Learning and Engagement. In *IUI '16*, pages 152–163, 2016.

[94] Philip J Guo and Katharina Reinecke. Demographic differences in how students navigate through MOOCs. In *L@S '14*, pages 21–30, 2014.

[95] Sherif Halawa, Daniel Greene, and John Mitchell. Dropout prediction in moocs using learner activity features. In *Experiences and best practices in and around MOOCs*, volume 7, pages 3–12, 2014.

[96] John Hattie. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge, 2008.

[97] III Henry L Roediger and Jeffrey D Karpicke. The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1(3):181–210, June 2016.

[98] E Tory Higgins. Promotion and prevention: Regulatory focus as a motivational principle. *Advances in experimental social psychology*, 30: 1–46, 1998.

[99] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. *Cultures and organizations: Software of the mind.* McGraw Hill, 1991.

[100] Nina Hood, Allison Littlejohn, and Colin Milligan. Context counts: How learners' Contexts Influence Learning in a MOOC. *Computers & Education*, 91:83–91, 2015.

[101] Danial Hooshyar, Rodina Binti Ahmad, Moslem Yousefi, Moein Fathi, Shi-Jinn Horng, and Heuiseok Lim. Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Computers & Education*, 94:18–36, March 2016.

[102] Hui-Wen Huang, Chih-Wei Wu, and Nian-Shing Chen. The effectiveness of using procedural scaffoldings in a paper-plus-smartphone collaborative learning context. *Computers & Education*, 59(2):250–259, September 2012.

[103] Pascal Huguet, Marie P Galvaing, Jean M Monteil, and Florence Dumas. Social presence effects in the stroop task: further evidence for an attentional view of social facilitation. *Journal of personality and social psychology*, 77(5):1011, 1999.

[104] Pascal Huguet, Florence Dumas, Jean M Monteil, and Nicolas Genestoux. Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European journal of social psychology*, 31(5):557–578, 2001.

[105] Maria L Hwang and Lena Mamykina. *Monster Appetite: Effects of Subversive Framing on Nutritional Choices in a Digital Game Environment.* Effects of Subversive Framing on Nutritional Choices in a Digital Game Environment. ACM, New York, New York, USA, May 2017.

[106] Maria-Blanca Ibanez, Angela Di-Serio, and Carlos Delgado-Kloos. Gamification for engaging computer science students in learning activities: A case study. *IEEE TLT*, 7(3):291–301, 2014.

[107] Birgit Imhof, Katharina Scheiter, Jörg Edelmann, and Peter Gerjets. Learning about locomotion patterns: Effective use of multiple pictures and motion-indicating arrows. *Computers & Education*, 65:45–55, July 2013.

[108] Susan Jang, Jonathan M Vitale, Robert W Jyung, and John B Black. Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, 106:150–165, March 2017.

[109] Lucas M Jeno, John-Arvid Grytnes, and Vigdis Vandvik. The effect of a mobile-application tool on biology students' motivation and achievement in species identification: A Self-Determination Theory perspective. *Computers & Education*, 107:1–12, April 2017.

[110] Hogyeong Jeong and Gautam Biswas. Mining student behavior models in learning-by-teaching environments. In *EDM '08*, pages 127–136, 2008.

[111] Ioana Jivet, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice. In *ECTEL '17*, pages 82–96, 2017.

[112] Yohan Jo, Gaurav Tomar, Oliver Ferschke, Carolyn Penstein Rosé, and Dragan Gasevic. Expediting Support for Social Learning with Behavior Modeling. In *EDM '16*, pages 400–405, 2016.

[113] Cheryl I Johnson and Richard E Mayer. A Testing Effect with Multimedia Learning. *Journal of Educational Psychology*, 101(3):621, 2009.

[114] David Joyner. Toward cs1 at scale: building and testing a mooc-for-credit candidate. In *L@S '18*, page 59. ACM, 2018.

[115] Anastasios Karakostas and Stavros Demetriadis. Adaptive domain-specific support to enhance collaborative learning: results from two studies. In *EC-TEL '11*, pages 210–219. Springer, 2011.

[116] Samad Kardan and Cristina Conati. Providing adaptive support in an interactive simulation for learning: an experimental evaluation. In *CHI '15*, pages 3671–3680, 2015.

[117] Jeffrey D Karpicke and Janell R Blunt. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018):772–775, 2011.

[118] Jeffrey D Karpicke and Henry L Roediger. The critical importance of retrieval for learning. *science*, 319(5865):966–968, 2008.

[119] Douglas F Kauffman. Self-regulated learning in web-based environments: Instructional tools designed to facilitate cognitive strategy use, metacognitive processing, and motivational beliefs. *Journal of educational computing research*, 30(1-2):139–161, 2004.

[120] Judy Kay. Learner know thyself: Student models to give learner control and responsibility. In *ICCE '97*, pages 17–24, 1997.

[121] Judy Kay. Stereotypes, student models and scrutability. In *International Conference on Intelligent Tutoring Systems*, pages 19–30, 2000.

[122] Mohammad Khalil and Martin Ebner. Moocs completion rates and possible methods to improve retention-a literature review. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 1305–1313, 2014.

[123] Nobel Khandaker and Leen-Kiat Soh. ClassroomWiki: A collaborative wiki for instructional use with multiagent group formation. *IEEE TLT*, 3(3):190–202, 2010.

[124] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. *Explaining the Gap: Visualizing One's Predictions Improves Recall and Comprehension of Data.* Visualizing One's Predictions Improves Recall and Comprehension of Data. ACM, New York, New York, USA, May 2017.

[125] Paul A Kirschner. Stop propagating the learning styles myth. *Computers & Education*, 106:166–171, 2017.

[126] Paul A Kirschner and Jeroen JG van Merriënboer. Do learners really know best? urban legends in education. *Educational psychologist*, 48 (3):169–183, 2013.

[127] René F Kizilcec and Christopher Brooks. Diverse big data and randomized field experiments in moocs. *Handbook of Learning Analytics*, pages 211–222, 2017.

[128] René F Kizilcec and Geoffrey L Cohen. Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences*, page 201611898, 2017.

[129] René F Kizilcec and Sherif Halawa. Attrition and Achievement Gaps in Online Learning. In *L@S '15*, pages 57–66. ACM, 2015.

[130] René F. Kizilcec and Emily Schneider. Motivation As a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. *TOCHI*, 22(2):6:1–6:24, 2015.

[131] René F Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *LAK 2013*, pages 170–179, 2013.

[132] René F Kizilcec, Mar Pérez-Sanagustín, and Jorge J Maldonado. Recommending self-regulated learning strategies does not improve performance in a mooc. In *L@S '16*, pages 101–104, 2016.

[133] René F Kizilcec, Glenn M Davis, and Geoffrey L Cohen. Towards equal opportunities in moocs: Affirmation reduces gender & social-class achievement gaps in china. In *L@S '17*, pages 121–130. ACM, 2017.

[134] René F Kizilcec, Mar Pérez-Sanagustín, and Jorge J Maldonado. Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & education*, 104: 18–33, 2017.

[135] René F Kizilcec, Andrew J Saltarelli, Justin Reich, and Geoffrey L Cohen. Closing global achievement gaps in moocs. *Science*, 355(6322): 251–252, 2017.

[136] René F. Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. Showing face in video instruction: effects on information retention, visual attention, and affect. In *CHI '14*, pages 2095–2102, 2014.

[137] René F. Kizilcec, Emily Schneider, Geoffrey L. Cohen, and Daniel A. McFarland. Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. In *eLearning Papers*, pages 17–26, 2014.

[138] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. Predicting MOOC dropout over weeks using machine learning methods. In *EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 60–65, 2014.

[139] M. Köck and A. Paramythis. Towards Adaptive Learning Support on the Basis of Behavioural Patterns in Learning Activity Sequences. In *Intelligent Networking and INCOS '10*, pages 100–107, 2010.

[140] Mirjam Köck and Alexandros Paramythis. Activity sequence modelling and dynamic clustering for personalized e-learning. *User Modeling and User-Adapted Interaction*, 21(1-2):51–97, 2011.

[141] Ken Koedinger. Practical learning research at scale. In *L@S '16*, pages 429–429. ACM, 2016.

[142] Kenneth R Koedinger, Albert T Corbett, and Charles Perfetti. The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science*, 36(5):757–798, 2012.

[143] Kenneth R. Koedinger, Elizabeth A. McLaughlin, Julianna Zhuxin Jia, and Norman L. Bier. Is the doer effect a causal relationship?: How can we tell and why it's important. In *LAK '16*, pages 388–397, 2016.

[144] Daphne Koller, Andrew Ng, Chuong Do, and Zhenghao Chen. Retention and intention in massive open online courses. *Educause Review*, 48(3):62–63, 2013.

[145] Johannes Konert, Henrik Bellhäuser, René Röpke, Eduard Gallwas, and Ahmed Zucik. Moodlepeers: Factors relevant in learning group formation for improved learning outcomes, satisfaction and commitment in e-learning scenarios using groupal. In *EC-TEL '16*, pages 390–396. Springer, 2016.

[146] Markus Krause, Marc Mogalle, Henning Pohl, and Joseph Jay Williams. A playful game changer: Fostering student retention in online education with social gamification. In *L@S '15*, pages 95–102, 2015.

[147] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott R. Klemmer. Talkabout: Making distance matter with small groups in massive classes. In *CSCW '15*, pages 1116–1128, 2015.

[148] Chinmay E. Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. Peerstudio: Rapid peer feedback emphasizes revision and improves performance. In *L@S '15*, pages 75–84, 2015.

[149] Bum Chul Kwon and Bongshin Lee. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *CHI '16*, pages 993–997. ACM, 2016.

[150] Hugues Labarthe, François Bouchet, Rémi Bachelet, and Kalina Yacef. Does a peer recommender foster students' engagement in moocs? In *EDM '16*, pages 418–423, 2016.

[151] Anne Lamb, Jascha Smilack, Andrew Ho, and Justin Reich. Addressing common analytic challenges to randomized experiments in MOOCs: Attrition and zero-inflation. In *L@S '15*, pages 21–30, 2015.

[152] Chung Hsien Lan, Sabine Graf, K. Robert Lai, and others. Enrichment of peer assessment with agent negotiation. *IEEE TLT*, 4(1):35–46, 2011.

[153] Charles Lang, Neil Heffernan, Korinn Ostrow, and Yutao Wang. The Impact of Incorporating Student Confidence Items into an Intelligent Tutor: A Randomized Controlled Trial. In *EDM '15*, pages 144–149, 2015.

[154] Peter Langfelder, Bin Zhang, and Steve Horvath. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Bioinformatics*, 24(5):719–720, 2007.

[155] Diana Laurillard. Technology enhanced learning as a tool for pedagogical innovation. *Journal of Philosophy of Education*, 42(3-4):521–533, 2008.

[156] Diana Laurillard. *Teaching as a design science: Building pedagogical patterns for learning and technology.* Routledge, 2013.

[157] Diana Laurillard and Dejan Ljubojevic. Evaluating learning designs through the formal representation of pedagogical patterns. *IGI Global*, 2011.

[158] Diana Laurillard and Patrick McAndrew. Reusable educational software: A basis for generic learning activities. *Reusing online resources: A sustainable approach to e-learning*, pages 81–93, 2003.

[159] Yi-Chieh Lee, Wen-Chieh Lin, Fu-Yin Cherng, Hao-Chuan Wang, Ching-Ying Sung, and Jung-Tai King. Using time-anchored peer comments to enhance social interaction in online educational videos. In *CHI '15*, pages 689–698, 2015.

[160] Manuel León, Ruth Cobos, Kate Dickens, Su White, and Hugh Davis. Visualising the MOOC experience: a dynamic MOOC dashboard built through institutional collaboration. In *EMOOCs '16*, pages 461–470, 2016.

[161] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10(8):707–710, 1966.

[162] John M Levine. Social comparison and education. *ERIC*, pages 3–29, 1983.

[163] Wei Li, Tovi Grossman, and George Fitzmaurice. CADament: a gamified multiplayer software tutorial system. In *CHI '14*, pages 3369–3378, 2014.

[164] Jens O Liegle and Thomas N Janicki. The effect of learning styles on the navigation needs of Web-based learners. *Computers in Human Behavior*, 22(5):885–898, 2006.

[165] Anthony M Limperos, Marjorie M Buckner, Renee Kaufmann, and Brandi N Frisby. Online teaching and technological affordances: An experimental investigation into the impact of modality and clarity on perceived and actual learning. *Computers & Education*, 83:1–9, April 2015.

[166] Robert V Lindsey, Jeffery D Shroyer, Harold Pashler, and Michael C Mozer. Improving students long-term knowledge retention through personalized review. *Psychological science*, 25(3):639–647, 2014.

[167] Allison Littlejohn, Nina Hood, Colin Milligan, and Paige Mustain. Learning in moocs: Motivations and self-regulated learning in moocs. *The Internet and Higher Education*, 29:40–48, 2016.

[168] Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. Understanding of internal clustering validation measures. In *Data Min-*

ing (ICDM), 2010 IEEE 10th International Conference on, pages 911–916. IEEE, 2010.

[169] Zhongxiu Liu, Rebecca Brown, Collin F Lynch, Tiffany Barnes, Ryan Baker, Yoav Bergner, and Danielle McNamara. Mooc learner behaviors by country and culture; an exploratory analysis. In *EDM '16*, pages 127–134, 2016.

[170] Penelope Lockwood, Dan Dolderman, Pamela Sadler, and Elinora Gerchak. Feeling better about doing worse: social comparisons within romantic relationships. *Journal of personality and social psychology*, 87 (1):80, 2004.

[171] Penelope Lockwood, Tara C Marshall, and Pamela Sadler. Promoting success or preventing failure: Cultural differences in motivation by positive and negative role models. *Personality and Social Psychology Bulletin*, 31(3):379–392, 2005.

[172] Jaclyn K Maass and Philip I Pavlik Jr. Modeling the influence of format and depth during effortful retrieval practice. In *EDM '16*, pages 143–150, 2016.

[173] Michael J Mahoney, Bert S Moore, Terry C Wade, and Nanci G Moura. Effects of Continuous and Intermittent Self-Monitoring on Academic Behavior. *Journal of Consulting and Clinical Psychology*, 41(1):65, 1973.

[174] Roberto Martinez Maldonado, Judy Kay, Kalina Yacef, and Beat Schwendimann. An interactive teachers dashboard for monitoring groups in a multi-tabletop learning environment. In *ITS '12*, pages 482–492, 2012.

[175] Jorge Maldonado-Mahauad, Mar Pérez-Sanagustín, René F Kizilcec, Nicolás Morales, and Jorge Munoz-Gama. Mining theory-based patterns from big data. *Computers in Human Behavior*, 2017.

[176] Anoush Margaryan, Manuela Bianco, and Allison Littlejohn. Instructional Quality of Massive Open Online Courses (MOOCs). *Computers & Education*, 80:77–83, 2015.

[177] Lauren E. Margulieux and Richard Catrambone. Improving problem solving performance in computer-based learning environments through subgoal labels. In *L@S '14*, pages 149–150, 2014.

[178] Hazel R Markus and Shinobu Kitayama. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, 98(2):224, 1991.

[179] Florence Martin and Jeffrey Ertzberger. Here and now mobile learning: An experimental study on the use of mobile technology. *Computers & Education*, 68:76–85, October 2013.

[180] Alexander McAuley, Bonnie Stewart, George Siemens, and Dave Cormier. The mooc model for digital practice. *Knowledge Synthesis Grants on the Digital Economy*, 2010.

[181] Antonija Mitrovic, Stellan Ohlsson, and Devon K Barrow. The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education*, 60(1):264–272, January 2013.

[182] Yohsuke R Miyamoto, Cody A Coleman, Joseph Jay Williams, Jacob Whitehill, Sergiy Nesterko, and Justin Reich. Beyond time-on-task: The relationship between spaced study and certification in MOOCs. *Journal of Learning Analytics*, 2(2):47–69, December 2015.

[183] Toni-Jan Keith Palma Monserrat, Shengdong Zhao, Kevin McGee, and Anshul Vikram Pandey. NoteVideo: facilitating navigation of blackboard-style lecture videos. In *CHI '13*, pages 1139–1148, 2013.

[184] Toni-Jan Keith Palma Monserrat, Yawen Li, Shengdong Zhao, and Xiang Cao. L. IVE: an integrated interactive video-based learning environment. In *CHI '14*, pages 3399–3402, 2014.

[185] Yishay Mor and Niall Winters. Design approaches in technology-enhanced learning. *Interactive Learning Environments*, 15(1):61–75, 2007.

[186] Behrooz Mostafavi, Michael Eagle, and Tiffany Barnes. Towards data-driven mastery learning. In *LAK '15*, pages 270–274, 2015.

[187] Patrick Mukala, JCAM Buijs, and WMP Van Der Aalst. Exploring students' learning behaviour in moocs using process mining techniques, 2015.

[188] Chang Woo Nam and Ronald D Zellner. The relative effects of positive interdependence and group processing on student achievement and attitude in online cooperative learning. *Computers & Education*, 56(3): 680–688, April 2011.

[189] Ilona Nawrot and Antoine Doucet. Building Engagement for MOOC Students: Introducing Support for Time Management on Online Learning Platforms. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 1077–1082. International World Wide Web Conferences Steering Committee, 2014.

[190] Steve Nebel, Sascha Schneider, Maik Beege, and Günter Daniel Rey. Leaderboards within educational videogames: The impact of difficulty, effort and gameplay. *Computers & Education*, 113:28–41, October 2017.

[191] Tricia Ngoon, Alexander Gamero-Garrido, and Scott Klemmer. Supporting Peer Instruction with Evidence-Based Online Instructional Templates. In *L@S '16*, pages 301–304, 2016.

[192] Quan Nguyen, Michal Huptych, and Bart Rienties. Linking students' timing of engagement to learning design and academic performance. In *LAK '17*, pages 141–150, 2018.

[193] Thi Thao Duyen T Nguyen, Thomas Garncarz, Felicia Ng, Laura A Dabbish, and Steven P Dow. Fruitful Feedback. In *the 2017 ACM Conference*, pages 1024–1034, New York, New York, USA, 2017. ACM Press.

[194] James Nicholson, Mark Huber, Daniel Jackson, and Patrick Olivier. Panopticon as an elearning support search tool. In *CHI '14*, pages 1221–1224, 2014.

[195] Omid Noroozi, Armin Weinberger, Harm JA Biemans, Stephanie D. Teasley, and Martin Mulder. Fostering multidisciplinary learning through computer-supported collaboration script: The role of a transactive memory script. In *EC-TEL '12*, pages 413–418, 2012.

[196] Laura Nota, Salvatore Soresi, and Barry J Zimmerman. Self-Regulation and Academic Achievement and Resilience: A Longitudinal Study. *International Journal of Educational Research*, 41(3):198–215, 2004.

[197] Daniel FO Onah, Jane Sinclair, and Russell Boyatt. Dropout rates of massive open online courses: behavioural patterns. *EDULEARN14*, pages 5825–5834, 2014.

[198] Vineet Pandey, Amnon Amir, Justine Debelius, Embriette R Hyde, Tomasz Kosciolek, Rob Knight, and Scott Klemmer. *Gut Instinct: Creating Scientific Theories with Online Learners.* Creating Scientific

Theories with Online Learners. ACM, New York, New York, USA, May 2017.

[199] Pantelis M Papadopoulos, Thomas D Lagkas, and Stavros N Demetriadis. How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course. *Computers & Education*, 59(2):182–195, September 2012.

[200] Kyparisia A Papanikolaou. Constructing interpretative views of learners interaction behavior in an open learner model. *IEEE Transactions on Learning Technologies*, 8(2):201–214, 2015.

[201] Abelardo Pardo, Jelena Jovanovic, Shane Dawson, Dragan Gašević, and Negin Mirriahi. Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 2017.

[202] Zachary A Pardos, Steven Tang, Daniel Davis, and Christopher Vu Le. Enabling real-time adaptivity in moocs with a personalized next-step recommendation framework. In *L@S '17*, pages 23–32. ACM, 2017.

[203] Jihyun Park, Kameryn Denaro, Fernando Rodriguez, Padhraic Smyth, and Mark Warschauer. Detecting changes in student behavior from clickstream data. In *LAK '17*, pages 21–30. ACM, 2017.

[204] Yeonjeong Park and Il-Hyun Jo. Development of the learning analytics dashboard to support students' learning performance. *J. UCS*, 21(1): 110–133, 2015.

[205] Harold Pashler and Eric-Jan Wagenmakers. Editors introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6):528–530, 2012.

[206] Ray Pastore. The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education*, 58(1):641–651, January 2012.

[207] Jackie M Poos, Karel van den Bosch, and Christian P Janssen. Battling bias: Effects of training and training context. *Computers & Education*, 111:101–113, August 2017.

[208] Rafael Pozo-Barajas, María del Pópulo Pablo-Romero, and Rafael Caballero. Evaluating a computer-based simulator program to teach the principles of macroeconomic equilibria. *Computers & Education*, 69: 71–84, November 2013.

[209] Anna N Rafferty, Rachel A Jansen, and Thomas L Griffiths. Using inverse planning for personalized feedback. In *EDM '16*, pages 472–477, 2016.

[210] Mitchel Resnick. Rethinking learning in the digital age, 2002.

[211] Blake A Richards and Paul W Frankland. The Persistence and Transience of Memory. *Neuron*, 94(6):1071–1084, 2017.

[212] Henry L Roediger and Andrew C Butler. The Critical Role of Retrieval Practice in Long-Term Retention. *Trends in cognitive sciences*, 15(1): 20–27, 2011.

[213] Todd Rogers and Avi Feller. Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting. *Psychological Science*, 27(3):365–374, 2016.

[214] Doug Rohrer and Kelli Taylor. The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20(9):1209–1224, 2006.

[215] Yigal Rosen, Ilia Rushkin, Andrew Ang, Colin Federicks, Dustin Tingley, and Mary Jean Blink. Designing Adaptive Assessments in MOOCs. In *the Fourth (2017) ACM Conference*, pages 233–236, New York, New York, USA, 2017. ACM Press.

[216] Marc J Rosenberg and Rob Foshay. E-learning: Strategies for delivering knowledge in the digital age. *Performance Improvement*, 41(5):50–51, 2002.

[217] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[218] Anne Rozinat and Wil MP van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33 (1):64–95, 2008.

[219] M Joan Saary. Radar plots: a useful way for presenting multivariate health care data. *Journal of clinical epidemiology*, 61(4):311–317, 2008.

[220] Gerald Sagotsky, Charlotte J Patterson, and Mark R Lepper. Training Children's Self-Control: A Field Experiment in Self-Monitoring and Goal-Setting in the Classroom. *Journal of Experimental Child Psychology*, 25(2):242–253, 1978.

[221] Michaéla C Schippers, Ad WA Scheepers, and Jordan B Peterson. A Scalable Goal-Setting Intervention Closes Both the Gender and Ethnic Minority Achievement Gap. *Palgrave Communications*, 1, 2015.

[222] Bertrand Schneider, Patrick Jermann, Guillaume Zufferey, and Pierre Dillenbourg. Benefits of a tangible interface for collaborative learning and interaction. *IEEE TLT*, 4(3):222–232, 2011.

[223] Beat A Schwendimann, María Jesús Rodríguez-Triana, Andrii Vozniuk, Luis P Prieto, Mina Shirvani Boroujeni, Adrian Holzer, Denis Gillet, and Pierre Dillenbourg. Understanding learning at a glance: An overview of learning dashboard studies. In *LAK '16*, pages 532–533, 2016.

[224] Gayane Sedrakyan, Monique Snoeck, and Stephan Poelmans. Assessing the effectiveness of feedback enabled simulation in teaching conceptual modeling. *Computers & Education*, 78:367–382, September 2014.

[225] Serkan Şendağ and H Ferhan Odabaşı. Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, 53(1):132–141, August 2009.

[226] David H Shanabrook, David G Cooper, Beverly Park Woolf, and Ivon Arroyo. Identifying high-level student behavior using sequence-based motif discovery. In *EDM*, pages 191–200, 2010.

[227] Kshitij Sharma, Patrick Jermann, and Pierre Dillenbourg. Displaying Teachers Gaze in a MOOC: Effects on Students Video Navigation Patterns. In *EC-TEL '15*, pages 325–338. Springer, 2015.

[228] Lorraine Sherry. Issues in distance learning. *International journal of educational telecommunications*, 1(4):337–365, 1995.

[229] Lee S Shulman. Signature pedagogies in the professions. *Daedalus*, 134 (3):52–59, 2005.

[230] Melody Siadaty, Dragan Gašević, and Marek Hatala. Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning. *Computers in Human Behavior*, 59:469–482, 2016.

[231] Melody Siadaty, Dragan Gašević, and Marek Hatala. Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics*, 3(1):183–214, 2016.

[232] George Siemens. Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, 2(1):3–10, 2005.

[233] George Siemens. Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10):1380–1400, 2013.

[234] Tanmay Sinha, Patrick Jermann, Nan Li, and Pierre Dillenbourg. Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. In *2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*, pages 1–12, 2014.

[235] Glenn Gordon Smith, Mimi Li, Jack Drobisz, Ho-Ryong Park, Deoksoon Kim, and Stanley Dana Smith. Play games or study? Computer games in eBooks to learn English vocabulary. *Computers & Education*, 69:274–286, November 2013.

[236] Mario Soflano, Thomas M Connolly, and Thomas Hainey. An application of adaptive games-based learning based on learning style to teach SQL. *Computers & Education*, 86:192–211, August 2015.

[237] Hyuksoon S Song, Martin Pusic, Michael W Nick, Umut Sarpel, Jan L Plass, and Adina L Kalet. The cognitive impact of interactive design features for learning complex materials in medical education. *Computers & Education*, 71:198–205, February 2014.

[238] Christoph Sonnenberg and Maria Bannert. Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, 2(1): 72–100, 2015.

[239] Sarah D. Sparks. Data Mining' Gains Traction in Education. *Education Week*, 2010.

[240] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explor. Newsl.*, pages 12–23, 2000.

[241] David J Stanley and Jeffrey R Spence. Expectations for replications: Are yours realistic? *Perspectives on Psychological Science*, 9(3):305–318, 2014.

[242] Matthew Streeter. Mixture modeling of individual learning curves. *International Educational Data Mining Society*, 2015.

[243] Karen Swan, Scott Day, Leonard Bogle, and Traci van Prooyen. Amp: A tool for characterizing the pedagogical approaches of moocs. *e-mentor*, 2:75–85, 2014.

[244] Bernardo Tabuenca, Marco Kalz, Stefaan Ternier, and Marcus Specht. Stop and think: Exploring mobile notifications to foster reflective practice on meta-learning. *IEEE TLT*, 8(1):124–135, 2015.

[245] Candace Thille, Emily Schneider, René F Kizilcec, Christopher Piech, Sherif A Halawa, and Daniel K Greene. The future of data-enriched assessment. *Research & Practice in Assessment*, 9:5–16, 2014.

[246] Meredith M. Thompson, Eric Braude, Christopher D. Canfield, Jay Halfond, and Aparaita Sengupta. Assessment of KNOWLA: Knowledge Assembly for Learning and Assessment. In *L@S '15*, pages 267–271, 2015.

[247] Jonathan H. Tomkin and Donna Charlevoix. Do professors matter?: Using an a/b test to evaluate the impact of instructor involvement on MOOC student outcomes. In *L@S '14*, pages 71–78, 2014.

[248] Jared Tritz, Nicole Michelotti, Ginger Shultz, Tim McKay, and Barsaa Mohapatra. Peer evaluation of student generated content. In *LAK '14*, pages 277–278, 2014.

[249] Andrew Trusty and Khai N. Truong. Augmenting the web for second language vocabulary learning. In *CHI '11*, pages 3179–3188, 2011.

[250] Selen Türkay. The effects of whiteboard animations on retention and subjective experiences when learning advanced physics topics. *Computers & Education*, 98:102–114, July 2016.

[251] Jaime Urquiza-Fuentes and J Ángel Velázquez. Toward the effective use of educational program animations: The roles of student's engagement and topic complexity. *Computers & Education*, 67:178–192, September 2013.

[252] Fabienne M van der Kleij, Theo J H M Eggen, Caroline F Timmers, and Bernard P Veldkamp. Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1):263–272, January 2012.

[253] Tamara van Gog. Effects of identical example–problem and problem–example pairs on learning. *Computers & Education*, 57(2):1775–1779, September 2011.

[254] Tamara van Gog, Ilse Verveer, and Lise Verveer. Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, 72:323–327, March 2014.

[255] Margot van Wermeskerken and Tamara van Gog. Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education*, 113:98–107, October 2017.

[256] Stephen Vassallo. Implications of Institutionalizing Self-Regulated Learning: An Analysis from Four Sociological Perspectives. *Educational Studies*, 47(1):26–49, 2011.

[257] Katrien Verbert, Erik Duval, Joris Klerkx, Sten Govaerts, and José Luis Santos. Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10):1500–1509, 2013.

[258] Pei-Yu Wang, Brandon K Vaughn, and Min Liu. The impact of animation interactivity on novices' learning of introductory statistics. *Computers & Education*, 56(1):300–311, January 2011.

[259] Michael L Wehmeyer, Susan B Palmer, Martin Agran, Dennis E Mithaug, and James E Martin. Promoting Causal Agency: The Self-Determined Learning Model of Instruction. *Exceptional Children*, 66 (4):439–453, 2000.

[260] Miaomiao Wen and Carolyn Penstein Rosé. Identifying latent study habits by mining learner behavior patterns in massive open online courses. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1983–1986. ACM, 2014.

[261] Eliane S Wiese, Michael Yen, Antares Chen, Lucas A Santos, and Armando Fox. Teaching Students to Recognize and Implement Good Coding Style. *L@S*, pages 41–50, 2017.

[262] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z. Gajos, Walter S. Lasecki, and Neil Heffernan. AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning. In *L@S '16*, pages 379–388, 2016.

[263] Joseph Jay Williams, Tania Lombrozo, Anne Hsu, Bernd Huber, and Juho Kim. Revising learner misconceptions without feedback: Prompting for reflection on anomalies. In *CHI '16*, pages 470–474. ACM, 2016.

[264] Philip H Winne. Learning analytics for self-regulated learning. *Handbook of Learning Analytics*, page 241, 2017.

[265] Fielding I Winters, Jeffrey A Greene, and Claudine M Costich. Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20(4):429–444, 2008.

[266] Gabriele Wulf and Charles H Shea. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic bulletin & review*, 9(2):185–211, 2002.

[267] Xiaolu Xiong, Yan Wang, and Joseph Barbosa Beck. Improving students' long-term retention performance: a study on personalized retention schedules. In *LAK '15*, pages 325–329, 2015.

[268] Masanori Yamada. The role of social presence in learner-centered communicative language learning using synchronous computer-mediated communication: Experimental study. *Computers & Education*, 52(4): 820–833, May 2009.

[269] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rose. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *2013 NIPS Data-driven education workshop*, volume 11, 2013.

[270] Tzu-Chi Yang, Sherry Y Chen, and Gwo-Jen Hwang. The influences of a two-tier test strategy on student learning: A lag sequential analysis approach. *Computers & Education*, 82:366–377, March 2015.

[271] Tzu Chi Yang, Dai Ling Shih, and Meng Chang Chen. An Investigation of the Effects of Online Test Strategy on Students' Learning Behaviors. In *L@S '16*, pages 281–284, 2016.

[272] Michael Yeomans and Justin Reich. Planning prompts increase and forecast course completion in massive open online courses. In *LAK '17*, pages 464–473, 2017.

[273] Yue Zhao, Dan Davis, Guanliang Chen, Christoph Lofi, Claudia Hauff, and Geert-Jan Houben. Certificate achievement unlocked: How does mooc learners' behaviour change? In *Adjunct Publication of the 25th*

Conference on User Modeling, Adaptation and Personalization, pages 83–88. ACM, 2017.

[274] Saijing Zheng, Mary Beth Rosson, Patrick C Shih, and John M Carroll. Understanding student motivation, behaviors and perceptions in moocs. In *CSCW '15*, pages 1882–1895, 2015.

[275] Zhilin Zheng, Tim Vogelsang, and Niels Pinkwart. The Impact of Small Learning Group Composition on Student Engagement and Success in a MOOC. In *EDM '14*, pages 500–503, 2014.

[276] Guojing Zhou, Jianxun Wang, Collin F Lynch, and Min Chi. Towards Closing the Loop: Bridging Machine-induced Pedagogical Policies to Learning Theories. In *the 10th International Conference on Educational Data Mining*, pages 112–119, 2017.

[277] Yeshuang Zhu, Yuntao Wang, Chun Yu, Shaoyun Shi, Yankai Zhang, Shuang He, Peijun Zhao, Xiaojuan Ma, and Yuanchun Shi. *ViVo: Video-Augmented Dictionary for Vocabulary Learning.* Video-Augmented Dictionary for Vocabulary Learning. ACM, New York, New York, USA, May 2017.

[278] Barry J Zimmerman. A Social Cognitive View of Self-Regulated Academic Learning. *Journal of educational psychology*, 81(3):329, 1989.

[279] Barry J Zimmerman. Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2):64–70, 2002.

[280] Barry J Zimmerman, Albert Bandura, and Manuel Martinez-Pons. Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting. *American educational research journal*, 29(3):663–676, 1992.

List of Figures

List of Tables

Origins of Chapters

8.3 Origin of Chapters

All main chapters in this thesis are based on existing peer-reviewed, published articles in journals or conferences related to the research domain of this work.

- **Part I**: Teaching & Learning Paths
  - **Chapter 2** is based on the article published at the 9th International Conference on Educational Data Mining (EDM 2016) [63],
  - **Chapter 3** is based on the article published at the 5th annual ACM Conference on Learning at Scale (L@S 2018) [69]
- **Part II**: Improving Learning & Teaching Behavior
  - **Chapter 4** is based on the article published in the journal Computers & Education, Volume 125 in 2018 [67].
- **Part III**: Study Planning
  - **Chapter 5** is based on the article published at the 11th European Conference on Technology-Enhanced Learning (EC-TEL 2016) [65],
  - **Chapter 6** is based on the article published at the 7th International Conference on Learning Analytics and Knowledge (LAK 2017) [66],
  - **Chapter 7** is based on the article published at the 13th European Conference on Technology-Enhanced Learning (EC-TEL 2018) [70]
- **Part IV**: Retrieval Practice

- **Chapter 8** is based on the article published at the 11th European Conference on Technology-Enhanced Learning (EC-TEL 2016) [65],
- **Chapter 9** is based on the article published at the 8th International Conference on Learning Analytics and Knowledge (LAK 2018) [68]

Curriculum Vitae

Publications

1. Davis, D., Chen, G., Hauff, C., Houben, G.J. (2018) Activating Learning at Scale: A Review of Innovations in Online Learning Strategies. Computers & Education. Volume 125, Pages 327-344.

2. Davis, D., Kizilcec, R.F., Hauff, C., Houben, G.J. (2018) Scaling Effective Learning Strategies: Retrieval Practice and Long-Term Knowledge Retention in MOOCs. In Journal of Learning Analytics. In Press.

3. Davis, D., Hauff, C., Houben, G.J. (2018) Evaluating Crowdworkers as a Proxy for Online Learners in Video-Based Learning Contexts. In Proceedings of the 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing, CSCW '18.

4. Davis, D., Triglianos, V., Hauff, C., Houben, G.J (2018) SRLx: An Personalized Learner Interface for MOOCs. In Proceedings of the 13th European Conference on Technology-Enhanced Learning, EC-TEL '18.

5. Davis, D., Seaton, D., Hauff, C., Houben, G.J. (2018) Toward Large-Scale Learning Design: Categorizing Course Designs in Service of Supporting Learning Outcomes. In Proceedings of the Fifth Annual ACM Conference on Learning at Scale, L@S '18.

6. Davis, D., Kizilcec, R.F., Hauff, C., Houben, G.J. (2018) The Half-Life of MOOC Knowledge: A Randomized Trial Evaluating the Testing Effect in MOOCs. In Proceedings of 8th International Conference on Learning Analytics and Knowledge, LAK '18.

7. Van der Zee, T., Davis, D., Saab, N., Giesbers, B., Ginn, J., van der Sluis, F., Paas, F., Admiraal, W. (2018) Evaluating Retrieval Practice in a MOOC: How writing and reading summaries of videos affects

student learning. In Proceedings of 8th International Conference on Learning Analytics and Knowledge, LAK ’18.

8. Bodily, R., Kay, J., Aleven, V., Davis, D., Jivet, I., Xhakaj, F., Verbert, K. (2018) Open learner models and learning analytics dashboards: A systematic review. In Proceedings of 8th International Conference on Learning Analytics and Knowledge, LAK ’18.
9. Khalil, M., Davis, D., Wong, J. (2018) Gamification in MOOCs: A General Overview. In Open Education Global Conference, OE Global ’18.
10. Seaton, D., Davis, D. (2018) Toward Large-Scale Learning Design: ăCategorizing Course Designs in Service of Supporting Learning Outcomes. In HybridEd Workshop. 2018
11. Zhao, Y., Davis, D., Chen, G., Lofi, C., Hauff, C., Houben, G.J. (2017) Certificate Achievement Unlocked: Exploring MOOC Learners Behaviour Before & After Passing. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP ’17
12. Pardos, Z., Tang, S.,ăDavis, D.,ăLe, C.V. (2017) Enabling Real-Time Adaptivity in MOOCs with a Personalized Next-Step Recommendation Framework. In Proceedings of the Fourth Annual ACM Conference on Learning at Scale, L@S ’17.
13. Davis, D., Jivet, I., Kizilcec, R.F., Chen, G., Hauff, C., Houben, G.J. (2017) Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale. In Proceedings of 7th International Conference on Learning Analytics and Knowledge, LAK ’17.
14. Chen, G., Davis, D., Krause, M., Aivaloglou, E., Hauff, C., Houben, G.J. (2017) Buying Time: Enabling Learners to become Earners with a Real-World Paid Task Recommender System. In Proceedings of 7th International Conference on Learning Analytics and Knowledge, LAK ’17.
15. Chen, G., Davis, D., Krause, M., Aivaloglou, E., Hauff, C., Houben, G.J. (2016) Can Learners be Earners? Investigating a Design to Enable MOOC learners to Apply their Skills and Earn Money in an Online Market Place. In IEEE Transactions on Learning Technologies (TLT).
16. Davis, D., Chen, G., van der Zee, T., Hauff, C., Houben, G.J. (2016) Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-Regulated Learning Strategies at Scale. InăProceedings of

the 11th European Conference on Technology-Enhanced Learning, EC-TEL '16. *Best Student Paper Award

17. Davis, D., Chen, G., Hauff, C., Houben, G.J. (2016) Gauging MOOC Learners Adherence to the Designed Learning Path. In Proceedings of the Ninth International Conference on Educational Data Mining, EDM '16.

18. Chen, G., Davis, D., Hauff, C., Houben, G.J. (2016) On the Impact of Personality in Massive Open Online Learning. In Proceedings of the 24th Conference on User Modeling, Adaptation, and Personalization, UMAP '16.

19. Chen, G., Davis, D., Hauff, C., Houben, G.J. (2016) MOOC Learners in Social Web Platforms. In Proceedings of the Eighth International ACM Web Science Conference, WebSci '16.

20. Davis, D., Chen, G., Jivet, I., Hauff, C., Houben, G.J. (2016) Encouraging Metacognition & Self-Regulation in MOOCs through Increased Learner Feedback. In the Sixth International Learning Analytics & Knowledge Conference, LAK '16, Learning Analytics for Learners Workshop.

21. Chen, G., Davis, D., Hauff, C., Houben, G.J. (2016) Learning Transfer: Does it take place in MOOCS?. In Proceedings of the Third Annual ACM Conference on Learning at Scale, L@S '16. *Best Paper Honorable Mention

22. van Staalduinen, J.P., Davis, D., Topolovec, S. (2016) Opening University Education to the World and Improving Education: Using MOOC-based Research as a Tool for Innovation. In Proceedings of the 9th annual International Conference of Education, Research and Innovation, ICERI 2016.

23. Davis, D., Hanacek, J., Myers, A., Mulroney, S., Vovides, Y. (2015) Capturing, Tracing, and Visualizing the Spread of Technology-Enhanced Instructional Strategies. In Proceedings of the Seventh International Conference on Education and New Learning Technologies, EDULEARN '15.

24. Vovides, Y., Youmans, T., Arthur, P., Davis, D., Ayo, E., Pongsajapan, R., McWilliams, M., Kruse, A., (2015) Examining Learners Cognitive Presence in MOOCs. In the Fifth International Learning Analytics

& Knowledge Conference, LAK '15, Learning Analytics Review Workshop.

25. Vovides, Y., Davis, D., Kruse, A., Pongsajapan, R., Youmans, T., Maloney, E. J. (2014) GeorgetownX: Applying an Integrative Learning Analytics Design Approach. In EDUCAUSE Annual Conference, EDUCAUSE '14.

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D. students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series.

2019-14 Daniel Davis (TUD) *Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses*
2019-13 Guanliang Chen (TUD) *MOOC Analytics: Learner Modeling and Content Generation*
2019-12 Jacqueline Heinerman (VU) *Better Together*
2019-11 Yue Zhao (TUD) *Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs*
2019-10 Qing Chuan Ye (EUR) *Multi-objective Optimization Methods for Allocation and Prediction*
2019-09 Fahimeh Alizadeh Moghaddam (UVA) *Self-adaptation for energy efficiency in software systems*
2019-08 Frits de Nijs (TUD) *Resource-constrained Multi-agent Markov Decision Processes*
2019-06 Chris Dijkshoorn (VU) *Nichesourcing for Improving Access to Linked Cultural Heritage Datasets*
2019-05 Sebastiaan van Zelst (TUE) *Process Mining with Streaming Data*
2019-04 Ridho Rahmadi (RUN) *Finding stable causal structures from clinical data*
2019-03 Eduardo Gonzalez Lopez de Murillas (TUE) *Process Mining on Databases: Extracting Event Data from Real Life Data Sources*
2019-02 Emmanuelle Beauxis- Aussalet (CWI, UU) *Statistics and Visualizations for Assessing Class Size Uncertainty*
2019-01 Rob van Eijk (UL) *Comparing and Aligning Process Representations*
2018-30 Wouter Beek (VU) *The K in semantic web stands for knowledge: scaling semantics to the web*
2018-29 Yu Gu (UVT) *Emotion Recognition from Mandarin Speech*
2018-28 Christian Willemse (UT) *Social Touch Technologies: How they feel and how they make you feel*
2018-27 Maikel Leemans (TUE) *Hierarchical Process Mining for Scalable Software Analysis*
2018-26 Roelof de Vries (UT) *Theory-Based And Tailor-Made: Motivational Messages for Behavior Change Technology*
2018-25 Riste Gligorov (VUA) *Serious Games in Audio-Visual Collections*
2018-24 Jered Vroon (UT) *Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots*
2018-23 Kim Schouten (EUR) *Semantics-driven Aspect-Based Sentiment Analysis*
2018-22 Eric Fernandes de Mello Araújo (VUA) *Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks*
2018-21 Aad Slootmaker (OU) *EMERGO: a generic platform for authoring and playing scenario-based serious games*
2018-20 Manxia Liu (RUN) *Time and Bayesian Networks*
2018-19 Minh Duc Pham (VUA) *Emergent relational schemas for RDF*
2018-18 Henriette Nakad (UL) *De Notaris en Private Rechtspraak*
2018-17 Jianpeng Zhang (TUE) *On Graph Sample Clustering*
2018-16 Jaebok Kim (UT) *Automatic recognition of engagement and emotion in a group of children*
2018-15 Naser Davarzani (UM) *Biomarker discovery in heart failure*
2018-14 Bart Joosten (UVT) *Detecting Social Signals with Spatiotemporal Gabor Filters*
2018-13 Seyed Amin Tabatabaei (VUA) *Computing a Sustainable Future: Exploring the added value of computational models for increasing* the use of renewable energy in the residential sector
2018-12 Xixi Lu (TUE) *Using behavioral context in process mining*
2018-11 Mahdi Sargolzaei (UVA) *Enabling Framework for Service-oriented Collaborative Networks*
2018-10 Julienka Mollee (VUA) *Moving forward: supporting physical activity behavior change through intelligent technology*
2018-09 Xu Xie (TUD) *Data Assimilation in Discrete Event Simulations*
2018-08 Rick Smetsers (RUN) *Advances in Model Learning for Software Systems*
2018-07 Jieting Luo (UU) *A formal account of opportunism in multi-agent systems*
2018-06 Dan Ionita (UT) *Model-Driven Information Security Risk Assessment of Socio-Technical Systems*
2018-05 Hugo Huurdeman (UVA) *Supporting the Complex Dynamics of the Information Seeking Process*
2018-04 Jordan Janeiro (TUD) *Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks*
2018-03 Steven Bosems (UT) *Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction*
2018-02 Felix Mannhardt (TUE) *Multi-perspective Process Mining*
2018-01 Han van der Aa (VU) *Comparing and Aligning Process Representations*
2017-47 Jie Yang (TUD), *Crowd Knowledge Creation Acceleration*
2017-46 Jan Schneider (OU), *Sensor-based Learning Support*

2017-45 Bas Testerink (UU), *Decentralized Runtime Norm Enforcement*
2017-44 Garm Lucassen (UU), *Understanding User Stories - Computational Linguistics in Agile Requirements Engineering*
2017-43 Maaike de Boer (RUN), *Semantic Mapping in Video Retrieval*
2017-42 Elena Sokolova (RUN), *Causal discovery from mixed and missing data with applications on ADHD datasets*
2017-41 Adnan Manzoor (VUA), *Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle*
2017-40 Altaf Hussain Abro (VUA), *Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems"*
2017-39 Sara Ahmadi (RUN), *Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR*
2017-38 Alex Kayal (TUD), *Normative Social Applications*
2017-37 Alejandro Montes Garca (TUE), *WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy*
2017-36 Yuanhao Guo (UL), *Shape Analysis for Phenotype Characterisation from High-throughput Imaging*
2017-35 Martine de Vos (VU), *Interpreting natural science spreadsheets*
2017-34 Maren Scheffel (OUN), *The Evaluation Framework for Learning Analytics*
2017-33 Brigit van Loggem (OU), *Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity*
2017-32 Thaer Samar (RUN), *Access to and Retrievability of Content in Web Archives*
2017-31 Ben Ruijl (UL), *Advances in computational methods for QFT calculations*
2017-30 Wilma Latuny (UVT), *The Power of Facial Expressions*
2017-29 Adel Alhuraibi (UVT), *From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT*
2017-28 John Klein (VU), *Architecture Practices for Complex Contexts*
2017-27 Michiel Joosse (UT), *Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors*
2017-26 Merel Jung (UT), *Socially intelligent robots that understand and respond to human touch*
2017-25 Veruska Zamborlini (VU), *Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search*
2017-24 Chang Wang (TUD), *Use of Affordances for Efficient Robot Learning*
2017-23 David Graus (UVA), *Entities of Interest— Discovery in Digital Traces*
2017-22 Sara Magliacane (VU), *Logics for causal inference under uncertainty*
2017-21 Jeroen Linssen (UT), *Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)*
2017-20 Mohammadbashir Sedighi (TUD), *Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility*
2017-19 Jeroen Vuurens (TUD), *Proximity of Terms, Texts and Semantic Vectors in Information Retrieval*
2017-18 Ridho Reinanda (UVA), *Entity Associations for Search*
2017-17 Daniel Dimov (UL), *Crowdsourced Online Dispute Resolution*
2017-16 Aleksandr Chuklin (UVA), *Understanding and Modeling Users of Modern Search Engines*
2017-15 Peter Berck, Radboud University (RUN), *Memory-Based Text Correction*
2017-14 Shoshannah Tekofsky (UvT), *You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior*
2017-13 Gijs Huisman (UT), *Social Touch Technology - Extending the reach of social touch through haptic technology*
2017-12 Sander Leemans (TUE), *Robust Process Mining with Guarantees*
2017-11 Florian Kunneman (RUN), *Modelling patterns of time and emotion in Twitter #anticipointment*
2017-10 Robby van Delden (UT), *(Steering) Interactive Play Behavior*
2017-09 Dong Nguyen (UT), *Text as Social and Cultural Data: A Computational Perspective on Variation in Text*
2017-08 Rob Konijn (VU), *Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery*
2017-07 Roel Bertens (UU), *Insight in Information: from Abstract to Anomaly*
2017-06 Damir Vandic (EUR), *Intelligent Information Systems for Web Product Search*
2017-05 Mahdieh Shadi (UVA), *Collaboration Behavior*
2017-04 Mrunal Gawade (CWI), *MULTI-CORE PARALLELISM IN A COLUMN-STORE*
2017-03 Daniël Harold Telgen (UU), *Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines*
2017-02 Sjoerd Timmer (UU), *Designing and Understanding Forensic Bayesian Networks using Argumentation*
2017-01 Jan-Jaap Oerlemans (UL), *Investigating Cybercrime*
2016-50 Yan Wang (UVT), *The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains*
2016-49 Gleb Polevoy (TUD), *Participation and Interaction in Projects. A Game-Theoretic Analysis*
2016-48 Tanja Buttler (TUD), *Collecting Lessons Learned*
2016-47 Christina Weber (UL), *Real-time foresight - Preparedness for dynamic innovation networks*
2016-46 Jorge Gallego Perez (UT), *Robots to Make you Happy*
2016-45 Bram van de Laar (UT), *Experiencing Brain-Computer Interface Control*
2016-44 Thibault Sellam (UVA), *Automatic Assistants for Database Exploration*
2016-43 Saskia Koldijk (RUN), *Context-Aware Support for Stress Self-Management: From Theory to Practice*
2016-42 Spyros Martzoukos (UVA), *Combinatorial and Compositional Aspects of Bilingual Aligned Corpora*
2016-41 Thomas King (TUD), *Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance*
2016-40 Christian Detweiler (TUD), *Accounting for Values in Design*
2016-39 Merijn Bruijnes (UT), *Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect*
2016-38 Andrea Minuto (UT), *MATERIALS THAT MATTER - Smart Materials meet Art & Interaction Design*
2016-37 Giovanni Sileno (UvA), *Aligning Law and Action - a conceptual and computational inquiry*
2016-36 Daphne Karreman (UT), *Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies*
2016-35 Zhaochun Ren (UVA), *Monitoring Social Media: Summarization, Classification and Recommendation*
2016-34 Dennis Schunselaar (TUE), *Configurable Process Trees: Elicitation, Analysis, and Enactment*
2016-33 Peter Bloem (UVA), *Single Sample Statistics, exercises in learning from just one example*
2016-32 Eelco Vriezekolk (UT), *Assessing Telecommunication Service Availability Risks for Crisis Organisations*
2016-31 Mohammad Khelghati (UT), *Deep web content monitoring*
2016-30 Ruud Mattheij (UvT), *The Eyes Have It*
2016-29 Nicolas Hning (TUD), *Peak reduction in decentralised electricity systems -Markets and prices for flexible planning*

2016-28 Mingxin Zhang (TUD), *Large-scale Agent-based Social Simulation - A study on epidemic prediction and control*
2016-27 Wen Li (TUD), *Understanding Geo-spatial Information on Social Media*
2016-26 Dilhan Thilakarathne (VU), *In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains*
2016-25 Julia Kiseleva (TU/e), *Using Contextual Information to Understand Searching and Browsing Behavior*
2016-24 Brend Wanders (UT), *Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach*
2016-23 Fei Cai (UVA), *Query Auto Completion in Information Retrieval*
2016-22 Grace Lewis (VU), *Software Architecture Strategies for Cyber-Foraging Systems*
2016-21 Alejandro Moreno Clleri (UT), *From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground*
2016-20 Daan Odijk (UVA), *Context & Semantics in News & Web Search*
2016-19 Julia Efremova (Tu/e), *Mining Social Structures from Genealogical Data*
2016-18 Albert Meroo Peuela (VU), *Refining Statistical Data on the Web*
2016-17 Berend Weel (VU), *Towards Embodied Evolution of Robot Organisms*
2016-16 Guangliang Li (UVA), *Socially Intelligent Autonomous Agents that Learn from Human Reward*
2016-15 Steffen Michels (RUN), *Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments*
2016-14 Ravi Khadka (UU), *Revisiting Legacy Software System Modernization*
2016-13 Nana Baah Gyan (VU), *The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach*
2016-12 Max Knobbout (UU), *Logics for Modelling and Verifying Normative Multi-Agent Systems*
2016-11 Anne Schuth (UVA), *Search Engines that Learn from Their Users*
2016-10 George Karafotias (VUA), *Parameter Control for Evolutionary Algorithms*
2016-09 Archana Nottamkandath (VU), *Trusting Crowdsourced Information on Cultural Artefacts*
2016-08 Matje van de Camp (TiU), *A Link to the Past: Constructing Historical Social Networks from Unstructured Data*
2016-07 Jeroen de Man (VU), *Measuring and modeling negative emotions for virtual training*
2016-06 Michel Wilson (TUD), *Robust scheduling in an uncertain environment*
2016-05 Evgeny Sherkhonov (UVA), *Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers*
2016-04 Laurens Rietveld (VU), *Publishing and Consuming Linked Data*
2016-03 Maya Sappelli (RUN), *Knowledge Work in Context: User Centered Knowledge Worker Support*
2016-02 Michiel Christiaan Meulendijk (UU), *Optimizing medication reviews through decision support: prescribing a better pill to swallow*
2016-01 Syed Saiden Abbas (RUN), *Recognition of Shapes by Humans and Machines*
2015-35 Jungxao Xu (TUD), *Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction*
2015-34 Victor de Graaf (UT), *Gesocial Recommender Systems*
2015-33 Frederik Schadd (TUD), *Ontology Mapping with Auxiliary Resources*
2015-32 Jerome Gard (UL), *Corporate Venture Management in SMEs*
2015-31 Yakup Koç (TUD), *On the robustness of Power Grids*
2015-30 Kiavash Bahreini (OU), *Real-time Multimodal Emotion Recognition in E-Learning*
2015-29 Hendrik Baier (UM), *Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains*
2015-28 Janet Bagorogoza (TiU), *KNOWLEDGE MANAGEMENT AND HIGH PERFORMANCE; The Uganda Financial Institutions Model for HPO*
2015-27 Sándor Héman (CWI), *Updating compressed colomn stores*
2015-26 Alexander Hogenboom (EUR), *Sentiment Analysis of Text Guided by Semantics and Structure*
2015-25 Steven Woudenberg (UU), *Bayesian Tools for Early Disease Detection*
2015-24 Richard Berendsen (UVA), *Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation*
2015-23 Luit Gazendam (VU), *Cataloguer Support in Cultural Heritage*
2015-22 Zhemin Zhu (UT), *Co-occurrence Rate Networks*
2015-21 Sibren Fetter (OUN), *Using Peer-Support to Expand and Stabilize Online Learning*
2015-20 Loïs Vanhée (UU), *Using Culture and Values to Support Flexible Coordination*
2015-19 Bernardo Tabuenca (OUN), *Ubiquitous Technology for Lifelong Learners*
2015-18 Holger Pirk (CWI), *Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories*
2015-17 André van Cleeff (UT), *Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs*
2015-16 Changyun Wei (UT), *Cognitive Coordination for Cooperative Multi-Robot Teamwork*
2015-15 Klaas Andries de Graaf (VU), *Ontology-based Software Architecture Documentation*
2015-14 Bart van Straalen (UT), *A cognitive approach to modeling bad news conversations*
2015-13 Giuseppe Procaccianti (VU), *Energy-Efficient Software*
2015-12 Julie M. Birkholz (VU), *Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks*
2015-11 Yongming Luo (TUE), *Designing algorithms for big graph datasets: A study of computing bisimulation and joins*
2015-10 Henry Hermans (OUN), *OpenU: design of an integrated system to support lifelong learning*
2015-09 Randy Klaassen (UT), *HCI Perspectives on Behavior Change Support Systems*
2015-08 Jie Jiang (TUD), *Organizational Compliance: An agent-based model for designing and evaluating organizational interactions*
2015-07 Maria-Hendrike Peetz (UvA), *Time-Aware Online Reputation Analysis*
2015-06 Farideh Heidari (TUD), *Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes*
2015-05 Christoph Bösch (UT), *Cryptographically Enforced Search Pattern Hiding*
2015-04 Howard Spoelstra (OUN), *Collaborations in Open Learning Environments*
2015-03 Twan van Laarhoven (RUN), *Machine learning for network data*
2015-02 Faiza Bukhsh (UvT), *Smart auditing: Innovative Compliance Checking in Customs Controls*
2015-01 Niels Netten (UvA), *Machine Learning for Relevance of Information in Crisis Response*
2014-47 Shangsong Liang (UVA), *Fusion and Diversification in Information Retrieval*
2014-46 Ke Tao (TUD), *Social Web Data Analytics: Relevance, Redundancy, Diversity*
2014-45 Birgit Schmitz (OU), *Mobile Games for Learning: A Pattern-Based Approach*
2014-44 Paulien Meesters (UvT), *Intelligent Blauw. Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden*
2014-43 Kevin Vlaanderen (UU), *Supporting Process Improvement using Method Increments*
2014-42 Carsten Eickhoff (CWI/TUD), *Contextual Multidimensional Relevance Models*
2014-41 Frederik Hogenboom (EUR), *Automated Detection of Financial Events in News Text*
2014-40 Walter Oboma (RUN), *A Framework for Knowledge Management Using ICT in Higher Education*

2014-39 Jasmina Maric (UvT), *Web Communities, Immigration and Social Capital*
2014-38 Danny Plass-Oude Bos (UT), *Making brain-computer interfaces better: improving usability through post-processing*
2014-37 Maral Dadvar (UT), *Experts and Machines United Against Cyberbullying*
2014-36 Joos Buijs (TUE), *Flexible Evolutionary Algorithms for Mining Structured Process Models*
2014-35 Joost van Oijen (UU), *Cognitive Agents in Virtual Worlds: A Middleware Design Approach*
2014-34 Christina Manteli (VU), *The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems*
2014-33 Tesfa Tegegne Asfaw (RUN), *Service Discovery in eHealth*
2014-32 Naser Ayat (UVA), *On Entity Resolution in Probabilistic Data*
2014-31 Leo van Moergestel (UU), *Agent Technology in Agile Multiparallel Manufacturing and Product Support*
2014-30 Peter de Kock Berenschot (UvT), *Anticipating Criminal Behaviour*
2014-29 Jaap Kabbedijk (UU), *Variability in Multi-Tenant Enterprise Software*
2014-28 Anna Chmielowiec (VU), *Decentralized k-Clique Matching*
2014-27 Rui Jorge Almeida (EUR), *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*
2014-26 Tim Baarslag (TUD), *What to Bid and When to Stop*
2014-25 Martijn Lappenschaar (RUN), *New network models for the analysis of disease interaction*
2014-24 Davide Ceolin (VU), *Trusting Semi-structured Web Data*
2014-23 Eleftherios Sidirourgos (UvA/CWI), *Space Efficient Indexes for the Big Data Era*
2014-22 Marieke Peeters (UU), *Personalized Educational Games - Developing agent-supported scenario-based training*
2014-21 Kassidy Clark (TUD), *Negotiation and Monitoring in Open Environments*
2014-20 Mena Habib (UT), *Named Entity Extraction and Disambiguation for Informal Text: The Missing Link*
2014-19 Vincius Ramos (TUE), *Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support*
2014-18 Mattijs Ghijsen (VU), *Methods and Models for the Design and Study of Dynamic Agent Organizations*
2014-17 Kathrin Dentler (VU), *Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability*
2014-16 Krystyna Milian (VU), *Supporting trial recruitment and design by automatically interpreting eligibility criteria*
2014-15 Natalya Mogles (VU), *Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare*
2014-14 Yangyang Shi (TUD), *Language Models With Meta-information*
2014-13 Arlette van Wissen (VU), *Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains*
2014-12 Willem van Willigen (VU), *Look Ma, No Hands: Aspects of Autonomous Vehicle Control*
2014-11 Janneke van der Zwaan (TUD), *An Empathic Virtual Buddy for Social Support*
2014-10 Ivan Salvador Razo Zapata (VU), *Service Value Networks*
2014-09 Philip Jackson (UvT), *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language*
2014-08 Samur Araujo (TUD), *Data Integration over Distributed and Heterogeneous Data Endpoints*
2014-07 Arya Adriansyah (TUE), *Aligning Observed and Modeled Behavior*
2014-06 Damian Tamburri (VU), *Supporting Networked Software Development*
2014-05 Jurriaan van Reijsen (UU), *Knowledge Perspectives on Advancing Dynamic Capability*
2014-04 Hanna Jochmann-Mannak (UT), *Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation*
2014-03 Sergio Raul Duarte Torres (UT), *Information Retrieval for Children: Search Behavior and Solutions*
2014-02 Fiona Tuliyano (RUN), *Combining System Dynamics with a Domain Modeling Method*
2014-01 Nicola Barile (UU), *Studies in Learning Monotone Models from Data*
2013-43 Marc Bron (UVA), *Exploration and Contextualization through Interaction and Concepts*
2013-42 Léon Planken (TUD), *Algorithms for Simple Temporal Reasoning*
2013-41 Jochem Liem (UVA), *Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning*
2013-40 Pim Nijssen (UM), *Monte-Carlo Tree Search for Multi-Player Games*
2013-39 Joop de Jong (TUD), *A Method for Enterprise Ontology based Design of Enterprise Information Systems*
2013-38 Eelco den Heijer (VU), *Autonomous Evolutionary Art*
2013-37 Dirk Börner (OUN), *Ambient Learning Displays*
2013-36 Than Lam Hoang (TUe), *Pattern Mining in Data Streams*
2013-35 Abdallah El Ali (UvA), *Minimal Mobile Human Computer Interaction*
2013-34 Kien Tjin-Kam-Jet (UT), *Distributed Deep Web Search*
2013-33 Qi Gao (TUD), *User Modeling and Personalization in the Microblogging Sphere*
2013-32 Kamakshi Rajagopal (OUN), *Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development*
2013-31 Dinh Khoa Nguyen (UvT), *Blueprint Model and Language for Engineering Cloud Applications*
2013-30 Joyce Nakatumba (TUE), *Resource-Aware Business Process Management: Analysis and Support*
2013-29 Iwan de Kok (UT), *Listening Heads*
2013-28 Frans van der Sluis (UT), *When Complexity becomes Interesting: An Inquiry into the Information eXperience*
2013-27 Mohammad Huq (UT), *Inference-based Framework Managing Data Provenance*
2013-26 Alireza Zarghami (UT), *Architectural Support for Dynamic Homecare Service Provisioning*
2013-25 Agnieszka Anna Latoszek-Berendsen (UM), *Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System*
2013-24 Haitham Bou Ammar (UM), *Automated Transfer in Reinforcement Learning*
2013-23 Patricio de Alencar Silva (UvT), *Value Activity Monitoring*
2013-22 Tom Claassen (RUN), *Causal Discovery and Logic*
2013-21 Sander Wubben (UvT), *Text-to-text generation by monolingual machine translation*
2013-20 Katja Hofmann (UvA), *Fast and Reliable Online Learning to Rank for Information Retrieval*
2013-19 Renze Steenhuizen (TUD), *Coordinated Multi-Agent Planning and Scheduling*
2013-18 Jeroen Janssens (UvT), *Outlier Selection and One-Class Classification*
2013-17 Koen Kok (VU), *The PowerMatcher: Smart Coordination for the Smart Electricity Grid*
2013-16 Eric Kok (UU), *Exploring the practical benefits of argumentation in multi-agent deliberation*
2013-15 Daniel Hennes (UM), *Multiagent Learning - Dynamic Games and Applications*
2013-14 Jafar Tanha (UVA), *Ensemble Approaches to Semi-Supervised Learning Learning*
2013-13 Mohammad Safiri (UT), *Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly*
2013-12 Marian Razavian (VU), *Knowledge-driven Migration to Services*
2013-11 Evangelos Pournaras (TUD), *Multi-level Reconfigurable Self-organization in Overlay Services*

2013-10 Jeewanie Jayasinghe Arachchige (UvT), *A Unified Modeling Framework for Service Design*
2013-09 Fabio Gori (RUN), *Metagenomic Data Analysis: Computational Methods and Applications*
2013-08 Robbert-Jan Merk (VU), *Making enemies: cognitive modeling for opponent agents in fighter pilot simulators*
2013-07 Giel van Lankveld (UvT), *Quantifying Individual Player Differences*
2013-06 Romulo Goncalves (CWI), *The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience*
2013-05 Dulce Pumareja (UT), *Groupware Requirements Evolutions Patterns*
2013-04 Chetan Yadati (TUD), *Coordinating autonomous planning and scheduling*
2013-03 Szymon Klarman (VU), *Reasoning with Contexts in Description Logics*
2013-02 Erietta Liarou (CWI), *MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing*
2013-01 Viorel Milea (EUR), *News Analytics for Financial Decision Support*
2012-51 Jeroen de Jong (TUD), *Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching*
2012-50 Steven van Kervel (TUD), *Ontologogy driven Enterprise Information Systems Engineering*
2012-49 Michael Kaisers (UM), *Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions*
2012-48 Jorn Bakker (TUE), *Handling Abrupt Changes in Evolving Time-series Data*
2012-47 Manos Tsagkias (UVA), *Mining Social Media: Tracking Content and Predicting Behavior*
2012-46 Simon Carter (UVA), *Exploration and Exploitation of Multilingual Data for Statistical Machine Translation*
2012-45 Benedikt Kratz (UvT), *A Model and Language for Business-aware Transactions*
2012-44 Anna Tordai (VU), *On Combining Alignment Techniques*
2012-42 Dominique Verpoorten (OU), *Reflection Amplifiers in self-regulated Learning*
2012-41 Sebastian Kelle (OU), *Game Design Patterns for Learning*
2012-40 Agus Gunawan (UvT), *Information Access for SMEs in Indonesia*
2012-39 Hassan Fatemi (UT), *Risk-aware design of value and coordination networks*
2012-38 Selmar Smit (VU), *Parameter Tuning and Scientific Testing in Evolutionary Algorithms*
2012-37 Agnes Nakakawa (RUN), *A Collaboration Process for Enterprise Architecture Creation*
2012-36 Denis Ssebugwawo (RUN), *Analysis and Evaluation of Collaborative Modeling Processes*
2012-35 Evert Haasdijk (VU), *Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics*
2012-34 Pavol Jancura (RUN), *Evolutionary analysis in PPI networks and applications*
2012-33 Rory Sie (OUN), *Coalitions in Cooperation Networks (COCOON)*
2012-32 Wietske Visser (TUD), *Qualitative multi-criteria preference representation and reasoning*
2012-31 Emily Bagarukayo (RUN), *A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure*
2012-30 Alina Pommeranz (TUD), *Designing Human-Centered Systems for Reflective Decision Making*
2012-29 Almer Tigelaar (UT), *Peer-to-Peer Information Retrieval*
2012-28 Nancy Pascall (UvT), *Engendering Technology Empowering Women*
2012-27 Hayrettin Gurkok (UT), *Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games*
2012-26 Emile de Maat (UVA), *Making Sense of Legal Text*
2012-25 Silja Eckartz (UT), *Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application*
2012-24 Laurens van der Werff (UT), *Evaluation of Noisy Transcripts for Spoken Document Retrieval*
2012-23 Christian Muehl (UT), *Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction*
2012-22 Thijs Vis (UvT), *Intelligence, politie en veiligheidsdienst: verenigbare grootheden?*
2012-21 Roberto Cornacchia (TUD), *Querying Sparse Matrices for Information Retrieval*
2012-20 Ali Bahramisharif (RUN), *Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing*
2012-19 Helen Schonenberg (TUE), *What's Next? Operational Support for Business Process Execution*
2012-18 Eltjo Poort (VU), *Improving Solution Architecting Practices*
2012-17 Amal Elgammal (UvT), *Towards a Comprehensive Framework for Business Process Compliance*
2012-16 Fiemke Both (VU), *Helping people by understanding them - Ambient Agents supporting task execution and depression treatment*
2012-15 Natalie van der Wal (VU), *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.*
2012-14 Evgeny Knutov (TUE), *Generic Adaptation Framework for Unifying Adaptive Web-based Systems*
2012-13 Suleman Shahid (UvT), *Fun and Face: Exploring non-verbal expressions of emotion during playful interactions*
2012-12 Kees van der Sluijs (TUE), *Model Driven Design and Data Integration in Semantic Web Information Systems*
2012-11 J.C.B. Rantham Prabhakara (TUE), *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*
2012-10 David Smits (TUE), *Towards a Generic Distributed Adaptive Hypermedia Environment*
2012-09 Ricardo Neisse (UT), *Trust and Privacy Management Support for Context-Aware Service Platforms*
2012-08 Gerben de Vries (UVA), *Kernel Methods for Vessel Trajectories*
2012-07 Rianne van Lambalgen (VU), *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions*
2012-06 Wolfgang Reinhardt (OU), *Awareness Support for Knowledge Workers in Research Networks*
2012-05 Marijn Plomp (UU), *Maturing Interorganisational Information Systems*
2012-04 Jurriaan Souer (UU), *Development of Content Management System-based Web Applications*
2012-03 Adam Vanya (VU), *Supporting Architecture Evolution by Mining Software Repositories*
2012-02 Muhammad Umair (VU), *Adaptivity, emotion, and Rationality in Human and Ambient Agent Models*
2012-01 Terry Kakeeto (UvT), *Relationship Marketing for SMEs in Uganda*
2011-49 Andreea Niculescu (UT), *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality*
2011-48 Mark Ter Maat (UT), *Response Selection and Turn-taking for a Sensitive Artificial Listening Agent*
2011-47 Azizi Bin Ab Aziz (VU), *Exploring Computational Models for Intelligent Support of Persons with Depression*
2011-46 Beibei Hu (TUD), *Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work*
2011-45 Herman Stehouwer (UvT), *Statistical Language Models for Alternative Sequence Selection*
2011-44 Boris Reuderink (UT), *Robust Brain-Computer Interfaces*
2011-43 Henk van der Schuur (UU), *Process Improvement through Software Operation Knowledge*
2011-42 Michal Sindlar (UU), *Explaining Behavior through Mental State Attribution*
2011-41 Luan Ibraimi (UT), *Cryptographically Enforced Distributed Data Access Control*
2011-40 Viktor Clerc (VU), *Architectural Knowledge Management in Global Software Development*

2011-39 Joost Westra (UU), *Organizing Adaptation using Agents in Serious Games*
2011-38 Nyree Lemmens (UM), *Bee-inspired Distributed Optimization*
2011-37 Adriana Burlutiu (RUN), *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference*
2011-36 Erik van der Spek (UU), *Experiments in serious game design: a cognitive approach*
2011-35 Maaike Harbers (UU), *Explaining Agent Behavior in Virtual Training*
2011-34 Paolo Turrini (UU), *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations*
2011-33 Tom van der Weide (UU), *Arguing to Motivate Decisions*
2011-32 Nees-Jan van Eck (EUR), *Methodological Advances in Bibliometric Mapping of Science*
2011-31 Ludo Waltman (EUR), *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*
2011-30 Egon van den Broek (UT), *Affective Signal Processing (ASP): Unraveling the mystery of emotions*
2011-29 Faisal Kamiran (TUE), *Discrimination-aware Classification*
2011-28 Rianne Kaptein (UVA), *Effective Focused Retrieval by Exploiting Query Context and Document Structure*
2011-27 Aniel Bhulai (VU), *Dynamic website optimization through autonomous management of design patterns*
2011-26 Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*
2011-25 Syed Waqar ul Qounain Jaffry (VU)), *Analysis and Validation of Models for Trust Dynamics*
2011-24 Herwin van Welbergen (UT), *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior*
2011-23 Wouter Weerkamp (UVA), *Finding People and their Utterances in Social Media*
2011-22 Junte Zhang (UVA), *System Evaluation of Archival Description and Access*
2011-21 Linda Terlouw (TUD), *Modularization and Specification of Service-Oriented Systems*
2011-20 Qing Gu (VU), *Guiding service-oriented software engineering - A view-based approach*
2011-19 Ellen Rusman (OU), *The Mind's Eye on Personal Profiles*
2011-18 Mark Ponsen (UM), *Strategic Decision-Making in complex games*
2011-17 Jiyin He (UVA), *Exploring Topic Structure: Coherence, Diversity and Relatedness*
2011-16 Maarten Schadd (UM), *Selective Search in Games of Different Complexity*
2011-15 Marijn Koolen (UvA), *The Meaning of Structure: the Value of Link Evidence for Information Retrieval*
2011-14 Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets*
2011-13 Xiaoyu Mao (UvT), *Airport under Control. Multiagent Scheduling for Airport Ground Handling*
2011-12 Carmen Bratosin (TUE), *Grid Architecture for Distributed Process Mining*
2011-11 Dhaval Vyas (UT), *Designing for Awareness: An Experience-focused HCI Perspective*
2011-10 Bart Bogaert (UvT), *Cloud Content Contention*
2011-09 Tim de Jong (OU), *Contextualised Mobile Media for Learning*
2011-08 Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues*
2011-07 Yujia Cao (UT), *Multimodal Information Presentation for High Load Human Computer Interaction*
2011-06 Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage*
2011-05 Base van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.*
2011-04 Hado van Hasselt (UU), *Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms*
2011-03 Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems*
2011-02 Nick Tinnemeier (UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*
2011-01 Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*
2010-53 Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access*
2010-52 Peter-Paul van Maanen (VU), *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention*
2010-51 Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources*
2010-50 Bouke Huurnink (UVA), *Search in Audiovisual Broadcast Archives*
2010-49 Jahn-Takeshi Saito (UM), *Solving difficult game positions*
2010-47 Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples*
2010-46 Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*
2010-45 Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services*
2010-44 Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*
2010-43 Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*
2010-42 Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach*
2010-41 Guillaume Chaslot (UM), *Monte-Carlo Tree Search*
2010-40 Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web*
2010-39 Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents*
2010-38 Dirk Fahland (TUE), *From Scenarios to components*
2010-37 Niels Lohmann (TUE), *Correctness of services and their composition*
2010-36 Jose Janssen (OU), *Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification*
2010-35 Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval*
2010-34 Teduh Dirgahayu (UT), *Interaction Design in Service Compositions*
2010-33 Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*
2010-32 Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*
2010-31 Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web*
2010-30 Marieke van Erp (UvT), *Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval*
2010-29 Stratos Idreos (CWI), *Database Cracking: Towards Auto-tuning Database Kernels*
2010-28 Arne Koopman (UU), *Characteristic Relational Patterns*
2010-27 Marten Voulon (UL), *Automatisch contracteren*
2010-26 Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*
2010-25 Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*
2010-24 Dmytro Tykhonov, *Designing Generic and Efficient Negotiation Strategies*
2010-23 Bas Steunebrink (UU), *The Logical Structure of Emotions*
2010-22 Michiel Hildebrand (CWI), *End-user Support for Access to Heterogeneous Linked Data*
2010-21 Harold van Heerde (UT), *Privacy-aware data management by means of data degradation*

2010-20 Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*
2010-19 Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems*
2010-18 Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation*
2010-17 Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*
2010-16 Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice*
2010-15 Lianne Bodenstaff (UT), *Managing Dependency Relations in Inter-Organizational Models*
2010-14 Sander van Splunter (VU), *Automated Web Service Reconfiguration*
2010-13 Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques*
2010-12 Susan van den Braak (UU), *Sensemaking software for crime analysis*
2010-11 Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning*
2010-10 Rebecca Ong (UL), *Mobile Communication and Protection of Children*
2010-09 Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging*
2010-08 Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments*
2010-07 Wim Fikkert (UT), *Gesture interaction at a Distance*
2010-06 Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI*
2010-05 Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems*
2010-04 Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*
2010-03 Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents*
2010-02 Ingo Wassink (UT), *Work flows in Life Science*
2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter*
2009-46 Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion*
2009-45 Jilles Vreeken (UU), *Making Pattern Mining Useful*
2009-44 Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations*
2009-43 Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*
2009-42 Toine Bogers (UvT), *Recommender Systems for Social Bookmarking*
2009-41 Igor Berezhnyy (UvT), *Digital Analysis of Paintings*
2009-40 Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language*
2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets*
2009-38 Riina Vuorikari (OU), *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*
2009-37 Hendrik Drachsler (OUN), *Navigation Support for Learners in Informal Learning Networks*
2009-36 Marco Kalz (OUN), *Placement Support for Learners in Learning Networks*
2009-35 Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling*
2009-34 Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach*
2009-33 Khiet Truong (UT), *How Does Real Affect Affect Affect Recognition In Speech?*
2009-32 Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors*
2009-31 Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text*
2009-30 Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage*
2009-29 Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications*
2009-28 Sander Evers (UT), *Sensor Data Management with Probabilistic Models*
2009-27 Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web*
2009-26 Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services*
2009-25 Alex van Ballegooij (CWI), *"RAM: Array Database Management through Relational Mapping"*
2009-24 Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations*
2009-23 Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment*
2009-22 Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence*
2009-21 Stijn Vanderlooy (UM), *Ranking and Reliable Classification*
2009-20 Bob van der Vecht (UU), *Adjustable Autonomy: Controling Influences on Decision Making*
2009-19 Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*
2009-18 Fabian Groffen (CWI), *Armada, An Evolving Database System*
2009-17 Laurens van der Maaten (UvT), *Feature Extraction from Visual Data*
2009-16 Fritz Reul (UvT), *New Architectures in Computer Chess*
2009-15 Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense*
2009-14 Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*
2009-13 Steven de Jong (UM), *Fairness in Multi-Agent Systems*
2009-12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin), *Operating Guidelines for Services*
2009-11 Alexander Boer (UVA), *Legal Theory, Sources of Law & the Semantic Web*
2009-10 Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications*
2009-09 Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems*
2009-08 Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*
2009-07 Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion*
2009-06 Muhammad Subianto (UU), *Understanding Classification*
2009-05 Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*
2009-04 Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering*
2009-03 Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT*
2009-02 Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques*
2009-01 Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models*
2008-35 Ben Torben Nielsen (UvT), *Dendritic morphologies: function shapes structure*
2008-34 Jeroen de Knijf (UU), *Studies in Frequent Tree Mining*
2008-33 Frank Terpstra (UVA), *Scientific Workflow Design; theoretical and practical issues*
2008-32 Trung H. Bui (UT), *Toward Affective Dialogue Management using Partially Observable Markov Decision Processes*
2008-31 Loes Braun (UM), *Pro-Active Medical Information Retrieval*

2008-30 Wouter van Atteveldt (VU), *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*
2008-29 Dennis Reidsma (UT), *Annotations and Subjective Machines - Of Annotators, Embodied Agents, Users, and Other Humans*
2008-28 Ildiko Flesch (RUN), *On the Use of Independence Relations in Bayesian Networks*
2008-27 Hubert Vogten (OU), *Design and Implementation Strategies for IMS Learning Design*
2008-26 Marijn Huijbregts (UT), *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*
2008-25 Geert Jonker (UU), *Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency*
2008-24 Zharko Aleksovski (VU), *Using background knowledge in ontology matching*
2008-23 Stefan Visscher (UU), *Bayesian network models for the management of ventilator-associated pneumonia*
2008-22 Henk Koning (UU), *Communication of IT-Architecture*
2008-21 Krisztian Balog (UVA), *People Search in the Enterprise*
2008-20 Rex Arendsen (UVA), *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven.*
2008-19 Henning Rode (UT), *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*
2008-18 Guido de Croon (UM), *Adaptive Active Vision*
2008-17 Martin Op 't Land (TUD), *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*
2008-16 Henriette van Vugt (VU), *Embodied agents from a user's perspective*
2008-15 Martijn van Otterlo (UT), *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains.*
2008-14 Arthur van Bunningen (UT), *Context-Aware Querying; Better Answers with Less Effort*
2008-13 Caterina Carraciolo (UVA), *Topic Driven Access to Scientific Handbooks*
2008-12 Jozsef Farkas (RUN), *A Semiotically Oriented Cognitive Model of Knowledge Representation*
2008-11 Vera Kartseva (VU), *Designing Controls for Network Organizations: A Value-Based Approach*
2008-10 Wauter Bosma (UT), *Discourse oriented summarization*
2008-09 Christof van Nimwegen (UU), *The paradox of the guided user: assistance can be counter-effective*
2008-08 Janneke Bolt (UU), *Bayesian Networks: Aspects of Approximate Inference*
2008-07 Peter van Rosmalen (OU), *Supporting the tutor in the design and support of adaptive e-learning*
2008-06 Arjen Hommersom (RUN), *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*
2008-05 Bela Mutschler (UT), *Modeling and simulating causal dependencies on process-aware information systems from a cost perspective*
2008-04 Ander de Keijzer (UT), *Management of Uncertain Data - towards unattended integration*
2008-03 Vera Hollink (UVA), *Optimizing hierarchical menus: a usage-based approach*
2008-02 Alexei Sharpanskykh (VU), *On Computer-Aided Methods for Modeling and Analysis of Organizations*
2008-01 Katalin Boer-Sorbán (EUR), *Agent-Based Simulation of Financial Markets: A modular, continuous-time approach*
2007-25 Joost Schalken (VU), *Empirical Investigations in Software Process Improvement*
2007-24 Georgina Ramírez Camps (CWI), *Structural Features in XML Retrieval*
2007-23 Peter Barna (TUE), *Specification of Application Logic in Web Information Systems*
2007-22 Zlatko Zlatev (UT), *Goal-oriented design of value and process models from patterns*
2007-21 Karianne Vermaas (UU), *Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005*
2007-20 Slinger Jansen (UU), *Customer Configuration Updating in a Software Supply Network*
2007-19 David Levy (UM), *Intimate relationships with artificial partners*
2007-18 Bart Orriens (UvT), *On the development an management of adaptive business collaborations*
2007-17 Theodore Charitos (UU), *Reasoning with Dynamic Networks in Practice*
2007-16 Davide Grossi (UU), *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*
2007-15 Joyca Lacroix (UM), *NIM: a Situated Computational Memory Model*
2007-14 Niek Bergboer (UM), *Context-Based Image Analysis*
2007-13 Rutger Rienks (UT), *Meetings in Smart Environments; Implications of Progressing Technology*
2007-12 Marcel van Gerven (RUN), *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*
2007-11 Natalia Stash (TUE), *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*
2007-10 Huib Aldewereld (UU), *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*
2007-09 David Mobach (VU), *Agent-Based Mediated Service Negotiation*
2007-08 Mark Hoogendoorn (VU), *Modeling of Change in Multi-Agent Organizations*
2007-07 Natasa Jovanovic (UT), *To Whom It May Concern - Addressee Identification in Face-to-Face Meetings*
2007-06 Gilad Mishne (UVA), *Applied Text Analytics for Blogs*
2007-05 Bart Schermer (UL), *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*
2007-04 Jurriaan van Diggelen (UU), *Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach*
2007-03 Peter Mika (VU), *Social Networks and the Semantic Web*
2007-02 Wouter Teepe (RUG), *Reconciling Information Exchange and Confidentiality: A Formal Approach*
2007-01 Kees Leune (UvT), *Access Control and Service-Oriented Architectures*
2006-28 Borkur Sigurbjornsson (UVA), *Focused Information Access using XML Element Retrieval*
2006-27 Stefano Bocconi (CWI), *Vox Populi: generating video documentaries from semantically annotated media repositories*
2006-26 Vojkan Mihajlovic (UT), *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*
2006-25 Madalina Drugan (UU), *Conditional log-likelihood MDL and Evolutionary MCMC*
2006-24 Laura Hollink (VU), *Semantic Annotation for Retrieval of Visual Resources*
2006-23 Ion Juvina (UU), *Development of Cognitive Model for Navigating on the Web*
2006-22 Paul de Vrieze (RUN), *Fundaments of Adaptive Personalisation*
2006-21 Bas van Gils (RUN), *Aptness on the Web*
2006-20 Marina Velikova (UvT), *Monotone models for prediction in data mining*
2006-19 Birna van Riemsdijk (UU), *Cognitive Agent Programming: A Semantic Approach*
2006-18 Valentin Zhizhkun (UVA), *Graph transformation for Natural Language Processing*
2006-17 Stacey Nagata (UU), *User Assistance for Multitasking with Interruptions on a Mobile Device*
2006-16 Carsten Riggelsen (UU), *Approximation Methods for Efficient Learning of Bayesian Networks*
2006-15 Rainer Malik (UU), *CONAN: Text Mining in the Biomedical Domain*
2006-14 Johan Hoorn (VU), *Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change*

2006-13 Henk-Jan Lebbink (UU), *Dialogue and Decision Games for Information Exchanging Agents*
2006-12 Bert Bongers (VU), *Interactivation - Towards an e-cology of people, our technological environment, and the arts*
2006-11 Joeri van Ruth (UT), *Flattening Queries over Nested Data Types*
2006-10 Ronny Siebes (VU), *Semantic Routing in Peer-to-Peer Systems*
2006-09 Mohamed Wahdan (UM), *Automatic Formulation of the Auditor's Opinion*
2006-08 Eelco Herder (UT), *Forward, Back and Home Again - Analyzing User Behavior on the Web*
2006-07 Marko Smiljanic (UT), *XML schema matching – balancing efficiency and effectiveness by means of clustering*
2006-06 Ziv Baida (VU), *Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling*
2006-05 Cees Pierik (UU), *Validation Techniques for Object-Oriented Proof Outlines*
2006-04 Marta Sabou (VU), *Building Web Service Ontologies*
2006-03 Noor Christoph (UVA), *The role of metacognitive skills in learning to solve problems*
2006-02 Cristina Chisalita (VU), *Contextual issues in the design and use of information technology in organizations*
2006-01 Samuil Angelov (TUE), *Foundations of B2B Electronic Contracting*
2005-21 Wijnand Derks (UT), *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*
2005-20 Cristina Coteanu (UL), *Cyber Consumer Law, State of the Art and Perspectives*
2005-19 Michel van Dartel (UM), *Situated Representation*
2005-18 Danielle Sent (UU), *Test-selection strategies for probabilistic networks*
2005-17 Boris Shishkov (TUD), *Software Specification Based on Re-usable Business Components*
2005-16 Joris Graaumans (UU), *Usability of XML Query Languages*
2005-15 Tibor Bosse (VU), *Analysis of the Dynamics of Cognitive Processes*
2005-14 Borys Omelayenko (VU), *Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics*
2005-13 Fred Hamburg (UL), *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*
2005-12 Csaba Boer (EUR), *Distributed Simulation in Industry*
2005-11 Elth Ogston (VU), *Agent Based Matchmaking and Clustering - A Decentralized Approach to Search*
2005-10 Anders Bouwer (UVA), *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*
2005-09 Jeen Broekstra (VU), *Storage, Querying and Inferencing for Semantic Web Languages*
2005-08 Richard Vdovjak (TUE), *A Model-driven Approach for Building Distributed Ontology-based Web Applications*
2005-07 Flavius Frasincar (TUE), *Hypermedia Presentation Generation for Semantic Web Information Systems*
2005-06 Pieter Spronck (UM), *Adaptive Game AI*
2005-05 Gabriel Infante-Lopez (UVA), *Two-Level Probabilistic Grammars for Natural Language Parsing*
2005-04 Nirvana Meratnia (UT), *Towards Database Support for Moving Object data*
2005-03 Franc Grootjen (RUN), *A Pragmatic Approach to the Conceptualisation of Language*
2005-02 Erik van der Werf (UM)), *AI techniques for the game of Go*
2005-01 Floor Verdenius (UVA), *Methodological Aspects of Designing Induction-Based Applications*
2004-20 Madelon Evers (Nyenrode), *Learning from Design: facilitating multidisciplinary design teams*
2004-19 Thijs Westerveld (UT), *Using generative probabilistic models for multimedia retrieval*
2004-18 Vania Bessa Machado (UvA), *Supporting the Construction of Qualitative Knowledge Models*
2004-17 Mark Winands (UM), *Informed Search in Complex Games*
2004-16 Federico Divina (VU), *Hybrid Genetic Relational Search for Inductive Learning*
2004-15 Arno Knobbe (UU), *Multi-Relational Data Mining*
2004-14 Paul Harrenstein (UU), *Logic in Conflict. Logical Explorations in Strategic Equilibrium*
2004-13 Wojciech Jamroga (UT), *Using Multiple Models of Reality: On Agents who Know how to Play*
2004-12 The Duy Bui (UT), *Creating emotions and facial expressions for embodied agents*
2004-11 Michel Klein (VU), *Change Management for Distributed Ontologies*
2004-10 Suzanne Kabel (UVA), *Knowledge-rich indexing of learning-objects*
2004-09 Martin Caminada (VU), *For the Sake of the Argument; explorations into argument-based reasoning*
2004-08 Joop Verbeek (UM), *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise*
2004-07 Elise Boltjes (UM), *Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes*
2004-06 Bart-Jan Hommes (TUD), *The Evaluation of Business Process Modeling Techniques*
2004-05 Viara Popova (EUR), *Knowledge discovery and monotonicity*
2004-04 Chris van Aart (UVA), *Organizational Principles for Multi-Agent Architectures*
2004-03 Perry Groot (VU), *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*
2004-02 Lai Xu (UvT), *Monitoring Multi-party Contracts for E-business*
2004-01 Virginia Dignum (UU), *A Model for Organizational Interaction: Based on Agents, Founded in Logic*
2003-18 Levente Kocsis (UM), *Learning Search Decisions*
2003-17 David Jansen (UT), *Extensions of Statecharts with Probability, Time, and Stochastic Timing*
2003-16 Menzo Windhouwer (CWI), *Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses*
2003-15 Mathijs de Weerdt (TUD), *Plan Merging in Multi-Agent Systems*
2003-14 Stijn Hoppenbrouwers (KUN), *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*
2003-13 Jeroen Donkers (UM), *Nosce Hostem - Searching with Opponent Models*
2003-12 Roeland Ordelman (UT), *Dutch speech recognition in multimedia information retrieval*
2003-11 Simon Keizer (UT), *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*
2003-10 Andreas Lincke (UvT), *Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture*
2003-09 Rens Kortmann (UM), *The resolution of visually guided behaviour*
2003-08 Yongping Ran (UM), *Repair Based Scheduling*
2003-07 Machiel Jansen (UvA), *Formal Explorations of Knowledge Intensive Tasks*
2003-06 Boris van Schooten (UT), *Development and specification of virtual environments*
2003-05 Jos Lehmann (UVA), *Causation in Artificial Intelligence and Law - A modelling approach*
2003-04 Milan Petkovic (UT), *Content-Based Video Retrieval Supported by Database Technology*
2003-03 Martijn Schuemie (TUD), *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*
2003-02 Jan Broersen (VU), *Modal Action Logics for Reasoning About Reactive Systems*
2003-01 Heiner Stuckenschmidt (VU), *Ontology-Based Information Sharing in Weakly Structured Environments*
2002-17 Stefan Manegold (UVA), *Understanding, Modeling, and Improving Main-Memory Database Performance*

2002-16 Pieter van Langen (VU), *The Anatomy of Design: Foundations, Models and Applications*

2002-15 Rik Eshuis (UT), *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*

2002-14 Wieke de Vries (UU), *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*

2002-13 Hongjing Wu (TUE), *A Reference Architecture for Adaptive Hypermedia Applications*

2002-12 Albrecht Schmidt (Uva), *Processing XML in Database Systems*

2002-11 Wouter C.A. Wijngaards (VU), *Agent Based Modelling of Dynamics: Biological and Organisational Applications*

2002-10 Brian Sheppard (UM), *Towards Perfect Play of Scrabble*

2002-09 Willem-Jan van den Heuvel (KUB), *Integrating Modern Business Applications with Objectified Legacy Systems*

2002-08 Jaap Gordijn (VU), *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*

2002-07 Peter Boncz (CWI), *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*

2002-06 Laurens Mommers (UL), *Applied legal epistemology; Building a knowledge-based ontology of the legal domain*

2002-05 Radu Serban (VU), *The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents*

2002-04 Juan Roberto Castelo Valdueza (UU), *The Discrete Acyclic Digraph Markov Model in Data Mining*

2002-03 Henk Ernst Blok (UT), *Database Optimization Aspects for Information Retrieval*

2002-02 Roelof van Zwol (UT), *Modelling and searching web-based document collections*

2002-01 Nico Lassing (VU), *Architecture-Level Modifiability Analysis*

2001-11 Tom M. van Engers (VUA), *Knowledge Management: The Role of Mental Models in Business Systems Design*

2001-10 Maarten Sierhuis (UvA), *Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design*

2001-09 Pieter Jan 't Hoen (RUL), *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*

2001-08 Pascal van Eck (VU), *A Compositional Semantic Structure for Multi-Agent Systems Dynamics.*

2001-07 Bastiaan Schonhage (VU), *Diva: Architectural Perspectives on Information Visualization*

2001-06 Martijn van Welie (VU), *Task-based User Interface Design*

2001-05 Jacco van Ossenbruggen (VU), *Processing Structured Hypermedia: A Matter of Style*

2001-04 Evgueni Smirnov (UM), *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*

2001-03 Maarten van Someren (UvA), *Learning as problem solving*

2001-02 Koen Hindriks (UU), *Agent Programming Languages: Programming with Mental Models*

2001-01 Silja Renooij (UU), *Qualitative Approaches to Quantifying Probabilistic Networks*

2000-11 Jonas Karlsson (CWI), *Scalable Distributed Data Structures for Database Management*

2000-10 Niels Nes (CWI), *Image Database Management System Design Considerations, Algorithms and Architecture*

2000-09 Florian Waas (CWI), *Principles of Probabilistic Query Optimization*

2000-08 Veerle Coupé (EUR), *Sensitivity Analyis of Decision-Theoretic Networks*

2000-07 Niels Peek (UU), *Decision-theoretic Planning of Clinical Patient Management*

2000-06 Rogier van Eijk (UU), *Programming Languages for Agent Communication*

2000-05 Ruud van der Pol (UM), *Knowledge-based Query Formulation in Information Retrieval.*

2000-04 Geert de Haan (VU), *ETAG, A Formal Model of Competence Knowledge for User Interface Design*

2000-03 Carolien M.T. Metselaar (UVA), *Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief.*

2000-02 Koen Holtman (TUE), *Prototyping of CMS Storage Management*

2000-01 Frank Niessink (VU), *Perspectives on Improving Software Maintenance*

1999-08 Jacques H.J. Lenting (UM), *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation.*

1999-07 David Spelt (UT), *Verification support for object database design*

1999-06 Niek J.E. Wijngaards (VU), *Re-design of compositional systems*

1999-05 Aldo de Moor (KUB), *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*

1999-04 Jacques Penders (UM), *The practical Art of Moving Physical Objects*

1999-03 Don Beal (UM), *The Nature of Minimax Search*

1999-02 Rob Potharst (EUR), *Classification using decision trees and neural nets*

1999-01 Mark Sloof (VU), *Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products*

1998-05 E.W.Oskamp (RUL), *Computerondersteuning bij Straftoemeting*

1998-04 Dennis Breuker (UM), *Memory versus Search in Games*

1998-03 Ans Steuten (TUD), *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*

1998-02 Floris Wiesman (UM), *Information Retrieval by Graphically Browsing Meta-Information*

1998-01 Johan van den Akker (CWI), *DEGAS - An Active, Temporal Database of Autonomous Objects*