AN EXPERIMENTAL STUDY OF SELF-REGULATED LEARNING STRATEGIES APPLICATION IN MOOCS

by

Shu-Yi Hsu

Dissertation Committee:

Professor Yoo Kyung Chang, Sponsor
Professor Paulo Blikstein

Approved by the Committee on the Degree of Doctor of Education

Date__________June 30th, 2021____________

Submitted in partial fulfillment of the
requirements of the degree of Doctor of Education
Teachers College, Columbia University

2021

ABSTRACT

AN EXPERIMENTAL STUDY OF

SELF-REGULATED LEARNING STRATEGIES

APPLICATION IN MOOCS

Shu-Yi Hsu

Online learning has been widely adopted in higher education to reach students who typically would not have a chance to complete accredited courses (Kentnor, 2015). Massive open online courses (MOOC), which is a type of online learning, makes it easier for people to take university courses with internet access and a fraction of cost compared to traditional residential programs (Reich, 2020). MOOCs also become popular for those who want to increase their professional profile or advance their academic career (Pheatt, 2017). However, online learning has long been criticized for its universally low completion rates, high dropout rate and poor learning performance (Almeda et al., 2018). This phenomenon is more exacerbated in MOOC environments. Historical studies have attempted to support learner self-regulated learning (SRL) activities in order to enhance completion rates and academic outcomes. Prior studies have conducted pre-course questionnaires as inexpensive SRL interventions to prompt learners as SRL support

(Kizilcec et al., 2017, Kizilcec & Cohen, 2017; Kizilcec et al., 2020; Yeomans & Reich, 2017). Yet, these one-time-only, short-term interventions only yield limited or no effects. This study implemented and evaluated the effectiveness of an alternative intervention, the self-regulated learning user interface (SRLUI), to support students' self-regulated learning (SRL) strategies in a MOOC environment. SRLUI is based on Zimmerman’s (2000) SRL model and develops learner’s SRL skills through longitudinal, recurring practice of multiple SRL dimensions activities (i.e., goal setting, self-evaluation, task planning, setting reminders) with content-specific information. The study utilized a randomized experimental design and implemented SRLUI in eight MOOCs with a total of 808 participants. The results indicated a higher usage rate of SRL support compared to the historical findings, which may be owing to the SRL support embedded into the learning activities throughout the course. Also, the study showed improved learning outcomes for a subgroup of participants, but there was no reduction in the number of dropouts.

Based on the findings of this study, it is recommended that a personalized SRL tool featuring content-specific information should be embedded in online courses. The research design also recorded direct cognitive records of learners' SRL activities, which yield stronger validity compared to trace and survey data. The result suggested SRLUI might only benefit a subgroup of learners with passing grades. Thus, it is recommended that future research identify various subgroups of learner profiles in MOOC environments and to consider how to reach and support learners in different subgroups.

DEDICATION

For my mother, Chin-Yun Cheng and my partner, Bo Pascual.

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful for my advisor, Dr. Yoo Kyung Chang, whose feedback and encouragement have been invaluable to completing my dissertation. Thank you for empowering me to think critically and independently through our rounds of discussion. Next, I want to express my appreciation to my mentor, Shane Tutwiler. You saw the potential for something that I am not sure would ever have been possible without you. Your mentorship helps me grow my scholarship. Thank you for being such a super teacher in my life. Paulo Blikstein has constantly introduced different perspectives to expand my arguments. Alex Bowers has offered me extensive mentorship and guidance through the dissertation process. Thank you for the constructive feedback and generosity in sharing resources to help me refine my research and presentation. Barbara Tversky has provided me with valuable feedback to extend my research beyond my dissertation.

I am grateful for my colleagues at the Columbia Engineering School and the Provost Office of Online Learning, especially Marina Zamalin, Soulaymane Kachani, Merrell Norden, Alexander Browne, and Abby Lee. Elizabeth Strauss. You are my role models, setting an example for me on remaining professional under tremendous pressure and completing the doctoral studies while working full time. Thank you for assisting me in conducting strategic planning on finishing the last leg of the doctoral journey. I also want to thank those who participated in the SRLUI project.

My doctoral colleagues at the Teachers College are like my family in the U.S – Abbas Abbasov, Nicola Law, Elliot Hu-Au, Sabine Zander, Paul Stengel, Dino Sossi, Eddie Lin, Sam Thanapornsangsuth, Monica Chan, Stephanie Yang, and Yipu Zheng. Your company and encouragement helped me every step in this journey.

My study group and my spiritual siblings – Michelle Kuhn, Austin Kuhn, Katelyn VanderVeen, Ian Miller, Madeline Benson, and Kirk Benson. Thank you for lifting me up when I feel low. Thank you for establishing my belief in God through you.

Thank you to my friends from Taiwan – Mingyun Hsu, Summer Hsia, Hsiao-Chen Cheng, Jia Ru Hsu, Michelle Yen. Jesus Trapero, Erin Lin, Chieh Chen, and Ming Chao Lin. I am incredibly lucky to have such extended friendship with you.

My partner, Bo Pascual. Thank you for your support, patience, and encouragement. Thank you for being my best cheerleader to celebrate me in every milestone. Finally, thank you to my parents. Thank you for letting me choose the path at every crossroad in my life. Thank you for believing in me and making no judgments on my choices. This work is dedicated to mom. I love you.

S.H.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I – INTRODUCTION

Statement of the Problem

The onset of the pandemic in the Spring of 2020 caused a global impact on education (Burgess & Sievertsen, 2020). All of a sudden, online became the only venue for teaching and learning. As a result, it has drawn the attention of multiple stakeholders; forcing not only researchers, teachers, and administrations, but also parents and students to consider the need to better support learning in online environments. However, online learning is not new in educational history, instead, online learning has persisted before, through, and after the pandemic for over 300 years. Prior to the digital era, distance education started with correspondence education back in the 18th century (Kentnor, 2015). The initiatives of distance education were to offer opportunities for those who were under-represented or those who normally wouldn't have access to traditional schools such as women and railroad and mine workers (Casey, 2008). Through the change of technologies, the medium for distance learning has evolved from parcels, radio shows, instructional television, computers to online learning (Liyanagunawardena et al., 2013). Distance learning allows people to pursue vocational training or academic degrees without leaving their full-time job or moving from home (Arkorful & Abaidoo, 2015).

Online learning has been widely adopted in higher education to increase access to higher education around the world in countries such as the U.S., China, India, and South Korea (Kumar et al., 2017). Starting from 2012, the Massive Open Online Course (MOOC) platform has opened a door for students to take university courses without going through the admission process (Reich, 2020). Anyone with internet access could register for a MOOC at a fraction of the cost for personal interest or advancing their academic or professional career (Pheatt, 2017). For example, one of the most successful MOOC programs was launched by Georgia Tech, which is also the first computer science master's program entirely on a MOOC platform, Udacity (OMSCS, n.d.). The program has received over 25,000 applications and enrolled more than 10,000 students up to 2021 (OMSCS). Up until now, boosted by the pandemic, total MOOC enrollments have reached 180 million students with 16,000 courses offered across 950 universities in 2020 (Class Central, 2020).

The large enrollment numbers in MOOCs have attracted widespread interest from educators and researchers for different reasons (Gardner & Brooks, 2018). The popular press and social media described MOOCs as having the potential to democratize education and as providing an economical alternative to higher education offerings (Shirky, 2013). Conversely, educators and researchers have expressed concerns about high dropout rates and information transmitted pedagogy with limited interactions on the MOOC platforms (Reich, 2020). Despite the concerns on the efficacy of MOOCs, MOOC enrollment numbers still continue to expand rapidly (Reich & Ruiperez-Valiente, 2019). Various research has been conducted over the past decade to help identify pedagogical (Quintana & Tan, 2019), affective (Green et al., 2015), and motivational supports (Stein & Allione, 2014; Xu & Yang, 2016) for the wide range of students who choose to engage in MOOC learning environments.

One area of active research in this field is the degree to which students' self-regulated learning (SRL) has been shown to be positively related to achievement in various educational settings (Azevedo & Aleven, 2013; Dbbagh & Kitsantas, 2012;

Kauffman, 2004). In accordance with this broader research base, SRL has also been identified as a critical skill to be successful in MOOC environments (Littlejohn et al., 2016; Milligan et al., 2013). In the following section, a vignette (based on actual student responses in a MOOC) will present a common learning challenge in MOOCs, which illustrates why SRL could be helpful in supporting students therein.

> *Hi, everyone. My name is Scott Rayner. (Pseudonym). I'm interested in this course as a means of furthering my professional development. I work as a lead architect at a small software company, and would like to be more capable especially as the field progresses into the future.*
> *I'm the father of two young boys under 4. Regrettably, this will probably reduce the amount of time I have available to spend on this course, but ces la vie.*
> *I'm hoping that this course will enable me to be more useful and productive for the people I build solutions for, and that I can finish the course work without sacrificing these precious years with my boys.*

Scott's story exemplifies a major obstacle that most MOOC learners encounter: finding time to study while juggling between other life responsibilities and work commitments (Kizilcec & Halawa, 2015). When free and open-source technologies provide thousands of accredited online courses, learning does not magically happen. Instead, Zimmerman's SRL theory (2000) claims that a learner has to be engaged, persistent, and able to identify learning resources and regulate their learnings to be able to achieve his or her learning goals. Scott would potentially benefit from having tools in place to help him set goals, track his progress, and engage in other reflective practices. This anecdote encapsulates the overall challenges of using MOOCs as a vehicle to enhance one's professional education. It also points out the need for various supports and scaffolds in MOOC environments.

Rationale of the Current study

Previous literature provides a theoretical basis to understand the potential affordance of SRL in learning persistence, motivation and learning performance (Azevedo & Hadwin, 2005; Newman, 2002; Perry & Winne, 2006). However, there is not sufficient empirical evidence to support whether and how SRL interventions can lead to more effective learning in MOOC environments. Another challenge of reviewing and comparing MOOC-based research is that MOOCs attract a wide range of learners with various backgrounds, characteristics, and motivations to register in a course. When coupled with diverse platform setups, pedagogies, and learner experiences, it complicates MOOC data collection, analysis, and interpretation.

For example, past experimental studies have used relatively brief pre-course interventions to prime students' self-regulation (Kizilcec et al., 2017a; Reich, 2014) to increase student's completion rate. These studies featured inexpensive installments and short-time interventions; however, they also showed limited effects when replicated with more participants (Kizilcec, 2020). Given the gap in the literature, more experimental research is needed to explore the educational affordance of self-regulated learning tools in MOOCs.

Applying with a personalized, longitudinal self-regulated learning interface (SRLUI), this study examined how learners interact with tools designed to support learners' SRL activities. This study proceeded with an assumption that learners might not participate in SRL activities which did not count towards the grades in an online environment. The historical data also suggested that learners' compliance rates with SRL tools were low (Davis et al., 2018; Jansen et al., 2020). What was under investigation here was what kind of instructional design of SRL artifacts can better support learners? The design of the treatment group provided learners opportunities to engage in SRL activities aligned with a full cycle of Zimmerman's (2000) SRL model from goal setting, task planning to self-reflection. In contrast, the control group followed with conventional, non-interactive SRL supports in a MOOC platform. Therefore, the first research question is to explore students' usage of SRLUI.

RQ1: Do learners interact with tools (SRLUI) that are designed to support their self-regulated learning strategies? If so, how do learners use SRLUI?

In the next step, this study explored the effectiveness of access to SRLUI on learner persistence and learning outcomes. The research questions are:

RQ2: To what extent does the usage of SRLUI have an effect on learner persistence?

RQ3: To what extent does the usage of SRLUI have an effect on learning outcomes?

To answer the first research question, a variety of descriptive techniques were applied to understand students' participation in SRLUI. Additionally, a hierarchical analysis with heatmap was used to visualize learner behavioral data to reveal any potential patterns or subgroups of learners in conjunction with their usage of SRLUI.

Survival analysis and multi-level regression were used to model the key relationships between access to SRLUI and the learning outcomes for 2nd and 3rd research questions respectively.

Summary of Chapters and Following Structure

This thesis is presented in five chapters. Chapter I provides an overview of the research background, the problem statement, and research questions. It also provides a roadmap of this dissertation and summarizes the findings. Chapter II is a targeted review of pertinent literature, which provides the conceptual foundations for the research design and the analysis of self-regulated learning (SRL) for this study. There are extensive SRL theory overviews and a discussion of best practices in support of SRL development and challenges in measuring SRL behaviors. An in-depth analysis of several SRL implications in MOOCs emerges the opportunities and unanswered questions for future studies to fulfill. Chapter III describes the research design and the quantitative models employed in this study. Relevant research studies are also provided to justify the decision-making process in choosing certain methods. Chapter IV reports the results of the analysis. It begins with a preliminary description of the sample and addresses the questions in sequential order. This study applies quantitative methods in analyzing three research questions, with some descriptive analysis of learner's behaviors on using SRLUI for the sub-questions under RQ1. Chapter V provides extensive discussions of the results, implications, limitations, and suggestions for future, related studies.

This study presents an opportunity to design and implement an SRL artifact in support of learners' SRL activities on a MOOC platform. Additionally, this study provides empirical evidence that learners are willing to use SRL tools and such tools could help a subgroup of learners achieve higher grades in MOOCs.

Definitions of Terms

- **MOOC:** Massive Open Online Course is an online course with unlimited enrollment and open access via the internet.
- **credential-based MOOCs:** It is a type of MOOC providing a certificate for a learner as proof for completion. Usually, it requires learners to complete a set of graded assessments, achieve a certain level of grades, and complete identity verification.
- **Instructor-paced MOOC:** It describes a type of MOOC with a specific course start and end dates. In this study, all data were collected from instructor-paced MOOCs.
- **Self-Regulated Learning:** A series of actions directed to acquire information or skill which include but are not limited to goal setting, environmental structuring, self-rewarding or self-punishment, and self-evaluation.
- **User interface:** It refers to the web pages in which learners interact on a MOOC platform.
- **Trace data:** A learner's log data on the learning management system. The log data includes but is not limited to timestamps, duration, and the pages learners visit on the course, and the pages a learner clicks on.
- **Learner persistence:** It describes a range of times a learner remains active during a course period. For example, a MOOC opens for a total of 14 weeks, which is 98 days. Thus, learner persistence could range from 1 - 98 days.
- **Learner dropout:** In general, it occurs when a learner leaves a course before a course ends. In this study, since the verified track learners are required to take the

final proctored exam in order to earn a certificate, learners who leave the course before the final proctored exam are considered dropouts.

II – LITERATURE REVIEW

Chapter Overview

Online learning, once heralded as the great democratizer of education, has long been criticized for its low completion rates (Reich, 2020). This study aims to combat some of the issues with online learning through a newly designed self-regulated learning (SRL) tool, particularly in massive open online course (MOOC) settings. In this section, I will provide a background overview of online learning prior to the COVID outbreak, which took place in March of 2020 and impacted higher education universally. I will also review the development of massive open online courses (MOOCs) and how they branched off from online learning. I will further discuss research trends and issues from the literature on self-regulated learning. Later, I will propose future research based on the guidance and recommendations of SRL theories and SRL empirical research findings.

History of Online Learning

Online learning which was also known as distant learning, can be traced back to the 1800s when students would receive course materials by mail (Spector et al., 2014). As production, distribution, and communication technologies evolved, the methods of distance learning changed in concert (Nipper, 1989). This systematic change in distance learning modalities can be broadly categorized into three phases: correspondence, during which most learning was done asynchronously via mail; multimedia, during which multiple modes of communication such as radio, television, and film were used to deliver content; and computer-mediated, during which the widespread availability and use of home computers allowed for synchronous and asynchronous delivery of content and communication (Sumner, 2000).

The transition to new dominant phases of distance learning did not completely stop the modalities of the previous phases, however. For example, mail-based correspondence courses persisted into the late 20th century, and universities often employ distance education methods that blend aspects of all three phases (Harting & Erthal, 2005). That being said, the current phase of distance learning is dominated by computer-based online learning, and continues to change as more powerful computers and high-speed internet access become more ubiquitous (Casey, 2008).

Online Learning Development

Online learning started to emerge in the1980s at the Western Behavioral Sciences Institute in La Jolla, California, where they began delivering executive business programs through computer conferencing (Harasim, 1993). Ever since, online education has been a growing segment of higher education. For example, the University of Phoenix began offering degree programs entirely online in 1989. Columbia University in the city of New York established its fully online Engineering School for graduate-level programs in 1986 (Columbia Video Network, n.d.). New York University, Western Governors University, and California Virtual University were established in 1998 (Miller et al., 2013). With the continuous growth of online education, 35% of students (at both the undergraduate and graduate levels) took at least one distance learning class in the fall of 2018 (National Center for Education Statistics, 2019), and 16.3% of students (3.2 million students) enrolled in only distance education classes during the fall of 2018.

Online learning has been widely adopted because it offers learning opportunities without geographic limits or constraints regarding the timely completion of course materials (Ally, 2008). This is especially suitable for learners who have full-time jobs, family responsibilities, or other social obligations (Simmons, 2002). Online education also allows teachers to reach more learners, resulting in economic savings for both schools and students (Aksal, 2009). However, students in online programs are reported to have lower grades and higher dropout rates than their cohorts in the residential programs (Almeda et al., 2018). This could be resulted from learning challenges such as students feeling isolated or less connected with their cohorts (Carr, 2000; Russo, 2005). The online learning challenges will be further discussed in a later section. The following section will discuss the learning theories, pedagogies, and the social aspects of online learning.

Online Learning Theory

Professor Ally from the University of Athabasca University, the Open University in Canada, once defined online learning as:

> [t]he use of the Internet to access learning materials; to interact with the content, instructor, and other learners; and to obtain support during the learning process, in order to acquire knowledge, to construct personal meaning, and to grow from the learning experience. (Ally, 2002, p. 7.)

In terms of the epistemology of online learning, Ally (2004) argues that online learning is influenced by three schools of thought: behaviorist, cognitivist, and constructivist theories. Ally explains that behaviorist strategies are used to teach content, cognitivist strategies influence the design process, and constructivist strategies inform the learning activities design.

In addition, designers of online learning often refer to Mayer's (2005) multimedia learning theory, which is originally focusing on effective web-based multimedia learning design. Heavily influenced by cognitive learning theory, Mayer investigates how to build appropriate mental representations to avoid cognitive overload. Mayer, Dow, and Mayer (2003) further develop three principles: the modality principle, the interactivity principle, and the self-explanation principle to better organize and deliver learning through texts, videos, photos, illustrations, and animations. Consequently, Mayer's multimedia learning theories become fundamental design principles for online learning (Ayres, 2015).

Online Learning and the Social Dynamic

Online learning also considers learning with peers. Garrison, Anderson, and Archer (2001) argue that online discussion is an important component of online learning experiences because knowledge-building occurs in a community through discourse (Scardamalia & Bereiter, 1996). Thus, online learning pedagogy relies on supporting classroom dialogue (Laurillard, 2002). Meyer (2002, 2003) claims that the socio-cognitive dynamics of online discourse not only help learners gain a thorough understanding of the content itself but that they also help students acquire critical thinking and inquiry skills. However, later empirical research reports that providing a discussion forum could not guarantee effective learning (Guo et al., 2014; Meyer, 2003; Smith, Ferguson & Caris, 2001). Through their research on professional development for teachers, Guo and his colleagues (2014) conclude that instructional feedback (i.e., asking thought-provoking questions, giving encouragement, and facilitating reflective discussion) positively enhances learners' cognitive engagement. An et al.'s (2009) study on instructors' participation in discussion forums suggests that students rarely voluntarily participate in any given discussion forum. When students are required to participate and respond to others' posts in the discussion forum, engagement does not rise above the minimum required posts. Contrary to intuition, this study finds that minimal instructor participation could allow students to express their opinion more freely (An et al.).

Taking these studies into consideration, we can conclude that the formation of an online learning community does not happen spontaneously or naturally, but can only arise through purposeful instructional design and facilitation from instructors (i.e., providing feedback). A well-designed online learning environment and a well-trained faculty are key to building a collaborative and effective online learning community.

Online Learning Technology Development and Delivery

An online learning program's design varies depending on the learning environment, target audience, learning objectives, and accessibility of the course. In terms of learning sequence and time constraints, online courses could be categorized as self-paced or instructor-paced (Moore et al., 2010). Most online courses offered by the residential programs are instructor-paced, which means instructors decide the learning sequence and pace for all learners; while a self-paced course has more flexibility for learner to decide when to start (Rhode, 2009).

In terms of the learning environments, there are various platforms such as learning management system (LMS), course management system (CMS), virtual learning environment (VLE) and knowledge management system (KMS) (Khan, 2001; Nichols, 2003). LMS, CMS, VLE and KMS had been used interchangeably over the past 10 years

(Asunka, 2008; Zhang & Kenny, 2010). Since 2012, LMS has been widely deployed in higher education for the residential programs or fully online programs (Reich, 2020). As for massive open online course (MOOC) providers such as Coursera, edX and Udacity, they utilize LMS to arrange and deliver courses. The next section will discuss MOOCs background and how it was developed in the first place.

MOOC Overview

In 2011, Stanford University offered three open online courses with the intent to democratize education by delivering free education to the public. Up until then, online classes maintained a traditional teacher to student ratio which limited the medium's potential for scalability. Stanford's public online courses became a huge success, with each course attracting more than 100,000 learners. This marked the first offering of public and scalable instructor-directed Massive Open Online Courses (MOOC) (Ng & Widom, 2014). These three courses were a Databases (DB) class, taught by Professor Jennifer Widom, the Chair of Computer Science Department at Stanford School of Engineering; a Machine learning (ML) course taught by Professor Andrew Ng, and an Artificial Intelligence (AI) course offered by Professors Sebastian Thrun and Peter Norvig. The first two courses were delivered on a platform developed by Professor Ng and his students, which later became known as Coursera. The AI course was taught on Udacity which was created by professor Thrun and is still operating today.

These three Stanford MOOCs, which were referenced as DB-, ML- and AI-courses, had a significant impact on the development of MOOCs. These courses featured short clips, fast-forward playback, auto-graded programming, and a discussion forum which allowed learners to post and answer questions (Ng & Widom, 2014). In addition, the ML and DB courses adopted a “mastery learning” approach which allowed students to have multiple attempts to do their assignments until they got it right. Upon completion of the course, students would receive “Statements of Accomplishment” issued by Stanford University (Ng & Widom, 2014). Although these courses captured the world’s attention in 2011, it seemed that neither Ng, Thrun nor Widom were aware of the concept of MOOCs, nor “connectivist MOOCs” (cMOOCs), which was initiated by Professor George Siemens and Stephdown Downes (Siemens, 2005). Consequently, a new term xMOOC has been proposed to distinguish them from cMOOCs (Ng & Widom, 2014).

MOOC Learning Theory

The majority of modern MOOCs aim to deliver highly scalable and accessible education. Influenced by Professor Daphne Koller and John Mitchell, who promoted “flipped classrooms” to better engage on-campus students, many MOOC platforms offer similar setups like short videos, fast-forward playback, subtitles, in-video quizzes, multiple-choice quizzes, auto-graded programming assignments, and discussion forums (Ng & Widom, 2014; Reich, 2020).

In terms of pedagogical design, most MOOCs adopt a student-content interaction framework (Miyazou & Anderson, 2013) where there is little teacher-student or student-student interaction. Ultimately, students in MOOCs are expected to learn by watching instructional videos and taking machine-graded assessments. This concept of learning through observation and repetitive action is advocated by schools of behaviorist thought such as those initiated by Skinner (1974). However, this model has been criticized for its failure to consider how emotional and psychological factors can influence learning (Reich, 2020).

Traditional classrooms are highly conducive to group activities since teachers and students can have in-person interaction and feedback. In contrast, the majority of learning activities in MOOCs are designed to be completed individually. This is caused by two main factors: the functionality of MOOC platforms and grading rubrics. In a MOOC setting, group projects are challenging to administer given the high dropout rate. In addition, grading becomes complicated because it is hard for instructors to evaluate individual student's efforts and contributions to a group project. The constraints of an entirely online learning environment often push course designers to convert the majority of learning assessments and projects in MOOCs into individual work for students to complete by themselves.

After this overview of the history and learning theory behind MOOCs, the following section will illustrate the educational challenges presented by MOOCs, as identified in prior literature.

MOOC Learning Challenges

Due to the massive number of enrollments, the lack of instructional presence, and the lack of personal feedback, students in MOOCs often struggle (Al-Freih, 2017), leading to large numbers of students dropping out (Hew & Cheung, 2014). In addition, allocating time for study is a challenge due to work or life responsibilities that may take priority (Kizilcec & Halawa, 2015). Research also suggests that high dropout rates in MOOCs are caused by a lack of self-efficacy, self-regulation, and self-motivation (Daniel, 2012; Koller et al., 2013, Lewin, 2013). Yousef (2014) proposes that learning analytics and learning assessment activities can be used to provide feedback to learners about their progress and promote self-awareness, self-confidence, and self-reflection in the MOOC environment.

According to a 2019 Class Central report, MOOCs have attracted over 110 million enrollments in more than 135,000 courses, including courses for 820 Micro-credential programs and 50 MOOC-based degrees (Shah, 2019, December). Despite this rapid expansion, many educators and researchers are skeptical of the effectiveness of MOOCs for the following reasons:

High dropout rate. The overall dropout rate for MOOCs was up to 90% or more (Cagiltary et al., 2020; McAuley et al., 2010). Completing a MOOC requires a learner to be highly motivated and disciplined, making time to study and complete assignments in an environment with little supervision and no real consequences. Learners often have to juggle work or personal responsibilities, which adds to the challenges of completing the course (Pheatt, 2017).

Lack of personalized feedback. Support via personalized feedback is often desired by students because it provides specific solutions to learners' questions and improves their comprehension (Saunders, 2018). However, due to the "massive" nature of enrollment in MOOCs, there is usually a highly unbalanced ratio of students to instructors and staff, making it difficult or impossible to provide personalized, synchronous feedback to students (Daradoumis et al., 2013; Yousef et al., 2014). Administratively, course staff use text-based discussion forums to address students en masse to manage the large volume of inquiries, and they may not be able to engage in individual conversations with learners.

Limited collaboration and social interaction with peers. Unlike traditional classrooms that rely on live dialogues between teachers and students, MOOCs operate primarily through text-based discussion forums (Onah, Sinclair & Boyatt, 2014). However, since students are not required to participate in the forums, only a small portion of students post questions, share learning tips, or answer other students' questions. The high attrition rates further decrease the pool of learners who might participate in the forums. With the open enrollment format, students can start the coursework at different times, making it harder for learners at different stages to engage in dynamic and spontaneous conversations with each other. Additionally, graded assessments are usually designed as individual tasks, which further limits opportunities for peer interaction and collaboration (Gamage, Fernando, & Perera, 2015).

One-way delivery of information. A MOOC usually features a guided curriculum with pre-recorded video lectures as the main learning task. Koedinger et al. (2015) argue that passive learning by watching lecture videos or reading texts could be the reason resulting in students dropping out. Particularly, when students have a question during a lecture, they do not have the means to ask questions or engage in a synchronous and spontaneous discussion like in a traditional classroom. Instead, a learner must simply watch the video repeatedly, hoping they can eventually figure out the concepts. Alternatively, students may reach out on the discussion forum to have their questions addressed. However, the response to their question may not be posted for another couple of hours to a few days, whether from the course staff or classmates.

Learner Characteristics and Psychological Challenges

Learner characteristics, such as a student's demographics, prerequisite knowledge, motivations, and other psychological characteristics, are found to be associated with MOOC learning outcomes (Chen et al., 2020; DeBoer et al., 2013; Gardener & Brooks, 2018; Kizilcec & Halawa, 2015). The following section will discuss two major challenges that are influenced by learners' characteristics: lack of prior knowledge and lack of study time.

Lack of prior knowledge. Lack of prior knowledge contributes to higher attrition rates and is found to be negatively associated with learners' grades in MOOCs (Coffrin et al., 2014; DeBoer et al., 2013; Gardener and Brooks, 2018; Kizilcec & Halawa, 2015). Unlike traditional school systems where there are admissions reviews to ensure admitted students have reached a certain academic threshold, MOOCs are open to anyone who has internet access (Pheatt, 2017). Although MOOCs present an opportunity for everyone to learn, learners who do not have the prerequisite knowledge often show a lower comprehension of the materials (Bransford & Johnson, 1972). That disadvantage, coupled with the scarcity of learning resources in MOOCs, makes it more difficult for learners to keep up with the course's pace. This often leads students to feel demotivated, and eventually dropping out (Chen et al., 2020).

Lack of time for study. Having no time to study is reported as a major hurdle for learners in MOOCs (Kizilcec & Halawa, 2015). Kizilcec & Halawa (2015) infer that learners who report not having enough time are hindered by "low volitional control" (p. 65). Terras and Ramsay (2014, 2015) also discuss the temporal experience of online learning and explain how it affects the psychological aspect. The issue with time and online learning lies in one’s perception of the difference between how physical (objective) time is spent versus how psychological (subjective) time is perceived (Terras & Ramsay, 2014). For example, “Not having enough time” could be a result of underestimating or overestimating the time required for a task. Another time management issue lies in the relationship between time and memory, namely “how long an event is perceived as lasting” (Terras & Ramsay, 2014, p. 111). In other words, “I don’t have time for studying” could be an excuse much like the common excuse “I don’t have time for gyms.”

Teras and Ramsay (2014) also argue that if a learner believes they do not have time to study, it could be due to an actual physical time limitation or due to a learner’s warped perception of their own schedule. The actual problem could be that the learner struggles with scheduling or tackling coursework efficiently, which results in them thinking, “I don’t have time for studying.”

In summary, time management has been reported as a major learning challenge in MOOCs. Particularly, it could be resulting from a student’s inability to complete tasks efficiently or a student’s inability to accurately estimate the time they need to study. When this issue is coupled with the “massive” nature of student enrollment, disproportionate student to faculty ratio, a low barrier-to-entry open enrollment option, and the diversity of learner backgrounds, that leaves a lot of responsibilities for learners. The nature of these challenges suggests that more learning support is needed to increase the effectiveness of the increasingly ubiquitous MOOC learning environment.

Self-Regulated Learning

Given the aforementioned learning challenges presented by MOOCs, learners have to be highly autonomous and exercise self-regulated learning strategies to achieve their learning goals. A self-regulated learner is a person who actively sets goals, stays motivated, and wisely manages their intellectual capacity as well as their mental/emotional health (Pintrich, 2000; Zimmerman, 2002). Zimmerman (2000) hypothesizes self-regulated learning (SRL) as a cyclical model in which learners attain their learning goals by repeating three phases of learning: forethought, performance, and self-reflection. In the forethought phase, a learner sets learning goals and conducts strategic planning. In the performance phase, a learner utilizes time management, help-seeking, and environmental structuring to execute learning tasks. In the self-reflection phase, a learner processes their performance and adjusts their strategies in order to attain the learning goals.

Zimmerman conducts several empirical studies to inspect self-regulatory development processes and the validity of the cyclical phase model in the context of academic training (Kitsantas, Zimmerman & Cleary, 2000) and athletic skills (Zimmerman & Kitsantas, 2002). These studies yield the following conclusions: 1. Goal setting is key to goal attainment 2. SRL is more than a concept or attitude, it is a set of "mentally and physically demanding activities" (Zimmerman, 2013, p141) which require time and guidance to develop. 3. SRL training is more effective when the three phases are taught all together, instead of focusing on one or two parts of the SRL model at a time. 4. Finally, Zimmerman (2013) also recommends the usage of computer-mediated environments to scaffold SRL development and allow learners to receive immediate feedback.

Recent experimental studies have found that MOOC learning performance is positively associated with a student’s SRL skills. (Al-Freih, 2017; Kizilcec et al., 2016; Reich, 2014). A considerable amount of studies have explored SRL applications and their influence on learning outcomes in MOOCs (Jansen et al., 2020). Many researchers have attempted to design interventions around SRL application in MOOCs; however, not many of them produce significant results (Borrella et al., 2019; Davis et al., 2018; Kizilcec et al., 2016). The following section explores a few well-cited SRL applications in MOOCs and discusses any potential opportunities and concerns yet to be addressed.

Examples of SRL Applications in MOOCs

Kizilcec, Perez-Sanagustin and Maldonado (2017b) conduct an observational study using a pre-course survey to investigate learners’ SRL strategies associated with their learning behaviors, outcomes, and characteristics. Using correlation analysis and logistic regression, the study suggests students who achieve their learning goals are much more likely to employ goal-setting strategies (Kizilcec et al., 2017b).

Following the recommendations of Kizilcec et al. (2017b), a few research studies also explore the effect of SRL intervention on goal setting. Yeomans and Reich (2017) utilize a pre-course survey to prompt learners to set goals and explore its effect on learner completion rates and the number of certificates purchased. Their results suggest long-term goal planning could increase completion rates; although this positive association is most apparent with a subgroup of students those who are affiliated with schools. Similarly, Kizilcec, Davis, and Cohen (2017a) also utilize a pre-course survey prompting learners to consider values generated by completing the course. The study is a randomized experiment and the value affirmation process is found to improve grades, persistence, and completion rates among a very specific subgroup of learners - lower class men (Kizilcec et al., 2017a).

Kizilcec et al. (2020) later replicate and scale up their prior MOOC studies focusing on long-term planning and value-relevance affirmation. However, there are no significant findings in either the long-term planning or the value-relevance interventions. They conclude that the effects of one-time SRL interventions in MOOCs are short-lived, and continuous support is needed to facilitate behavioral change. They also suggest that future research should consider integrating content-specific information in the artifact (Kizilcec et al., 2020).

The aforementioned studies (Kizilcec et al., 2017a; Kizilcec et al., 2020; Yeomans and Reich, 2017) implement randomized experimental designs to explore SRL applications in MOOC learning. There are three important things to note about the design of these studies: First of all, these studies only focus on promoting one SRL strategy, such as goal setting or value-affirmation, instead of incorporating the entire SRL process. Secondly, all the interventions are one-time, short treatment interventions (10-15 minutes). And finally, they all use self-report surveys to make inference of students' SRL abilities or their usage of SRL activities in the course.

When reviewing the design of the studies mentioned above, it is clear they do not take into account the findings from previous experimental studies and SRL models. The SRL models from Zimmerman (2000), Pintrich (1999), and Winne and Hadwin (1998) all describe SRL as a multi-phases procedure intertwined with cognitive, social, metacognitive, and behavioral domains. When Zimmerman and his colleagues design experimental research to verify his theory (Kitsantas, Zimmerman & Cleary, 2000), they find that learners receiving all three of the cyclical phases of SRL training are able to be more adaptive than those who receive partial phase training. Specifically, Zimmerman (2000) argues SRL skills require time and facilitation to be developed if they are to become a useful tool for students. This could be the reason why the replicated and scaled SRL application interventions (Kizilcec et al., 2020) do not produce any significant findings.

Additionally, self-reported surveys have been criticized for being an inaccurate way of recording a learner's actual SRL behaviors. Winne (2013) further warns that surveys and interventions around SRL strategies are often unreliable because SRL is more than a static aptitude, motivation, or perception. He emphasizes that learners have the agency to control and change their actions; thus, it is not sufficient to use historical behavior to predict a learner's future behavior in terms of exercising SRL strategies (Winne, 2013). Winne (2017) proposes analyzing learner trace data (or clickstreams) to measure SRL activities. However, there are still challenges regarding the interpretation of these ambient data on how learners acquire and process learning content (Winne et al., 2019). Specifically, Winne and his colleagues (2019) created nStudy, which is a web-based software, to gather learner's data on their cognition, metacognition, and motivation processes; as it also could support learner's ongoing self-regulated learning behaviors.

Following the same trend of thought, Jansen et al. (2020) conduct an experimental study using instructional videos to educate learners on the concepts of SRL and surveyed learners' perceptions of SRL strategies in MOOCs. In terms of measurement, Jansen and her colleagues (2020) utilize both trace data and survey responses to analyze learners' SRL abilities change and course completion. Compared to prior SRL studies which treated SRL as behavioral science, using one-time-only, short-term intervention, the design of Jansen et al.'s (2020) study provides more comprehensive SRL supports interventions. However, due to a high dropout rate, that leaves only a small amount of learners interacting with the artifacts, the results show a trend of increase in the completion rate, but no improvement is found in learners' SRL skills (Jansen et al., 2020). This study has a high ecological validity, since it integrated the SRL supports within the course in a MOOC environment (Jansen et al., 2020). However, this study leaves more questions that need to be answered, such as how to go about improving learner compliance rates, and whether SRL interventions affect learning in any other ways. Additionally, Jansen et al. (2020) also raise concerns that trace data don't provide clear and direct evidence on how learners process content which leaves a design challenge on how to capture learner's SRL activities.

The prior research findings inform this study to consider developing SRL support in the full cycle of SRL phases instead of partial support. In addition, to achieve high ecological validity, a SRL intervention should be embedded in the course with content related features.

Measures of Learner Behavior and Persistence

Categorizing learners into subgroups of learning profiles. Several MOOC studies attempt to explore salient engagement patterns and categorize learners into subgroups to better understand the learning dynamics among the diverse student profiles (Kizilcec et al., 2015; Khalil & Ebner, 2017). Khalil and Ebner utilize trace data such as frequency of video views, participation in the discussion forum, and quiz attempts to analyze learners' behavior patterns. They sort students into four groups, which are gamer students[1], perfect students, social students, and dropouts. The overall goal of the study is to improve student engagement so that learners would watch all the lecture videos and complete assignments as "perfect students" instead of "gamers" (Khalil & Ebner). It is worth noting that the course included in the study is a credit-bearing MOOC designed to be a required course for the students to fulfill their curriculum requirements at Graz University.

Cluster analysis is a common approach in multiple disciplines that allows researchers to categorize subjects based on their shared traits (Bowers, 2010b; Romesburg, 1984). The biggest challenge in employing cluster analysis is that different clustering methods can result in different numbers of clusters. In structured cluster analysis, researchers usually make certain assumptions about the character of the groups, while in unstructured cluster analysis, the groups are determined by the structure of the data itself (Bowers, 2007).

Particularly, hierarchical cluster analysis (HCA) is a type of unstructured cluster analysis that enables researchers to explore complex data matrices from a bottom-up approach. Hierarchical cluster analysis (HCA) could be valuable in studying novel topics and providing descriptive analysis of underlying patterns (Hawn, 2019). For example, Jorion et al. (2020) use HCA to analyze a complex and multidimensional dataset on gaming behaviors based on log data captured through touch events. The study collects data from a museum exhibit, *Oztoc*, which is a collaborative, interactive game table

---

[1] Gamers are students who skipped lecture videos and only completed assignments. They are students who did minimal coursework and still passed the course.

allowing up to 4 players at a time. Their results identify three groups of gaming behaviors. When combined with heatmap visualization, they are able to graph individual player trajectories and their associated outcomes (i.e. the number of circuits completed or numbers of players) (Jorion et al., 2020). In this study, in order to explore how learners interact with the artifact, this study also applied HCA to understand student-level learning activities, and report learner's interaction with SRLUI based on the clustering results.

Visualizing learner data. Combining HCA with heatmap visualization is a powerful way of displaying patterns and associated outcomes within and across individual cases at a granular level (Bower, 2010; Jorion et al., 2020; Lee et al., 2016). Below are two examples of HCA analysis with heatmap visualization. Figure 2.1 is a template with an explanation of relevant indices, and Figure 2.2 is a partial image of a clustergram showing the longitudinal summary of learners' online learning behaviors. There is another study attempting to visualize learner behaviors in MOOCs. Coffrin et al. (2014) conduct an exploratory analysis and create a state transition diagram, using HTML and D3, to indicate learners' entry point, exit point, and transitions between tasks (see Figure 2.3).

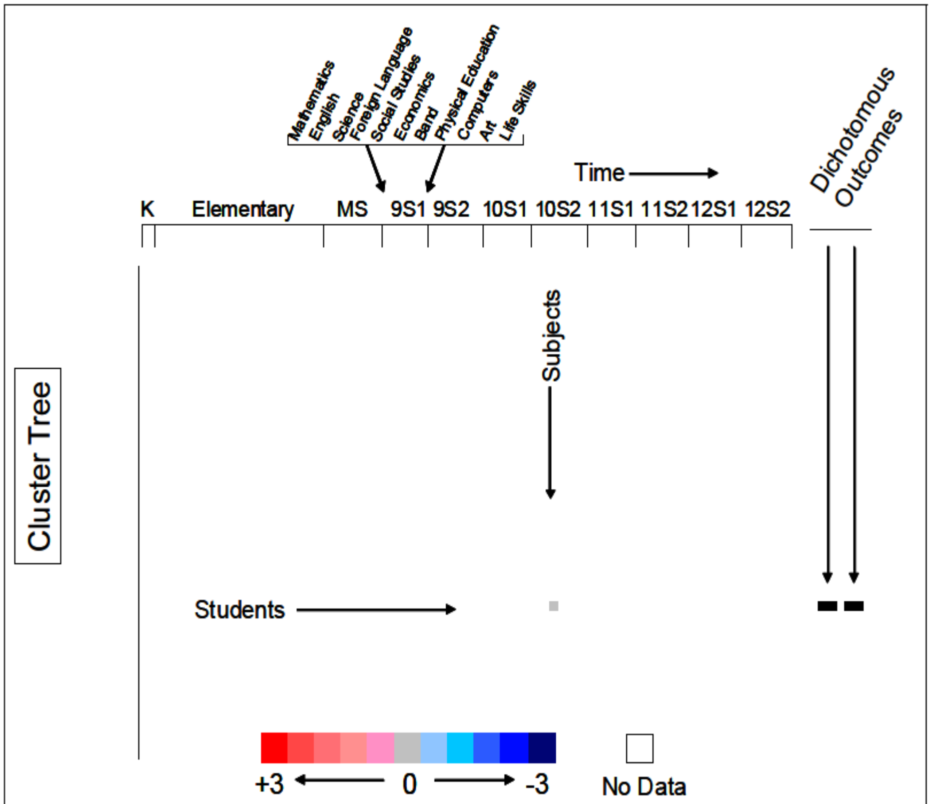


Figure 2: Clustergram Template.

Figure 2.1. *Clustergram Template*

(permission is acquired from the author to display the figure)

Learners are listed in rows and each block represents a subject from Kindergarten to K12. Each block is color-coded, and standardized using z-score. If the grade is on average, the block is grey. Dark red represents the highest grade and dark blue is the lowest. On the right-side of the chart are annotations, which show associated learning outcomes such as dropout or earning passing grades. In annotation, black color means “Yes”, and white means “No”. (Bowers, A. J., 2010, p. 7)

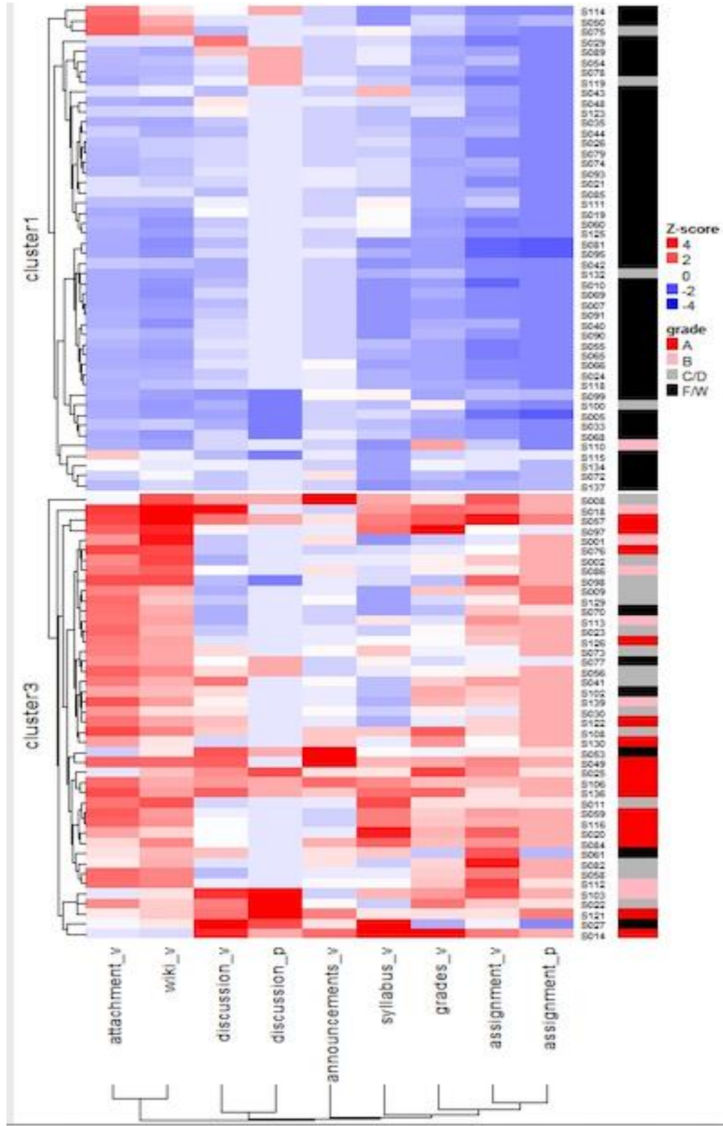


Figure 2.2. *Hierarchical Analysis and Heatmap on Students' Behavior Data*

(permission is acquired from the author to display the figure)

This figure showcased a section of the clustergram using longitudinal semester-long data to analyze students' behavior in the LMS system (for full image: goo.gl/Y7VFHJ). There are a total of 3 clusters of learner behaviors identified based on nine indices: number of views in attachment, wiki, discussion forum, announcements, syllabus, grades, assignments, participation in the discussion forum, and assignments.

(Lee et al., 2016, p. 604)

The limitation of the state transition diagram is that the visualization is based on an aggregated result, so it is not easy to discern an individual student's learning trajectory. In addition, one figure is only allowed to show one subgroup at a time. To compare multiple subgroups' learning behavior patterns, one has to generate multiple figures, which is not efficient or effective for comparison.

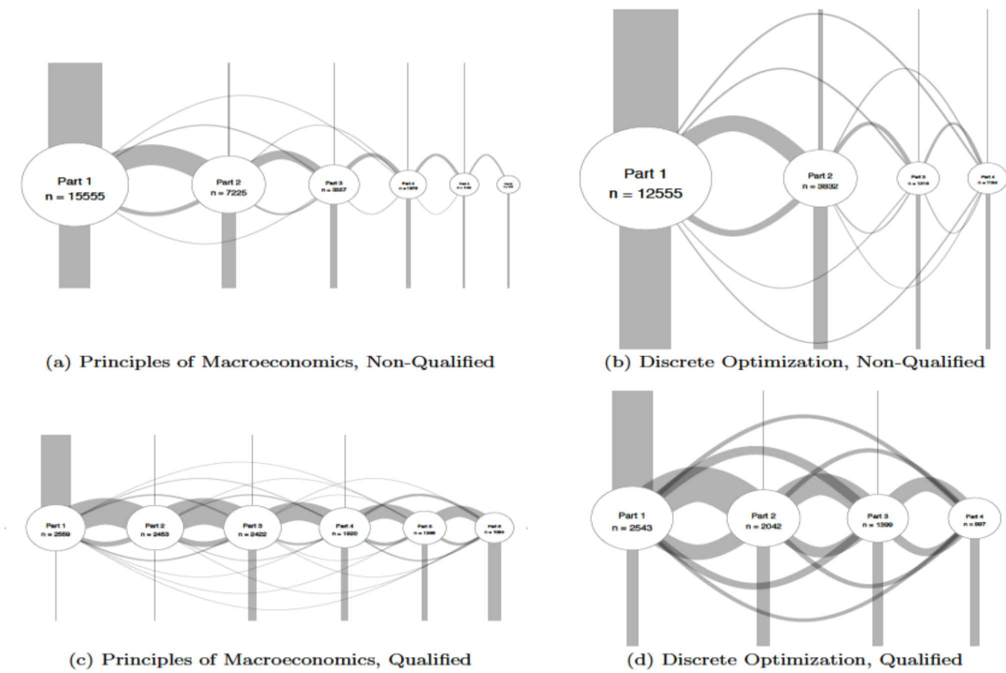


Figure 2.3. *State Transition Diagrams*

(permission is acquired from the author to display the figure)

The vertical lines represent the number of students accessing the content modules (on top of the circle) or leaving the course (below the circle). The upper curved lines represent learners' moving onto the next part of the content; the lower curved lines represent learners going back to review the previous section. (Coffrin et al., 2014, p 90).

Based on these examples, this study will employ a hierarchical cluster analysis heatmap approach to analyze learner behavioral data to identify learner subgroups based on their learning trajectories and their usage of the self-regulated learning user interface (SRLUI) created in the study.

Learner persistence and dropout

The pilot study (Hsu, 2020) informs us that learner persistence has a strong correlation with learning outcomes in MOOCs. To advance a more analytical approach towards learner dropout, Willet and Singer (1991) propose using survival functions and discrete-time hazard models based on longitudinal educational data to analyze learner dropout for the following reasons:

1. Survival functions do not need special software, just standard statistical packages.
2. Survival functions allow researchers to answer questions such as if a treatment reduces the number of dropouts.
3. Unlike traditional methods, survival functions and discrete-time hazard modeling can include more predictors (i.e. time-varying, time-invariant) or interactions between predictors in the model.
4. Each person could be censored at a particular time during the study.

Bowers (2010) confirms that survival analysis and discrete-time hazard analysis (using logistic regression with the person period dataset) are superior in analyzing longitudinal data when compared to traditional methods. In his study, Bowers (2010) is able to appropriately control time constant and time-varying predictors to investigate the utility of teacher-assigned grades as a predictor of student dropout. He is able to conclude that teacher-assigned grades are significant in predicting student dropout in any given subgroup (Bowers, 2010a).

Chen et al. (2020) also utilize survival analysis to explore the effect of scientific misconceptions on student persistence in MOOCs. Chen and his colleagues (2020) include time-constant variables (i.e. gender, age, education, English skills), time-varying predictors (scores in the previous milestone), and interactions between predictors in the model to predict student dropout.

To facilitate this research on learner persistence and dropout, this study will utilize survival function to explore the efficacy of SRL interventions in supporting learner persistence and reducing dropout rates.

Research Design

Prior literature has found that the major drawbacks of MOOC learning environments are a low completion rate, lack of personalized feedback, limited collaboration and social interaction with peers, and a one-way delivery system of information. Initial MOOC research focuses on learner characteristics (such as gender, educational level, and prior knowledge) and their influence on learning outcomes (Reich, 2014; Kizilcec & Halawa, 2015). More recent research explores the influence of internal motivational constructs (i.e. motivation, interest, goal orientation) and other psychological challenges (i.e. time management and volition control) (Kizilcec et al., 2015; Terras and Ramsay, 2014) on learning outcomes in MOOCs.

Previous SRL application studies in MOOCs provide valuable guidelines and suggestions in terms of research design and measurement for this proposed study. For example, Davis et al. (2018) recommend that using an experimental design could potentially help show a causal relationship between SRL applications and learning outcomes. Davis and his colleagues (2018) also suggest that SRL components be designed as compulsory activities, otherwise students would not necessarily participate. In addition, SRL interventions should not be one-time events (Jansen et al., 2020; Kizilcec et al., 2020), rather they should be treated as skills which need to be facilitated and developed over time (Borrella et al., 2019). Schraw (2009) also suggests using multiple measurements of learning outcomes, not only to avoid bias but to be able to discern inter-relationship.

Although SRL strategies and their applications have been widely discussed in classroom-based literature, there are only a few studies with empirical evidence investigating the effects of SRL activities on learning outcomes in a MOOC environment (Cobos & Ruiz-Garcia, 2020; Jansen et al., 2020; Kizilcec et al., 2017a). Additionally, these prior studies have limitations and challenges in generalizing their findings due to (a) low compliance rates, (b) using indirect measures of learner SRL activities, (c) only one or a few courses included in the sample data. This study is designed to provide empirical evidence with longitudinal SRL intervention design in MOOCs. Unlike past studies which only implement partial, one-time SRL interventions, this study will utilize an online, continuous SRL support embedded throughout the online learning course. In this intervention, learners are prompted to set weekly learning goals, learning tasks, and complete self-evaluations based on learning analytics and a visualization dashboard. This self-regulated learning user interface is designed as a repetitive, longitudinal intervention to support learners' SRL throughout the entire course.

Building on previous MOOC research, this study explores learner behaviors by utilizing cluster analysis and heatmap visualization to investigate potential learning patterns and provide an alternative way of visualizing learner data from a granular perspective.

This study addresses the following research questions:

RQ1: Do learners interact with tools (SRLUI) that are designed to support their self-regulated learning strategies? If so, how do learners use SRLUI?

RQ2: To what extent does the usage of SRLUI have an effect on learner persistence?

RQ3: To what extent does the usage of SRLUI have an effect on learning outcomes?

III – METHODS OF THE STUDY

Overview of Chapter

This chapter will provide a detailed overview of the intervention -- self-regulated learning user interface (SRLUI) and the research design. The design of SRLUI engages the learners through a multi-step process of SRL. The entire learning journey is presented with the design of the features and the theories supporting the design decisions. The next several sections will elaborate on the research design, data source, and data analysis. Since the data source is from three different data pipelines, edX, edX Insights, and SRLUI database, there are extensive explanations of how the raw data are processed in preparation for RQ1 (if and how learners use SRLUI) and RQ2 (learner persistence). The data analysis section is dedicated to the multilevel regression model utilized in RQ3.

Self-Regulated Learning User Interface (SRLUI) Design

The Self-Regulated Learning User Interface (SRLUI) aims to foster SRL strategy skills in 8 MOOCs. SRLUI is created as a standalone module entitled "Weekly Reflection" (see Figure 3.1) at the beginning of weekly content. SRLUI is composed of three pages: course progress, goal planning, and study tip. The treatment group is provided with interactive features in self-evaluation, goal setting, task planning, and reminder setting, whereas the control group is provided with non-interactive components. SRLUI is designed based on Zimmerman's SRL (2000) model to create opportunities for learners to engage in self-regulated activities repeatedly (i.e. self-evaluation, goal settings, task planning, and share weekly takeaways) on a weekly basis. Table 3.1

illustrates the differences in the user interface between the treatment and the control group.

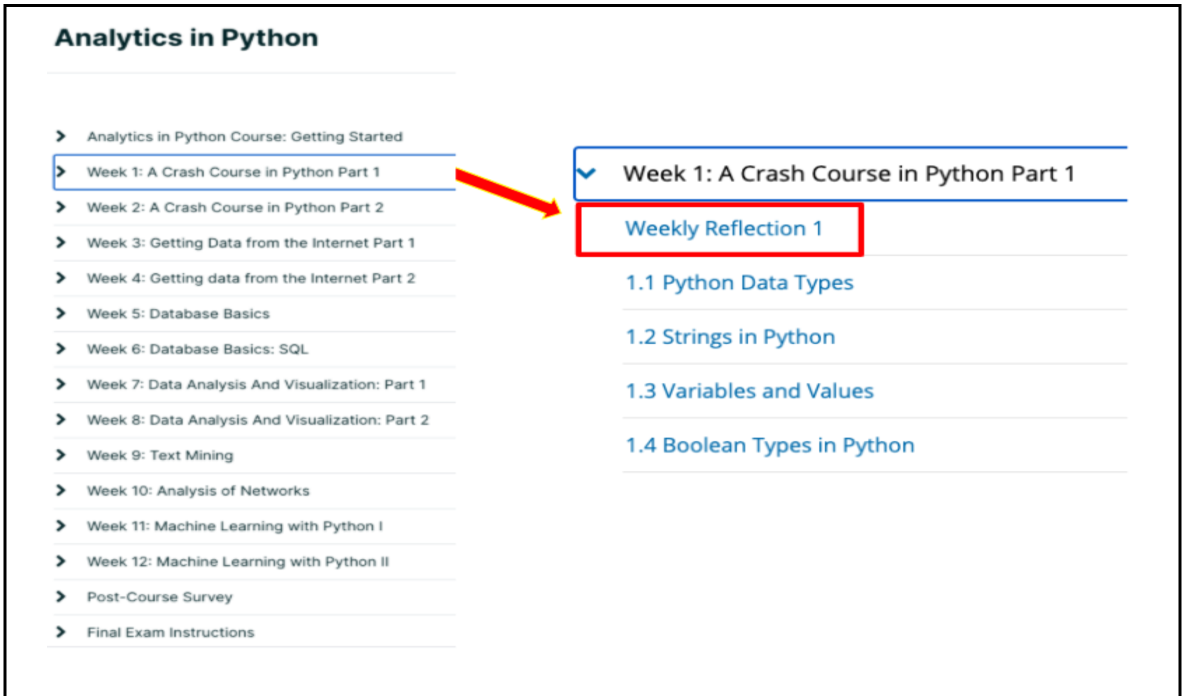


Figure 3.1 *Weekly Reflection*.
This is an example of how SRLUI appears on the course page.

Course progress page. To facilitate the self-evaluation progress, learners are provided with their last week's learning activity regarding the total number of videos watched, quiz problems attempted, and discussion forum activities together with learning goals set in the prior week. In the treatment group (see Figure 3.2), learners are prompted to self-evaluate on a scale from 0%-100% what percentage they completed of their planned learning goals. In the control group, learners do not have the option to self-evaluate. Still, they are provided with learning analytics of the prior week's learning activity and an overview (class progress) of accumulated learning activities (See Figure 3.3)

Table 3.1

User Interface Comparison Between the Treatment and Control Group

Page	SRL activity	Treatment group	Control group
Course Progress	Learning analytics	last week’s learning analytics	last week’s learning analytics
		last week's learning goals	NA
	Self-evaluation	* participants are asked to rate the completion of their study plan, input value range from 0-100%	NA
	learning analytics	historical report of learning analytics	historical report of learning analytics
Study Planning	goal setting	*prompt participants to write down up to 4 learning goals for the next week	provide the learning topics of the upcoming week
	task planning	*prompt participants to write down up to 3 learning tasks	NA
	setting reminders	*set email reminders according to the task planning schedule	NA
Study Tips	general SRL tip	provide a general learning tip with an illustration	provide a general learning tip with an illustration

* means the page requires participants to input information

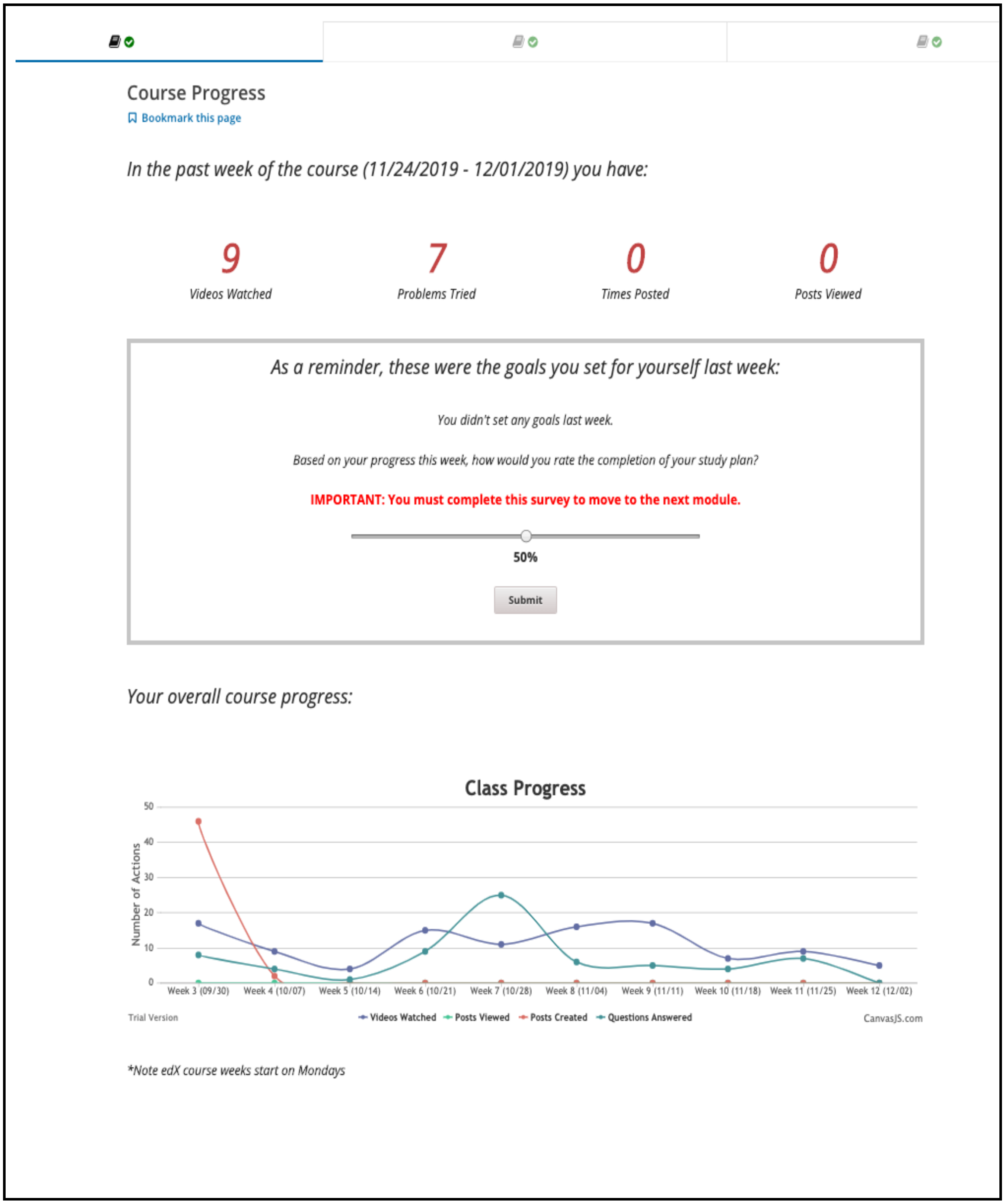


Figure 3.2. *Course Progress Page for the Treatment Group*

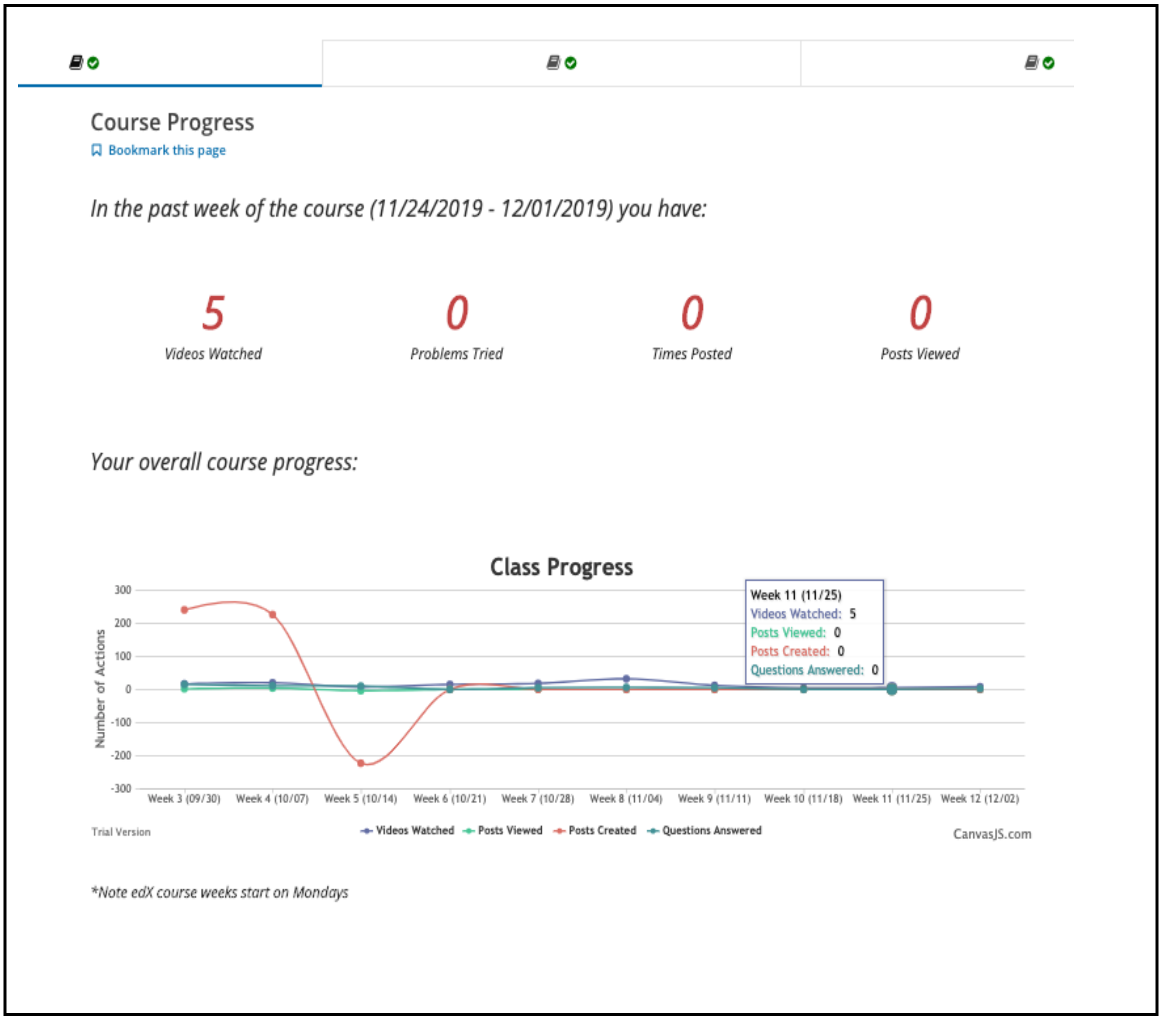


Figure 3.3. *Course Progress Page for the Control Group*

Study planning page. Goal setting is a fundamental step to activate learning (Zimmerman, 2000; Pintrich, 2000). To proceed, learners can type down their learning goals and hit submit (Figure 3.4). Next, learners are prompted to plan their learning tasks and time to study (Figure 3.5). An option to set a reminder allows learners to configure an email reminder to nudge themselves to study. In comparison, learners in the control group are only informed about upcoming learning topics (Figure 3.6).

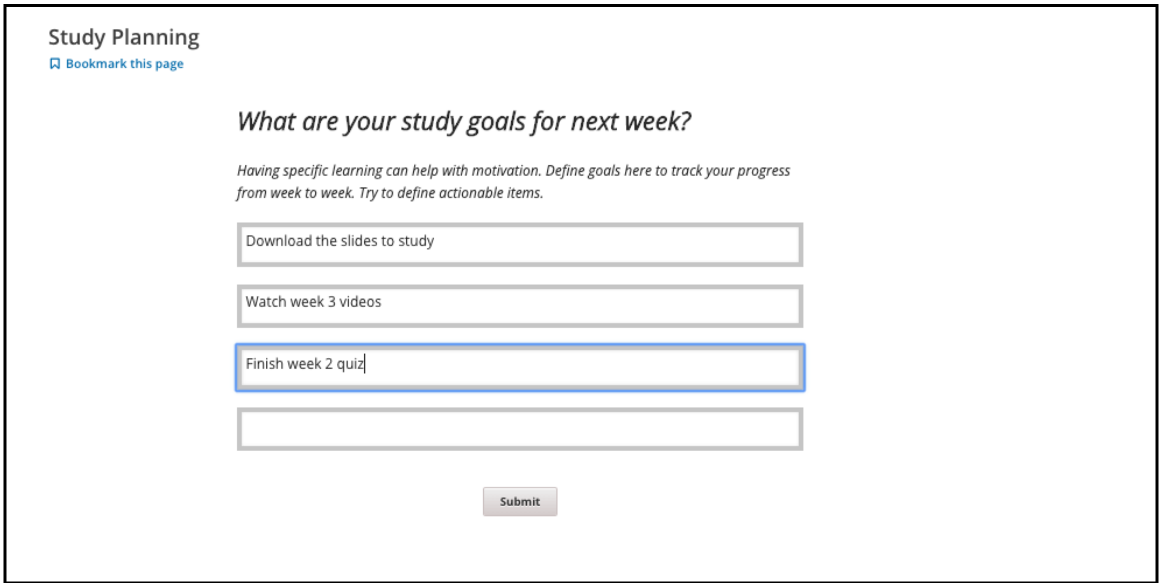


Figure 3.4. *Study Planning Page for the Treatment Group.*
Learners are prompted to set learning goals for the upcoming week.

Study tip page. For the last part of the Weekly Reflection, both the treatment and control groups are provided a study tip as a reminder to exercise SRL strategy skills. This section is designed to showcase learners a variety of general SRL tips to help them adapt their learnings to be more efficient. For example, a study tip says, "Avoid watching a lot of lectures all at once. Break it down into smaller sessions so you don't overwhelm your brain." (Figure 3.7)

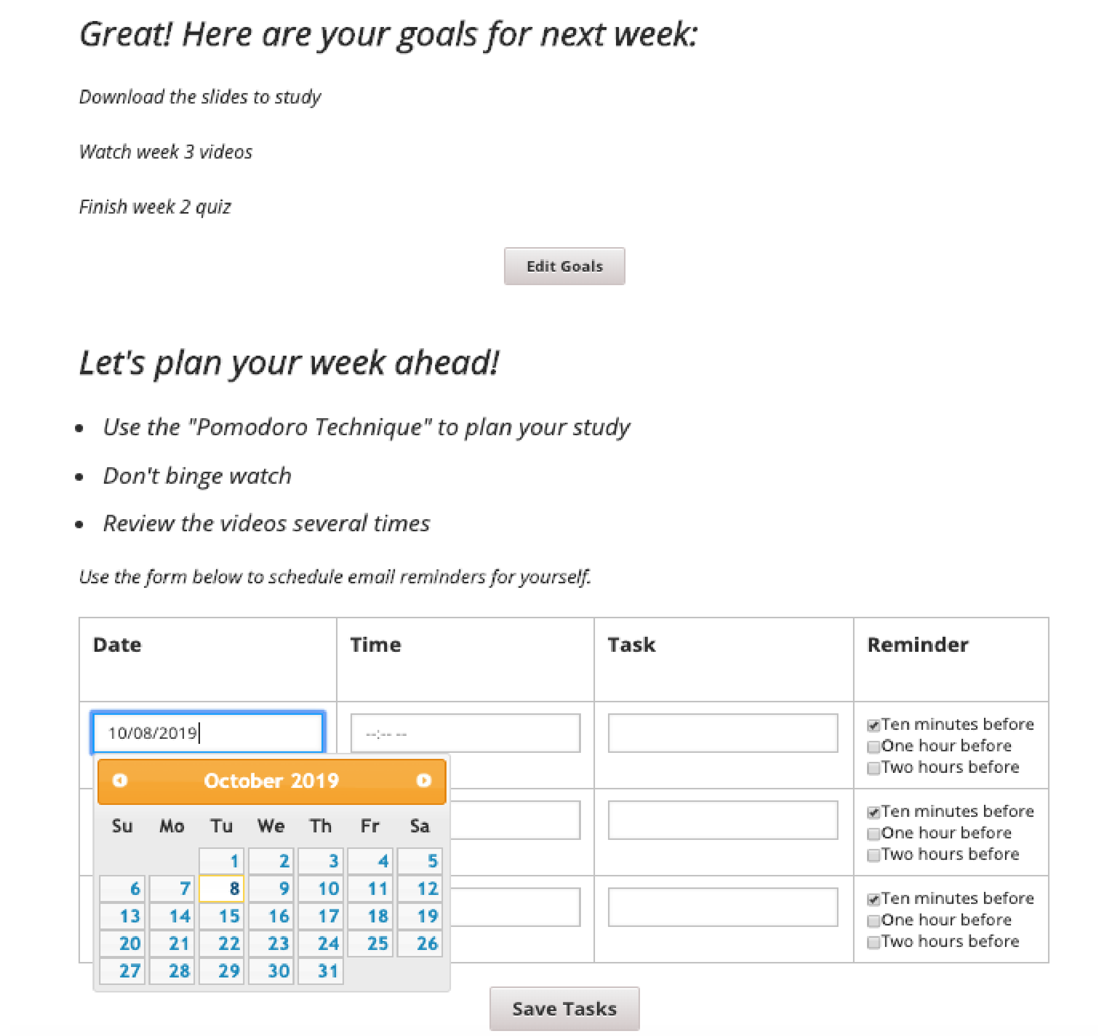


Figure 3.5. *Study Planning Page for the Treatment Group (continued).*
After setting the goals learners in the treatment group are prompted to do task planning and setting up reminders.

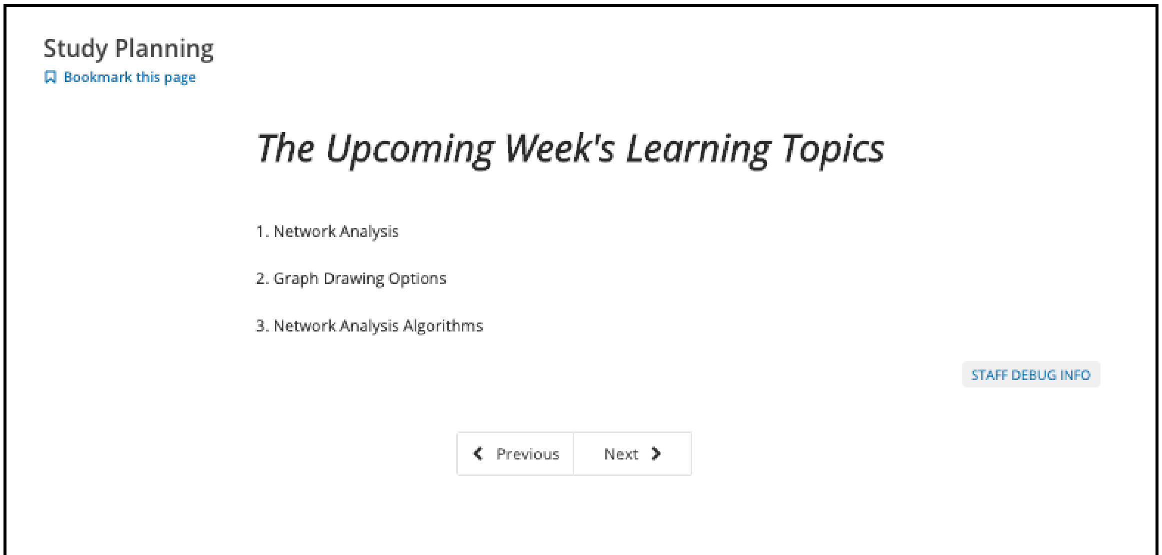


Figure 3.6 *Study Planning Page for the Control Group*
Learners are provided with upcoming learning topics.

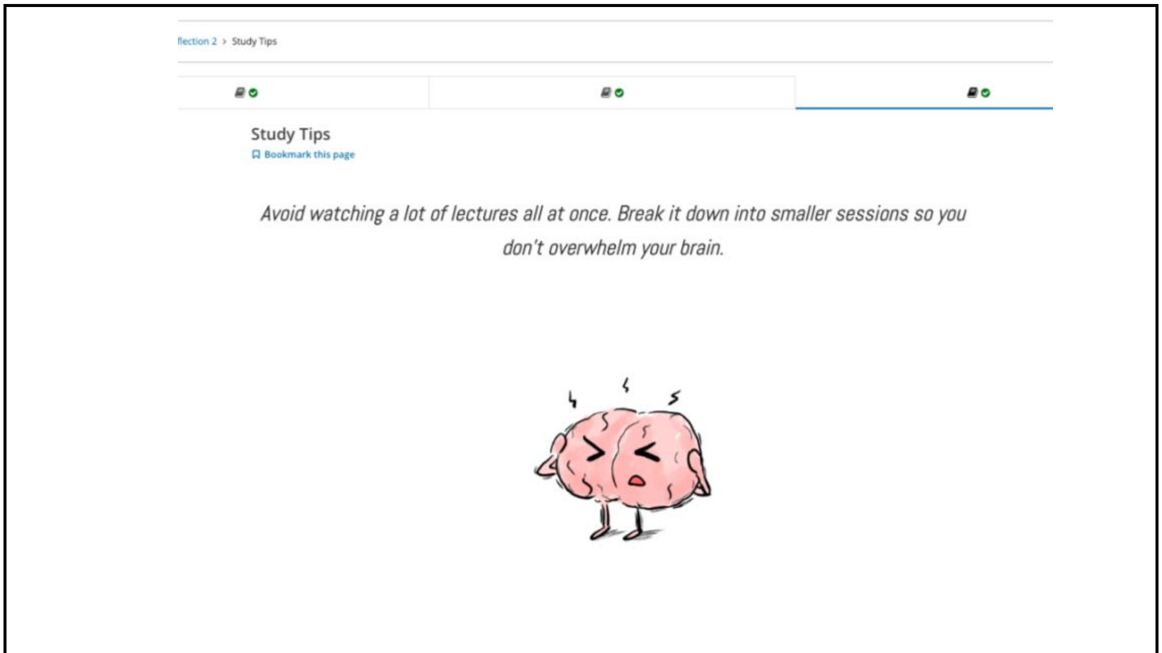


Figure 3.7. *Study Tips Page*
This page provides general study tips for the learners.

Reflection and share key takeaways. Another mechanism of the research design is to invite learners to summarize their learning and share it on the discussion forum to increase social interaction with the community. Thus, a category of KeyTakeaway on the forum is created for such a purpose. Figure 3.8 shows an example of the KeyTakeaway for BA1: Analytics in Python course.

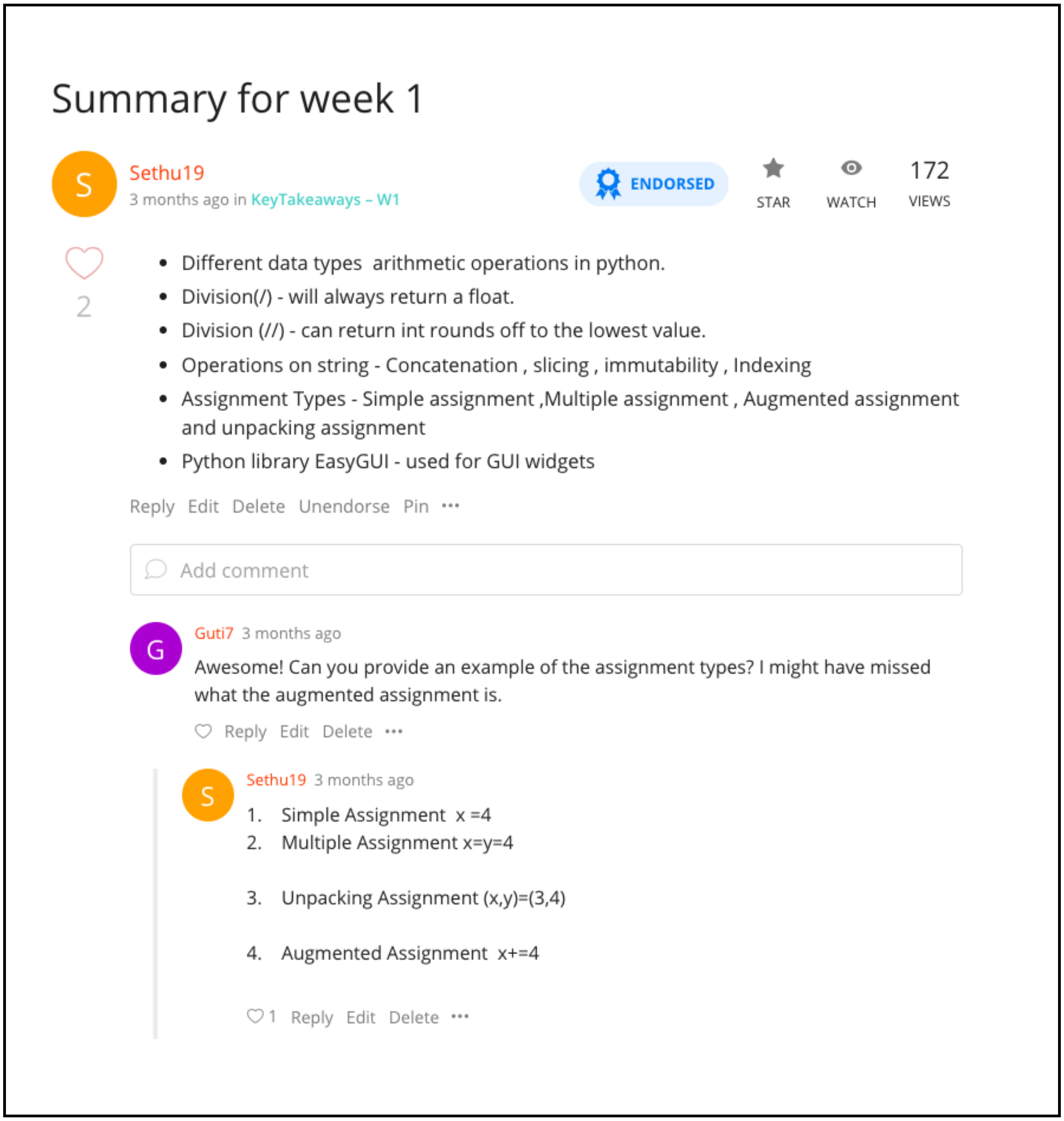


Figure 3.8. A *Key Takeaway Post in the Discussion Forum*

Methods

Participants. Participants of this study were learners registered for two MicroMasters Program MOOCs, each consisting of four courses from 09/15/2019 - 12/26/2019. The computer science (CS) MicroMasters program consists of Artificial Intelligence (CSMM101), Machine Learning (CSMM102), Robotics(CSMM103), and Animation and CGI Motion(CSMM104). The business analytics (BA) MicroMasters program includes Business Analytics in Python(BAMM101), Data, Model and Decisions in Business Analytics (BAMM102), Demand, and Supply Analytics (BAMM103) and Marketing Analytics (BAMM104). Each course was offered for 14 weeks, including 12 weeks of course and 2 additional weeks for the final exam, except Robotics (CSMM103) course was shorter (10 weeks of lecture and 2 weeks for the final exam). The course was open enrollment during the course period (12 weeks). The criteria to earn a course certificate included: paying a course fee, verifying identification, taking a final proctored exam, and achieving a total grade of 60% or above. An initial sample of 1314 verified track learners was included in the study. Since the intervention was implemented from the 5th week of the courses onward, learners who left the course by week four were excluded from the study. That resulted in a total of 808 learners for the final sample. Table 3.2 illustrates the breakdown of participants from each course.

Recruitment. This study was approved by the Teachers College Institutional Review Board (IRB) under protocol 20-189 and Columbia Video Network (CVN) at Columbia University to access the data for research. No participants were recruited solely for the purpose of the study, and the data used as part of the study came from the data collected as part of the curriculum.

Table 3.2
Number of Participants in Eight MOOCs

Course	Number of Participants
CSMM101	235
CSMM102	150
CSMM103	58
CSMM104	40
BAMM101	180
BAMM102	46
BAMM103	34
BAMM104	65
Total	808

Table 3.3 presents the descriptive statistics of demographic information including the total number, mean, standard deviation, minimum, and maximum. Gender information is not available for all participants. However, among the participants who identified their gender, 46% were male and 11% were female. In terms of education, more than 50% of the participants had undergraduate degrees or above and only about 6% of learners had a high school diploma or under. Learners were randomly assigned into either the treatment group or the control group depending on their user identification number (ID): an odd number was assigned to the treatment group and an even number was assigned to the control group. That resulted in a total of 430 participants in the treatment group and 378 in the control group.

Table 3.3

Descriptive Data of Demographic Information (n=808)

Variable	*N*	*Mean*	*SD*	*Min*	*Max*
Gender					
male	372	0. 46	0.5	0	1
female	94	0.11	0.32	0	1
Age	499	33.74	9.66	11	68
Education					
high school or under	50	0.06	0.24	0	1
undergraduate	224	0.28	0.45	0	1
post graduate	184	0.23	0.42	0	1
Treatment Group	430				
Control Group	378				
Class	8				

Data Sources

This section elaborates on the types of data collected for this study. There are three main data sources: (1) learning outcome data, (2) learning behavioral data, and (3) demographics. All the data were collected while learners participated in the courses and the data was exported using edX and edX Insights.

edX is a MOOC learning platform and it also archives learners' learning outcome data from quizzes, assignments and final exams.

edX Insights is a website providing course analytics to the course teams to improve course performance. For example, it records learner behavioral data such as how many videos are watched, how many problems are answered correctly, and how many posts they are made. In addition, edX Insights also reports learner background information on age, gender and education level.

Learning outcome data. Learning outcome data consisted of quizzes, assignments and final exams.

Learning behavior Data. Learning behavior data consisted of all the interactions that the learners engaged with the course, specifically the learning tasks. The learning tasks include watching lecture videos, participating in the graded assessments and the discussion forum. Each activity is also collected on a daily basis through edX Insights and learners grade data through edX Instructor page. The following list is the metrics of learner behaviors data included in the RQ1: the number of video views, the number of problems attempted, quiz attempted, project attempted, average quiz grade, average project grade and final exam grade.

Demographics. Participants' demographic data including the age, education level was collected through *edX Insights*. Figure 3.9 is an example of learning behavior log data collected on a daily basis in *edX Insights* and Figure 3.10 is a visualization based on the log data.

Activity Over Time

Date	Discussion Contributions	Problems Correct	Videos Viewed
December 11, 2019	0	32	0
December 10, 2019	0	0	0
December 9, 2019	0	12	0
December 8, 2019	0	29	1
December 7, 2019	0	0	0
December 6, 2019	0	0	6
December 5, 2019	0	0	0
December 4, 2019	0	0	3
December 3, 2019	0	3	5
December 2, 2019	0	48	9

Figure 3.9. *Example of Log Data from edX Insights*
This is the log data from edX Insights recording a participant's log data in BAMM101 course. A visualization of learner log data is shown in Figure 3.10.

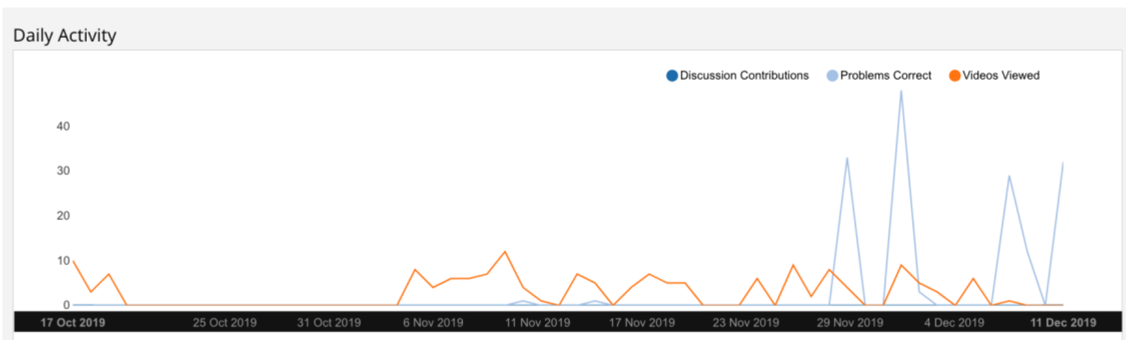


Figure 3.10. *Learning Daily Activity from edX Insights*
edX Insights provides a chart based on a participant's log data from Figure 3.9. The X axis represents the date, and each mark is on a 7 day interval. The Y axis indicates the number of the corresponding activities. Orange line indicates the numbers of videos viewed and blue line indicates the problem answered correctly.

SRLUI usage. To store and retrieve learners' interactions with SRLUI, an HTTPs server was built using Node.js in a MongoDB database. The SRLUI server also tracked and stored learners' participation in the discussion forum by using the discussion forum, edstem's API. Additionally, customized and secured APIs were developed to present the following learning behaviors in the course planning page: self-evaluation of learning goals, goal settings, posting and viewing counts of discussion forum activity, video watching and problem attempts and notifications of task reminders.

To answer RQ1: "How do learners use SRLUI?" I created a series of measures to account for the frequency and the amount of text entry in SRLUI Activity (Table 3.3).

Table 3.4
SRLUI Activity Data and Measure Description

Variables	Description	Coding or List of Possible Data Entry
useSRLUI	If a subject use SRLUI features (goal, tasks, reminders, self-evaluation) at least once, it is considered use SRLUI.	Yes=1, No=0
goals count	The total number of goals learners enter on goal setting	*Possible Responses (2 goal counts):* Watch Week 1 Video, Finish week 1 quiz
goals avg wordcount	The average goal word count is devised based on the total goals word counts divided by the number of goals	*Example:* Goal: Watch Week 1 Video, Finish week 1 quiz There are in total of 8 words, 2 goals, the goals avg wordcount is 8/2 = 4
tasks count	The total number of tasks enter in tasks planning	*Possible Responses (3 task counts):* project wk2, review videos week 3, finish assignment 4
tasks avg wordcount	The average task word count is devised based on the total task word counts divided by the number of tasks	*Example:* project wk2, review videos week 3, finish assignment 4 There are in total of 9 words, 3 tasks, the tasks avg wordcount is 9/3 = 3

Table 3.4 SRLUI Activity Data and Measure Description (Continued)

Variables	Description	Coding or List of Possible Data Entry
self-evaluation count	The total number of self-evaluation entry	*Possible Responses (2 self-evaluation counts):* Based on your progress this week, how would you rate the completion of your study plan? (Answer can range from 0 - 100%)
reminders count	The total number of reminder emails learners schedule to send out based on time planned for the tasks	*Possible Responses (1 reminder count):* [X]10 mins prior [ ] 1 hr prior [ ] 2 hrs prior
Other course-related info		
Create time	Date and time of each user action	*Example:* 11/6/19 6:41:15
Course ID	Course ID	*Example:* CSMM101
userID	a unique numeric number associated with each learner	*Example:* 10646139

Note: a learner could enroll in more than one course. Thus, a subject in the sample is composed of an unique userID with a unique courseID. For example, if learner ID, 10646139, enrolled in 2 courses, it is considered two unique subjects in the dataset.

SRLUI Usage Patterns

To further explore RQ1.2 on "How learners use SRLUI?", random learners are selected to demonstrate variability in how learners use SRLUI. In order to choose learners to represent the diversity of the populations in the sample MOOCs, hierarchical cluster analysis with heatmap visualization is utilized to identify any potential patterns or subclusters of learning behaviors. Based on the findings of the cluster analysis, a learner is selected from each cluster to provide descriptive analysis on their usage of SRLUI. To facilitate the HCA analysis, a set of metrics including learning behaviors, participations of the graded assessments, and the formative assessment grades are generated from the log files. These included the number of video views, number of problems attempted, project attempt, quiz attempt, average quiz grade, average project grade and final exam grade.

The HCA analysis in this study is structured based on Bowers (2010) and Hawn (2019) suggestions and the key steps include: 1. All data are standardized through z-scoring to prevent overweighting in the subsequent similarity matrix; 2. Euclidean measure is utilized to calculate the distance between each pair of observations; 3. The linkage between the groups of subjects is established with Ward's method; 4. Learners are clustered in rows and the behavioral data in columns. Each cell is represented by colored blocks of the heatmap to showcase the full range of the individual variation; 5. Annotations, binary color-coded vertical bars, are utilized to elaborate learner information (i.e. dropout, use of SRLUI, certificate and grades).

In short, to select example learners to demonstrate their usage of SRLUI, an agglomerative clustering with Ward's minimum variance method (Murtagh & Legendre, 2014) is employed through an iterative process to group the most similar observations and groups of observations. HCA and the heatmap are generated using R version 3.5.1 in Rstudio 1.3.1093. The heatmap visualization uses the ComplexHeatmap package in R (Gu, 2016).

Learner Persistence with Survival Analysis

To answer RQ2 on how SRLUI has an impact on learner persistence, a nonparametric estimator of the survival function (Kaplan & Meier, 1958) is utilized to estimate and graph survival probabilities as a function of time (Min et al., 2011). Survival analysis has been long studied in the medical field (Cox, 1972; Kaplan & Meier, 1958) and Willet and Singer (1991) advocated how such an approach could be used to investigate educational measures such as teacher attrition and student dropout rate.

Survival analysis, by definition, is the probability of when and whether an event occurs during an observed period of time (Willet and Singer, 1991). A censored survival time means that an event does not occur during the study time (Emmert-Streib & Dehmer, 2019). According to Emert-Streib & Dehmer (2019), there are three types of censoring: Type I censoring indicates all subjects begin and end the study at the same time; Type II censoring means all subjects begin the study at the same time, but the study ends when a predetermined condition meets; Type III censoring suggest that participants start the study at different times, but the length of the study is fixed. The survival analysis adopted in this research belonged to type III censoring because learners could join the course anytime during the 12 weeks and the course ended by the 14th week for all learners.

To address the 2nd research question: to what extent does SRLUI have an effect on learner persistence, this study applies Kaplan-Meier survival analysis to compare learner persistence between the treatment versus the control group. In this analysis, the event of interest is learner persistence and learner dropout. Specifically, learner persistence was calculated as the number of days based on learner’s first date of learning activity to the last day of learning activity (within the course beginning and end dates). Any activities beyond the course end date are not included in the analysis. In other words, this survival analysis is both right- and left-censored. The longest survival days are up to 98 days (14 weeks) and the dropout occurs if a learner leaves the course before the final exam week starts (the 13th week). Learner dropout is decided whether a subject stays active until the final exam week.

Figure 3.11 shows an example of how dropout and survival days are calculated. For example, learner A joins the course from the middle of week 4 until week 7. Since learner A leaves the course before the final exam starts, learner A is considered a dropout (coded as 1) and the survival days are 17 days. Learner B accesses the course before week 1, yet the total survival days are computed from the start date of the course. Given that learner B leaves the course in week 13, the participant’s survival days are 91 days, and is coded 0, no dropout. Learner C starts the course from week 7 and remains active after the course has archived. Learner C’s survival days are calculated from week 7 until the course is archived (week 14); thus the survival days is 49 and is coded 0, no dropout. Learner D starts from week 2 and leaves the course at week 10. Learner D survives for a total of 56 days, and is coded 1 for dropout. Rstudio and the survival and survminer packages were used in the model (R Core Team, 2020; Therneau, 2020; Kassambara, Kosinski & Biecek, 2020).

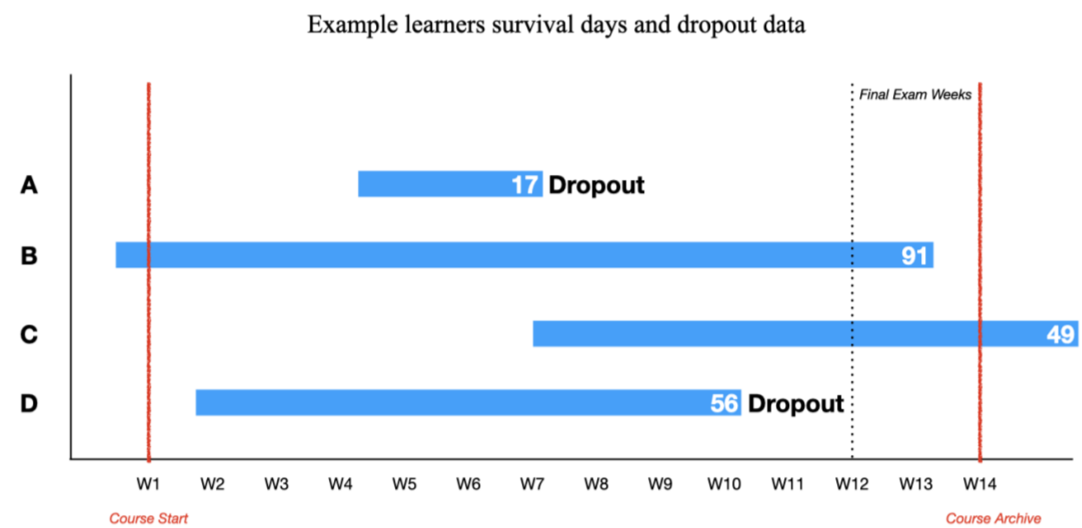


Figure 3.11. *Examples of Learner Data for Survival Days and Dropout*

Data Analysis

Hierarchical Linear Model. To explore RQ3 on how SRLUI has an effect on learning outcome, a random intercept two-level hierarchical linear model (HLM) is used to appropriately estimate the independent effects of the student variables including learner age, gender, educational level and usage of SRLUI. This approach is appropriate because the clustering of students by class violates common assumptions of independence of residuals in linear regression models (Bowers & Urick, 2011; Hox, Moerbeek & Schoot, 2017; Raudenbush and Bryk, 2002). The multilevel regression model assumes a hierarchical data structure with subjects who are nested within pre-existing groups; for example, dependent variables or response variables are situated in the student-level, whereas explanatory variables at all existing levels (Hox, Moerbeek & Schoot, 2017). Traditional methodologies such as regression and ANOVA would be inefficient to answer questions with explanation variables across different levels (McCoach & Adelson, 2010). Ignoring these clustering effects or nonindependence could result in issues such as incorrectly reducing the standard error, falsely increasing the confidence in estimated parameters, or outcomes; thus increasing type I error (O'Connell and McCoach, 2008). HLM is a well-established method to allow researchers to model multiple levels of a hierarchy and examine relationships and interactions among variables across multiple levels (McCoach & Adelson, 2010).

For example, Bowers and Urick (2011) employed a two-level HLM model, nesting students within schools, to estimate the direct effects of facility maintenance and disrepair on longitudinal high schooler mathematical achievement with a large nationwide dataset. Additionally, Kizilcec et al. (2020) also utilized the multilevel modeling approach in MOOC research.

Considering the sample data was clustered in 8 MOOCs; thus, HLM was applied to examine the independent effect of SRLUI on learning outcome data from 8 MOOCs. The HLM equation can be expressed in Equation 1 :

Level 1: $Y_{ij} = \pi_{0j} + \pi_{1j}SRLUI_{ij} + \omega X_{ij} \ldots + \varepsilon_{ij}$ Equation (1)

Level 2: $\pi_{0j} = \gamma_{00} + u_{0j}$

Level 2: $\pi_{1j} = \gamma_{10} + u_{1j}$

Where:

Y_{ij} = Dependent outcome variable for student i in course j, here learner grade

ω = Vector of fixed effects of the student level covariates

X_{ij}= Vector of student level covariates

γ_{00}= The value of the intercepts varying across courses

γ_{10} = The slope of the effect of SRLUI across courses

ε_{ij} = Level 1 residuals

u_{0j} = Level 2 residuals for the intercept

u_{1j} = Level 2 residuals for the slope

To answer research question 3, the magnitude, direction and precision of γ_{10} will be examined. If the estimate is positive and statistically significant, then it demonstrates a positive effect of SRLUI on learning outcome.

The student-level variables include learner characteristics and the treatment variable, useSRLUI, to indicate if a learner is in a treatment group. Dummy-coded measures were created to account for indicators: gender (male=1, female=0), educational level indicator, postgraduate(yes=1, no=0). Learners in the treatment group are labeled SRLUI treatment (yes=1, no=0). Given that learners can choose to interact with SRLUI or not, a variable "useSRLUI" (yes=1, no=0) is devised to indicate learners in the treatment group who use SRLUI. This allows subsequent analysis on intent to treat (ITT) and treatment on the treated (TOT) HLM modeling[2]. The learning outcome is based on a scale 0-100, referred to summative grades of learning assessments including quizzes, projects and a final exam. No class-level variables are included in the HLM modeling. For all HLM models, lme4 R packages and R studio are used for the statistical analysis (Bates, Mächler, Bolker and Walker, 2015).

[2] In a randomized experiment, to account for subject dropping out or not interacting with the artifact of the study, the HLM analysis examines the results of intent to treat (ITT), indicating subjects who are in the treatment group; versus treatment on the treated (TOT), meaning subjects who actually interact with SRLUI.

IV – RESULTS OF THE STUDY

RQ1.1: Do learners interact with tools (SRLUI) that are designed to support their self-regulated learning strategies?

A previous MOOC study reports that learner participation in SRL tools in MOOCs was low, with about 10% to 30% of students interacting with SRL tools at least one time (Davis et al., 2018; Jansen et al., 2020). To address the concern that learners might not access or take advantage of the SRL tools in MOOC environments, this research question aims to explore the extensive variety of learning behaviors interacting with SRLUI in eight MOOCs.

According to Table 4.1, 342 people (78%) out of 430 learners in the treatment group engage with SRLUI at least one time--here defined as having engaged with at least one of the following features of SRLUI: self-evaluation of plan completion, setting weekly goals, planning learning tasks, or setting a self-reminder. Specifically, goal-setting (n=241) and sliders (n=335) are the most heavily used compared to task planning (n=73) and reminders (n=73).

Our finding suggests that MOOC learner participation in our SRL tools (78%) is higher than in the previous study (10%-30%) (Davis et al., 2018; Jansen et al., 2020).

Table 4.1
Descriptive Data of Treatment Group (n=430) Interaction with SRLUI

SRLUI Activity	*N*	*Mean*	*SD*	*Min*	*Max*
useSRLUI	342	0.78	0.41	0	1
goals count	241	7.44	7.54	1	41
goals avg wordcount	241	4.59	2.71	1	37
tasks count	73	6.64	7.72	1	45
tasks avg wordcount	73	2.75	1.23	1	5.25
self-evaluation count	335	1.35	0.85	1	8
reminders count	73	6.01	7.14	0	42
Number of courses	8				

RQ1.2: How do learners use SRLUI?

In order to understand how learners interact with SRLUI, this section provides a descriptive analysis of learner profiles by exploring their learning behaviors and usage of SRLUI. In order to select samples from the treatment group to represent the diversity of learner profiles, a hierarchical cluster analysis (HCA) and heatmap are applied to investigate any potential subgroups or shared patterns of clickstream or behavioral data (Bowers, 2010b).

The variables included in the cluster analysis are number of video views, problem attempt, quiz attempt, average quiz grade, average project grade, as well as final exam grade. Based on the principles for the creation of a clustergram (Bowers, 2010b; Jorion et al., 2020), all the variables are z-scored, and illustrated as a single color block. The color gradient for the color blocks in the heatmap ranges from a more intense, “colder”, blue for learning activities -2 standard deviations below the mean, to grey as mean, to a more intense, “hotter”, red for learning activities +2 standard deviations above the mean, with missing data represented in white.

Figure 4.1 shows the result of the HCA analysis. The annotations on the right side of the heatmap are labeled of learners’ association with learning outcomes and usage of SRLUI. The annotations are not used in the heatmap to cluster learners. “Grade” represents the total grade and it uses a 10 scale gradient level from 0-1 to represent 0-100 points. For example, Level 0 is white, representing grade 0-9 points, and level 1 is black, indicating 90-100 points. “Dropout” uses dark color to represent a learner dropout. “useSRLUI” applies dark color to indicate a learner used SRLUI. “Cert” uses dark color to indicate a learner earned a certificate (with grades >= 60).

Sub-clusters features. Figure 4.2 displays the trends of four sub-clusters based on their learning behaviors from the hierarchical cluster analysis. On average, groups A and C have passing grades, whereas groups B and D have non-passing grades. In a closer inspection, Group A and C both have above average levels of participation in the graded assessments; however, group C has a higher number of video views and problem attempts in comparison to group A. On the other hand, group B has below-average learning activities, while group D shows the lowest participation of all the learning activities. In terms of the grade, group A (mean=78.7) and C (mean=75), both have, on average, passing grades. In contrast, group B, on average, scores 20, and group D is close to 0 (mean = 2.4).

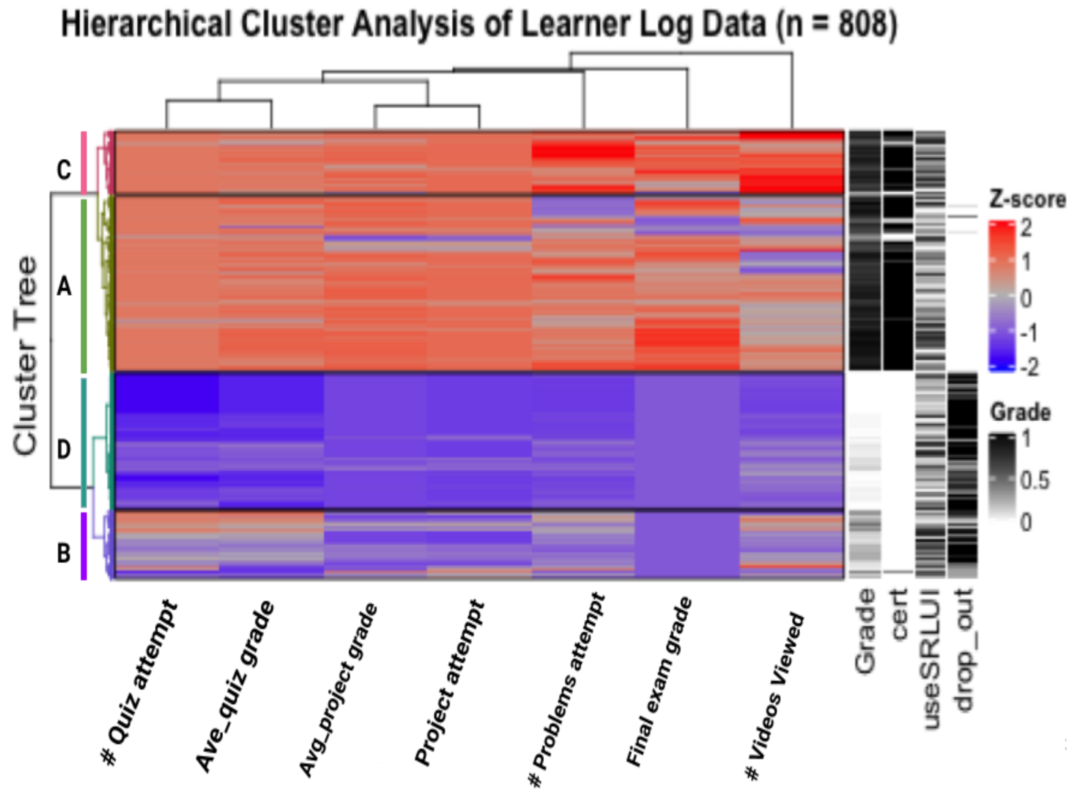


Figure 4.1. *Hierarchical Cluster Analysis of Learner Log Data (n=808)*
Hierarchical cluster analysis is based on learning behaviors (i.e number of video viewing and problems attempted) and participation of graded assessments. Hierarchical clusters are represented by a cluster tree (left) to indicate the clusters. Each learner is aligned along the horizontal axis, with learning behaviors aligned along the vertical axis. Learning behaviors are z-scored; higher number indicated by an increasing intensity of red, lower number indicated by an increasing intensity of blue, the mean indicated by grey (center). The solid black lines through the heatmaps indicates the division line between four major clusters in the full dataset.

The annotations on the right side of the heatmap use black bars to represent dichotomous categorical variables: certification, useSRLUI, and dropout. Grade uses 0-1 with a 10-level gradient to showcase the total grade (from 0-100). “Drop_out" uses dark color to represent a learner dropout. “useSRLUI" applies dark color to indicate a learner used SRLUI. “Cert” uses dark color to indicate a learner earned a certificate (with grades >= 60).

Four vertical colored bars between the cluster tree and the heatmap (left) denote four clusters (A, B, C, and D) based on the HCA analysis. Further explanation of the clusters can be found in Figure 4.2.

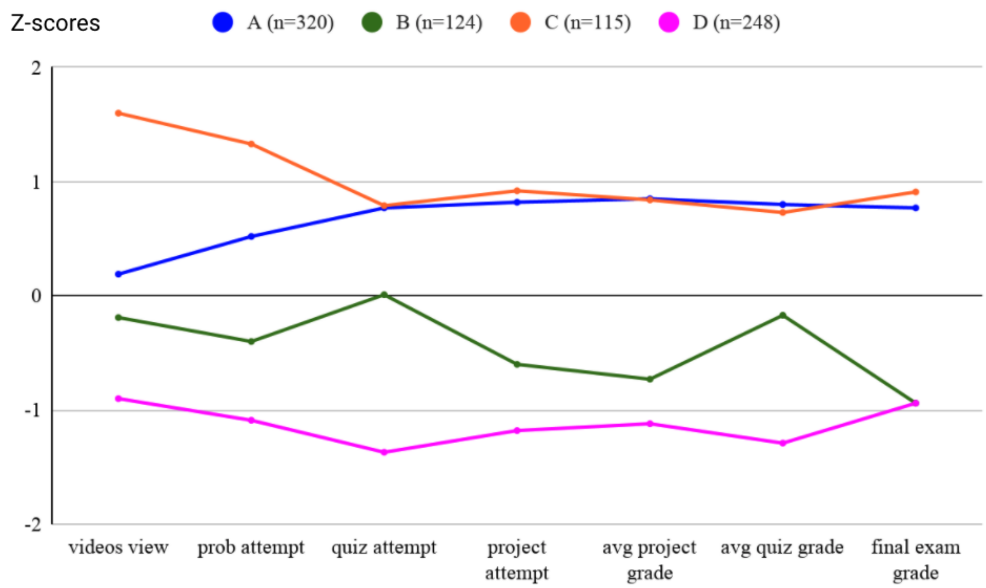


Figure 4.2. *Z-scores of Four Clusters*
This figure illustrates the means of learning behaviors (z-scored) trends of four sub-clusters based on the hierarchical cluster analysis.

Prior literature suggested that MOOCs attract learners from diverse backgrounds with various motivations, demographics, prior knowledge, and learning goals (Deboer et al., 2013; Gardner & Brooks, 2018). In the following section, a randomly selected learner from each cluster would be used as examples to showcase the diversity of learner profiles on how they use SRLUI and their learning trajectories in the sample dataset.

Example learner from cluster a. Alex (pseudonym) is a sample learner from cluster A who starts CSMM101: Artificial Intelligence course from week 1 and persists until the last week 14, for a total of 98 days with no dropout. Learners in cluster A on average, have passing grades and higher than average participation in the course and attempts in the graded assessments. Alex, specifically, achieves 42 points for the total grades and shows consistent interaction with the lecture videos and graded assignments throughout the course. Figure 4.3 shows Alex uses the goal planning and reminders features heavily, but only participates in self-evaluation once. Table 4.2 details the goals planning and it covers active learning (i.e. watching videos, reading handouts, and doing assignments), reaching out to peers (i.e. participating in the forum) and strategic goals (i.e. asking for a vacation day from the employer, and completing easy questions as strategies to complete the course). In short, Alex shows consistent interaction with SRLUI in goal setting and reminders throughout the course.

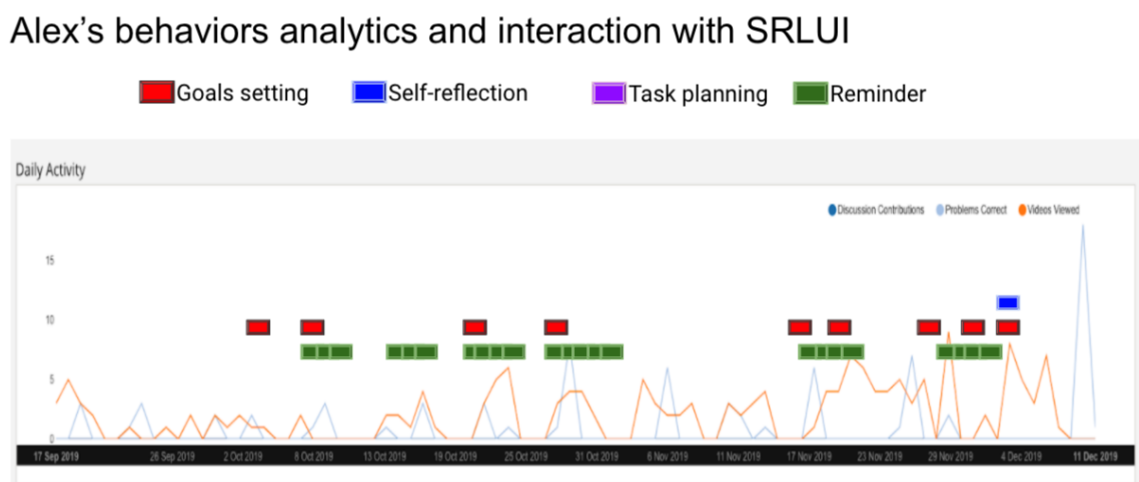


Figure 4.3. *Alex's behaviors analytics and interaction with SRLUI*
The x axis is the date of the log data and y axis represents the total number of learning activities. The orange line indicates the number of video views and the light blue is the number of problems answered correctly. Interaction with SRLUI is indicated with four colored blocks, and each color represents a specific SRLUI feature. Goal planning is illustrated with red, self-reflection is blue, task planning is purple and reminders are green. The position of the blocks also correspond with the createtime of the log entry.

Alex has an active participation in the course videos and problems; and multiple interactions with goals and reminders. Learner A interacted only once with self-reflection, and none with the task planning feature.

Table 4.2
Alex's Log Data with SRLUI on Goals Setting

Week	**Goals**
4	Get to know the various applications of the AI; First steps in the programming language Python ;
4	Get information on how to get an executable Python environment installed; Familiarization with Python syntax, data and program structures; Collect learning materials for programmers switching to Python
6	Make up for unprocessed lessons from previous weeks; In order to have the necessary time to participate in the discussion forums, apply for a vacation day at the employer; Watch all videos of the week attentively
8	read suggested readings; Refresh Knowledge in Probability and Statistics; Compare Handout with textbooks in my native language German
10	A discussion contribution to the topic of course participation as a foreign language speaker; To find out about proctored exams or regular exam.; Adding missing points that the following lessons are no longer locked.
11	watch videos; Quiz answer based on the handouts
11	Re-view old books on formal logic.; Answer the simple quiz questions and guess the others.
12	Get informed about the Protected Exam; Course content catching up
12	Answer and submit open quiz questions from previous lessons; Preparation and information about the protected exam

Example learner from cluster b. Bianca (pseudonym) is a sample learner from cluster B. Cluster B learners, on average, a total score of 20 out of 100, and has below-average learning activities and the graded assessments. Bianca enrolls in CS101 Artificial Intelligence course, starts from week 1 and persists for a total of 65 days. Bianca drops out at week 10, achieving a total grade of 11 points. Figure 4.4 illustrates Bianca's overall learning behavior and interactions with SRLUI. Bianca participates most frequently with goal setting but less frequently with task planning and reminder features. Table 4.3 suggests that Bianca is caught up in the coding assignment 1 and comments it "took longer than expected". In week 7 goals, it shows that Bianca adjusts her study plan by allocating "more time aside for assignments". One of SRLUI design principles is to support learners with the ability to evaluate and reflect their learning process to make necessary adjustments to achieve their learning goals (Zimmerman, 1998).

It seems that Bianca identifies more time as needed for study; however, she left the course around week 10. The prior MOOC literature suggests that dropout could be resulting from the content being hard, other life events taking priority over, or losing interest (Kizilcec & Halawa, 2015). Furthermore, dropping out from MOOC has much lower risk and consequences compared to traditional residential programs in accredited institutions (Gardners & Brooks, 2018). That could also explain the potential cause for the dropout behavior here.

Figure 4.4. *Bianca's Behaviors Analytics and Interaction with SRLUI*
The x axis is the date of the log data and y axis represents the total number of log entries. The orange line indicates the number of video views and the light blue is the number of problems answered correctly. Interaction with SRLUI is indicated with four colored blocks, and each color represents a specific SRLUI feature. Goal planning is illustrated with red, self-reflection is blue, task planning is purple and reminders are green. The position of the blocks also corresponds with the create time of the log entry. Bianca has consistent and active participation in the videos and problems; and interacts mainly on goals, then task planning and reminders; and once with self-evaluation.

Table 4.3
Bianca's Log Data with SRLUI on Goals and Tasks Planning

Week	Goals	Tasks
4	Do the prerequisite readings; Review algorithms and notes; Solidify understanding	Review previous videos, watch new ones; Coding Assignment Attempt Part 2
4	Catch up on Week 3 Modules; Make association between algorithms and coding	Review previous videos, watch new ones; Coding Assignment Attempt Part 2
4	Use Columbus day to catch up on some modules	Review previous videos, watch new ones; Coding Assignment Attempt Part 2
4	Try and Complete the first Programming Assignment in the next 1-2 weeks; Catch up to Week 5 by EOW; Find out why my code is currently breaking in first coding assignment.	Review previous videos, watch new ones; Coding Assignment Attempt Part 2
6	Week 1 through 4 Review; Week 5 videos done by Thursday; Week 6 Videos Start Friday	Week 1 through 4 Review; Week 5 Videos Pt. 2; Week 6 Start
7	Watch Week 6 Videos and do the quiz; More time aside for assignments. Assignments taking longer than expected	
9	Review Week 1-6; Start Chapter 7 and 8; Start Chapter 9/10 by next Monday	

Example learner from cluster c. Cecil (pseudonym) is a sample data from cluster C. Cluster C is the most active group of learners in the course in terms of video watching and participation in the graded assessments. Cluster C learners also have a passing score averagely.

According to Figure 4.5, Cecil enrolls in BAMM104 marketing analytics course starting from week 1 and persists until the end of the course for 86 days without a dropout. Cecil scores a total of 76 points which is slightly higher than the average of Cluster C (mean=75). Figure 4.5 shows Cecil's log data with consistent participation for video watching and attempts in problems. Cecil also interacts with SRLUI heavily in goals, tasks, and reminder features. Table 4.4 lists Cecil's weekly goals and tasks based on the SRLUI log data. Figure 4.6 presents the reminder feature which allows users to schedule emails sent out in advance to remind oneself time to study. Figure 4.7 is an example of the reminder email sent to Cecil. In addition, Cecil uses self-evaluation once.

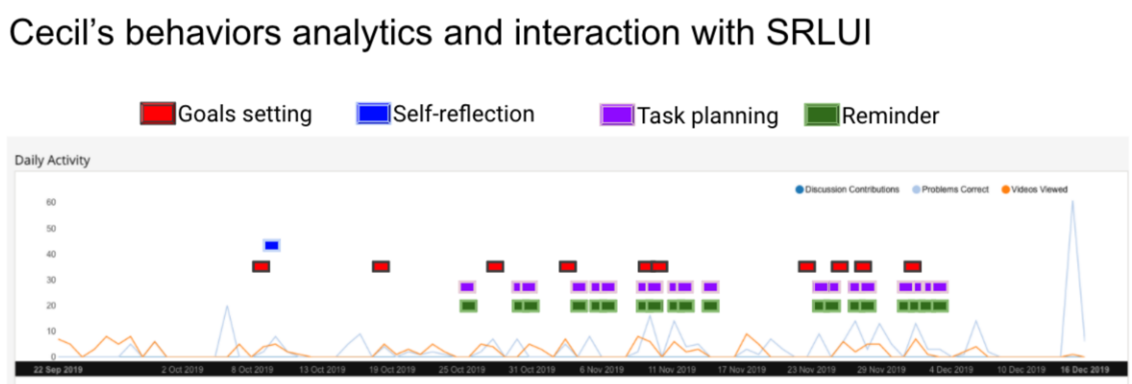


Figure 4.5. *Cecil's Behaviors Analytics and Integration with SRLUI*
The x axis is the date of the log data and y axis represents the total number of log entries. The orange line indicates the number of video views and the light blue is the number of problems answered correctly. Interaction with SRLUI is indicated with four colored blocks, and each color represents a specific SRLUI feature. Goal planning is illustrated with red, self-reflection is blue, task planning is purple and reminders are green. The position of the blocks also corresponds with the create time of the log entry.

Cecil has a continuous participation in the course videos and problems; and multiple interactions with the SRLUI on goals, task planning and reminders. She only uses once with self-reflection.

Table 4.4
Cecil Log Data with SRLUI on Goal and Tasks Planning

Week	Goals	Tasks
4	Watch Week 3 and 4 Videos; Complete week 3 quiz; Complete week 2 project	
5	watch all videos; complete week 3 programming assignment;	Study
7	watch week 5 videos	Study; Study
8	Watch week 6 videos; work on project 2	Study
8	Watch week 7 videos; Complete week 7 quiz; Start week 8 videos	Study
9	Watch week 8; complete week 8 quiz; start project 2	Study; Study; Study
10	Watch Week 9 and 10 lectures; Week 9 & 10 Quizes; Start final project	Study; Study; Study
11	Watch all week 10, 11 videos; complete week 10 and 11 quizes; complete final project	watch lectures; watch lectures; Study
11	Watch week 11 and 12 lectures; complete week 11 and 12 quizes; study for final	watch lectures; watch lectures; Study
12	Watch week 12 videos; Complete week 12 quiz; Complete last project	Study; watch lectures; Quiz

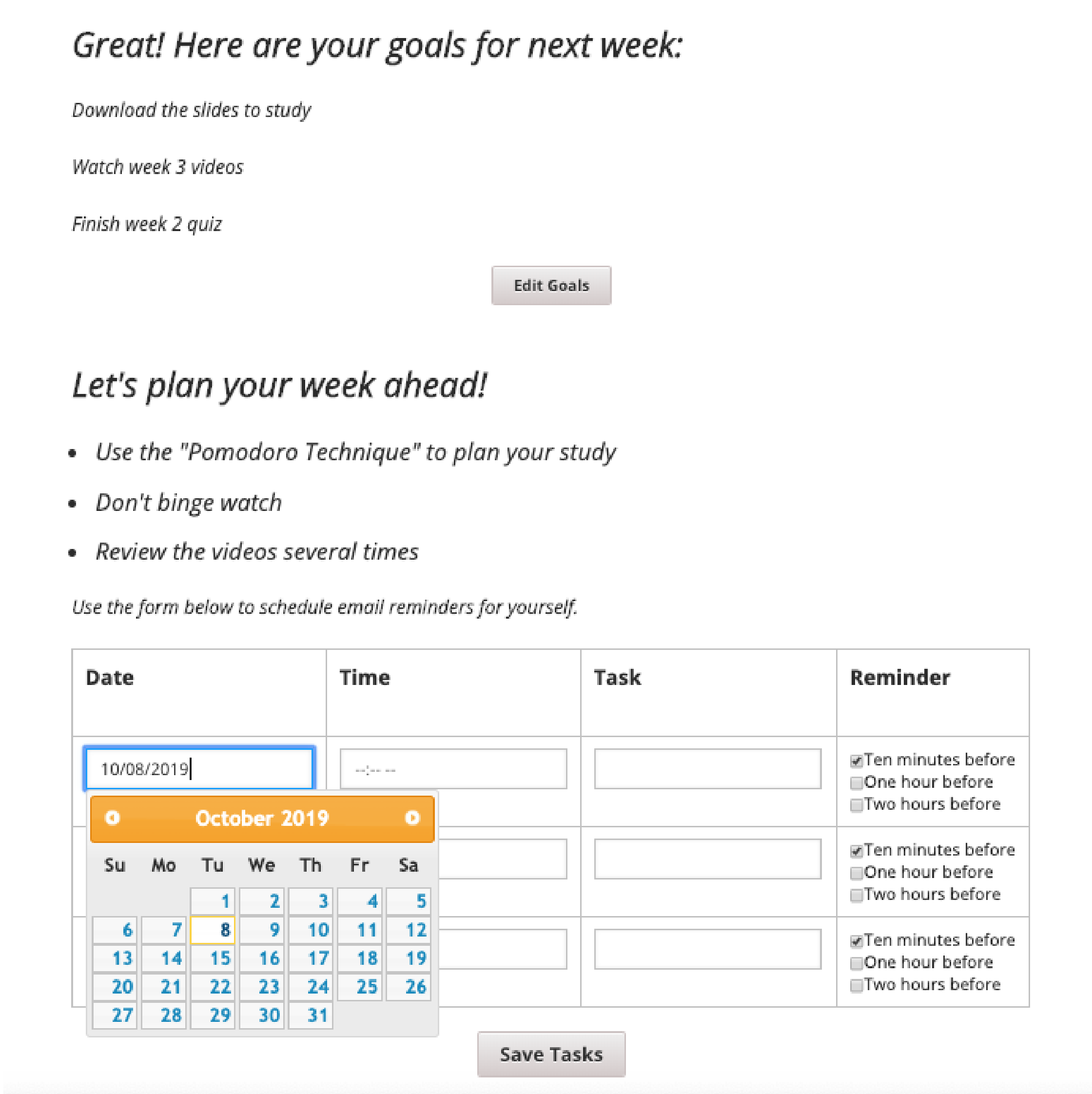


Figure 4.6. *Study Planning Page Part 2*
After setting learning goals, learners would be prompted to do tasks planning. By putting down the learning tasks and the time to study, learners would receive reminder emails based on their input on SRLUI.

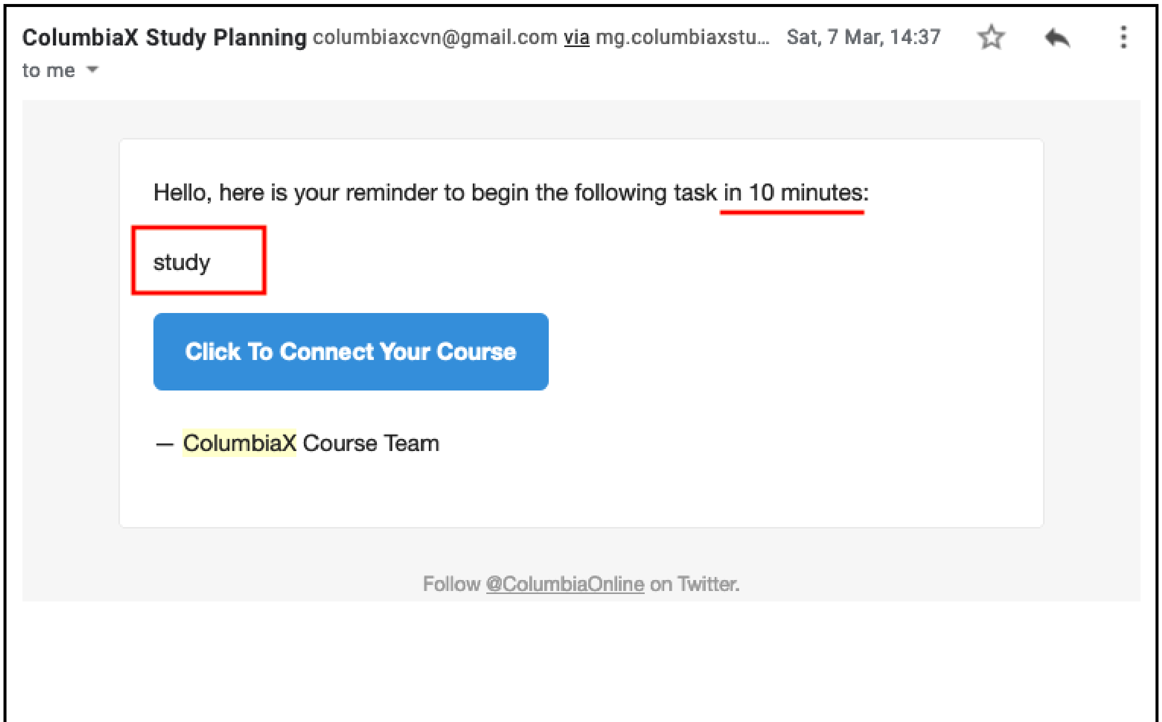


Figure 4.7. *Email Reminder Example*
This is an example email Cecil schedules repeatedly to send as a reminder for the study plan. The first line explains that in 10 minutes, the study should be starting. And the task was “study”.

Example learner from cluster d. Daarun (pseudonym) is an example learner from cluster D. Cluster D is, on average, the least active group of learners across all learning activities and their total grade is close to zero. Daarun enrolls in the BAMM102 Data Model and Decision course from week 3 and stays until week 6, then drops out. Daarun’s total score is 14 and his survival days are 21. Daarun interacts with SRLUI twice on goal setting and once on self-reflection.

In table 4.5, it shows that Daaron set his goal to study two week a time initially, and it seems that Daaron realizes the content requires more time to study and he alters his goal to study half week content maximum. Unfortunately, Daaron doesn’t continue with the course and leaves at week 6.

Interestingly, according to the course record, Daaron is also enrolled in BAMM101: Analytics in Python course at the same time. He starts BAMM101 from week 1 and completes projects and quizzes within 3 weeks and earns up to 76 grades. Daaron leaves the BAMM101 course in week 3. Since SRLUI rolls at week 5 so all learners who leave the course prior to week 4 are not included in the sample data. Therefore, Daaron's enrollment in BAMM101 is not included in this study. Still, from his learning activities in BAMM101, it seems that Daaron is a person who has a higher level of understanding of the course content prior to enrollment, so he is able to complete the quizzes and projects in 3 weeks and earn a passing grade. His ability to complete a course within 3 weeks can be inferred that he has a certain level of autonomy and self-regulated learning that he can execute and complete a course within a short period of time.

Using these two course performances and learning activities, Darron is an example to illustrate that grades and drop-out could be a poor proxy to indicate whether a learner is at risk in MOOC environments because researchers and educators could only collect partial information of learners from a MOOC. In other words, it is insufficient to judge a learners' competency based on a MOOC learning performance. It also points out this study's limitation for not including entire learners' trace data due to the implementation of SRLUI takes place on the 5th week. More discussions on the research findings will continue in Chapter 5.

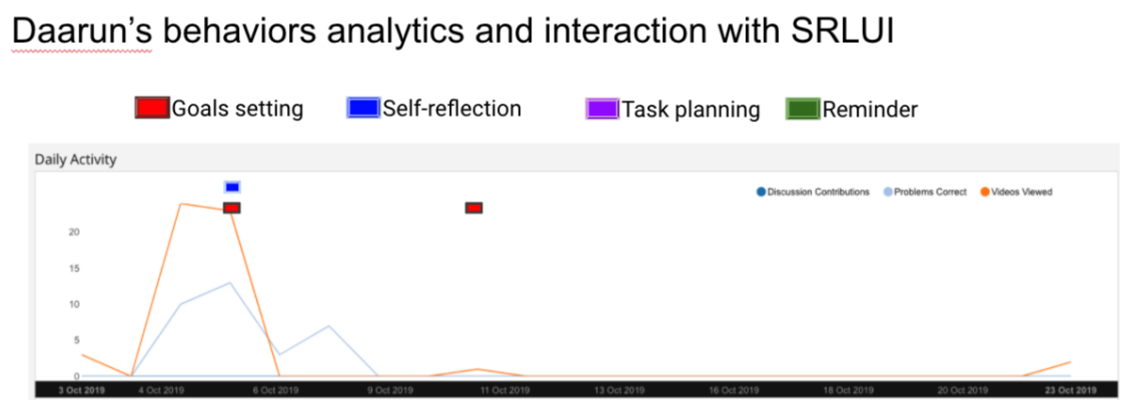


Figure 4.8. *Daarun's Behaviors Analytics and Integration with SRLUI.*
The x axis is the date of the log data and y axis represents the total number of log entries. The orange line indicates the number of video views and the light blue is the number of problems answered correctly. Interaction with SRLUI is indicated with four colored blocks, and each color represents a specific SRLUI feature. Goal planning is illustrated with red, self-reflection is blue, task planning is purple and reminders are green. The position of the blocks also corresponds with the createtime of the log entry.

Daarun is active in the course for 3 weeks, and he uses the self-reflection twice and the goal setting twice.

Table 4.5
Daarun Log Data with SRLUI on Goal Setting

Week	Goals
W3	Finish week 1 to 2
W4	Taking a holiday, 1/2 a week content max

RQ2: To what extent does the usage of SRLUI have an effect on learner persistence?

A prior MOOC study (Hsu, 2020) of 31 credential-seeking MOOC students suggests that learner persistence is a strong predictor of academic grades. Thus, the second research question investigates to what extent the usage of SRLUI may affect learner persistence by using survival analysis.

Since learners in the treatment group could choose to use or not to use SRLUI, an intent to treat (ITT) and treatment on the treated (TOT) effects are computed based on a variable useSRLUI: 1 for learners who used SRLUI at least once, 0 for did not use SRLUI at all. Similar issues also occur in medical research, economic or policy research while participants can self-select or decide to interact with the treatment which causes issues for researchers to make inferences for the average causal effect (ACE) (Geneletti & Dawid, 2007).

Survival functions for dropout are estimated and plotted for the control and the treatment groups in Figure 4.9. The survival function indicates that the control group has statistically significant higher survival rates than the treatment group ($p<0.05$) and over 50% of learners in both groups do not drop out. Specifically, the control group only has 30% of dropout rates compared to the treatment group (40%).

Figure 4.10 provides further evidence that the probability of the dropout of the control group and the treatment on the treated (TOT) are very similar to each other, and there is no significant difference between them ($p>0.05$). These results (Figure 4.8 and Figure 4.9) suggest that there are no observed effects of the usage of SRLUI on learner dropout rate. Overall, both the control and the treatment on the treatment (TOT) and the intent to treat (ITT) have demonstrated higher learning persistence (60%-70%) compared to the pilot study of the credential-based MOOCs (40%-50%) (Hsu, 2020).

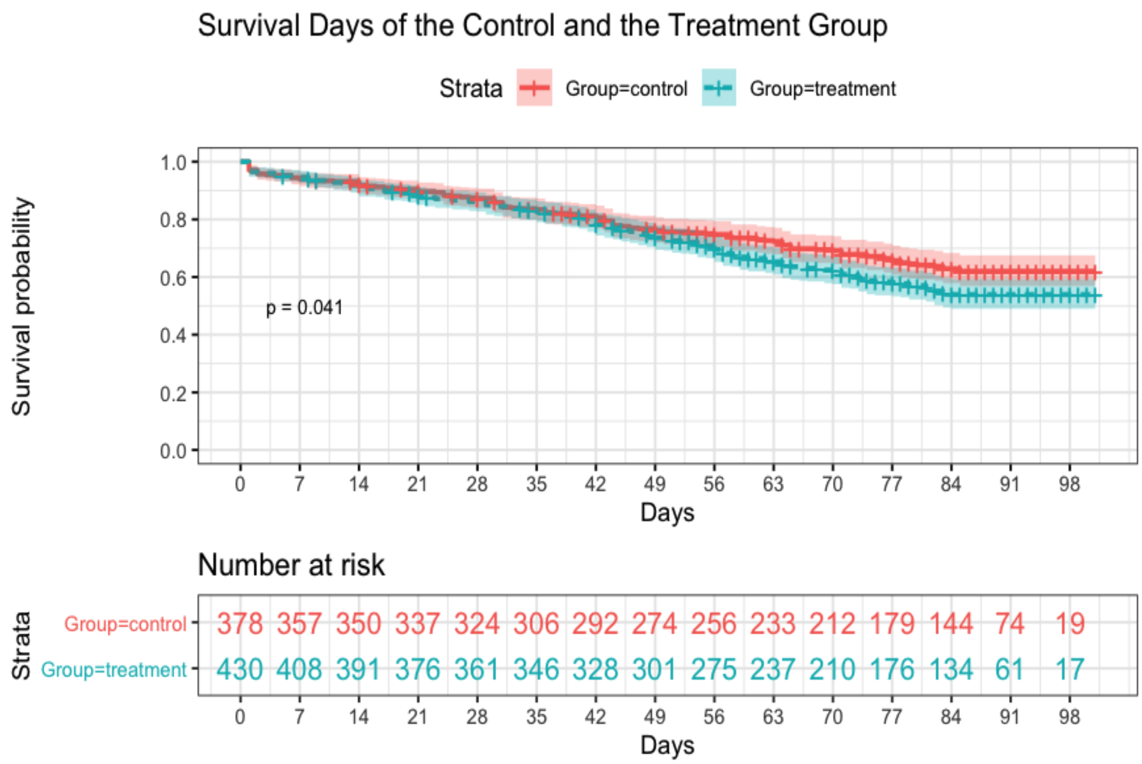


Figure 4.9. *Survival Days of the Control and the Treatment Group.*
The estimated survival function shows a graduate decline of survival rate. Overall, the control group (n=378) has a statistically significant higher survival rate than the treatment (n=430) (p<0.05). More than 50% of learners did not drop out for both groups. And the course is right-censored on day 98.

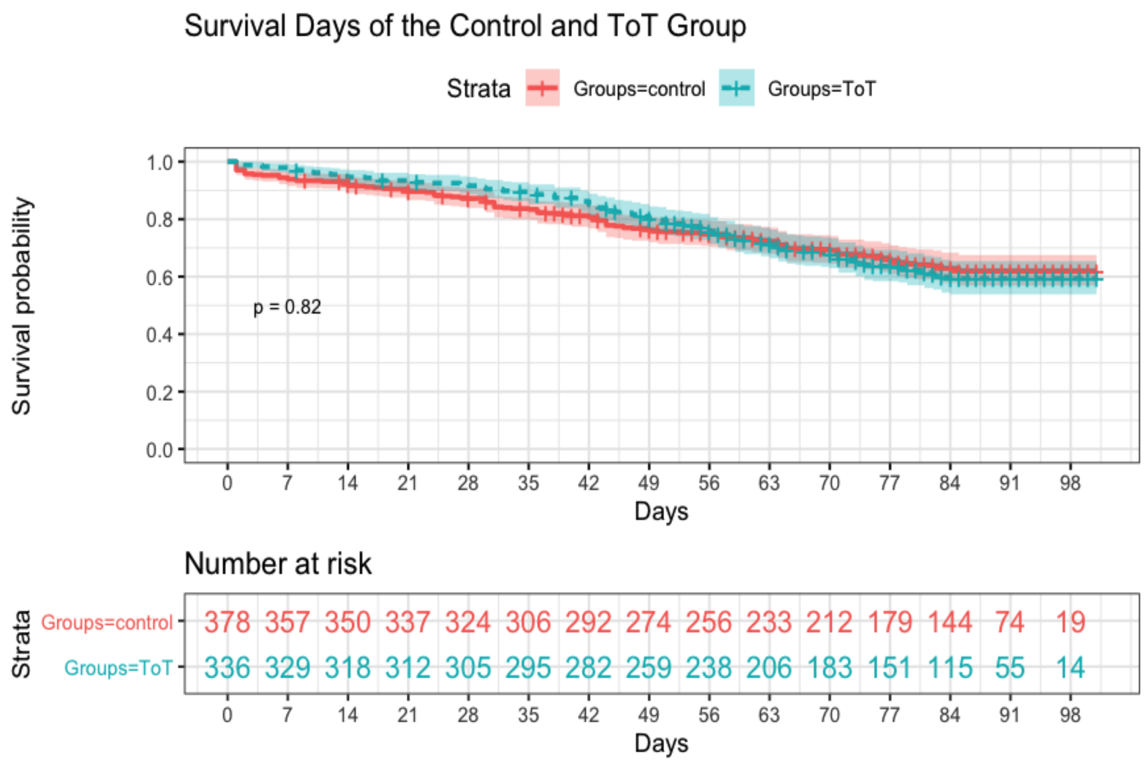


Figure 4.10. *Survival Days of the Control and the ToT Group.*
Survival function of the control group versus the treatment on the treated (TOT). There is no statistical significant difference between the two groups and the median survival time is greater than the observation window, meaning over 50% learners did not drop out.

RQ3: To what extent does the usage of SRLUI have an effect on learning outcomes?

The purpose of the 3rd research question is to investigate if SRLUI has any effect on learner's grades. To appropriately account for the nested nature of students within classes, a two-level hierarchical linear model (HLM) is utilized to model the SRLUI effect on learning outcomes.

Statistical Analysis

Descriptions of samples and variables coding. To assess the relationship of the usage of SRLUI on learning outcome, a subset of the sample with complete data on each of the variables was analyzed. The measures include learner's demographic data such as age, gender and educational levels in the subsequent models as independent variables. Since not all the learners in the treatment group interacted with SRLUI, useSRLUI variable is computed to signify intent to treat (ITT) and treatment on the treated (TOT) effects in a sequence HLM models. There are a total of 448 learners from 8 classes included in the RQ3 dataset.

Assumption testing. The initial inspection of the dependent variable, student grade, shows bimodal distribution with zero-inflation (n=108). Also, fitting an exploratory linear model to the bimodal outcome results in clear violation of normality and linearity of the residuals. As a result, the sample dataset is divided into two subgroups: a passing group (Grade>=60) and non-passing group (Grade<60) to proceed with the HLM analysis.

Effect size across eight courses. Figure 4.11 shows the effect sizes and the 95% confidence intervals of lower bound and upper bound of learning outcomes across eight courses. The effect size indicates the difference and the directions of treatment groups versus the control group. Since all the effect sizes are between -0.3 to 1, the differences between the treatment group and the control group are minimal within each course.

Course	95 % CI		Effect size
CSMM101 (n=235)	-0.02	-0.27	0.24
CSMM102 (n=150)	-0.05	-0.37	0.28
CSMM103 (n=58)	-0.27	-0.8	0.25
CSMM104 (n=40)	0.54	-0.11	1.19
BAMM101 (n=180)	0	-0.29	0.3
BAMM102 (n=46)	-0.11	-0.71	0.49
BAMM103 (n=34)	-0.31	-1.01	0.39
BAMM104 (n=65)	-0.03	-0.53	0.47

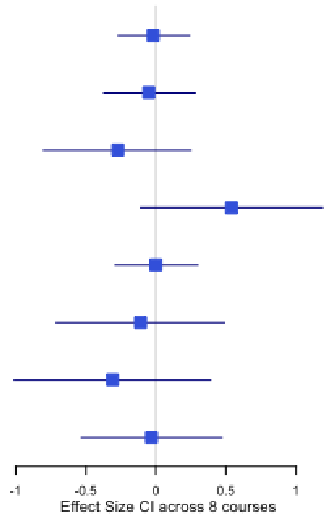


Figure 4.11. *Effect Size, Low and High Bound of CI.*
On the right hand side is a visual plot of learning outcomes across eight courses

Table 4.6 illustrates the HLM model result for the passing group. The intraclass correlation (ICC) in table 4.6 (the passing group) is 0.05, indicating that 5% of the variance in the grades is at the class level; while 95% of the variance in the grades is at the student level for the passing group. Table 4.6 shows, on average, the intent to treat (ITT) estimate of 2.5 points higher than the control group ($p<0.01$, effect size=.26) in the learning outcome; whereas the treatment on the treated (TOT) is 3.16 points higher ($p<0.01$) than the control group. Because the effect size is greater than .25, it could be considered substantively meaningful (What Works clearinghouse, 2011).

In contrast, similar results are not found in the non-passing group (See Table 4.7). Table 4.7 indicates there is no significant difference between the treatment group compared to the control group ($p>0.05$, effect size = -.18). Moreover, the trend indicates that learners receiving the treatment perform less than those in the control group. The intraclass correlation (ICC) is 0.05. It suggests that about 5% of the variance in the grades was at the class level and 95% of the variance in the grades was at the student level. The absolute value of effect size estimating the impact of SRLUI treatment is lower than .25, therefore, it should not be considered as substantively meaningful (What Works clearinghouse, 2011).

Table 4.6

Results of Passing Group (>=60) Hierarchical Linear Model

	ITT		TOT			
Student (n=247)	*Coeff.*	*SE*	*Coeff.*	*SE*	*p* value	ITT effect size
Intercept	75.89***	2.75	75.58***	2.70		
SRLUI treatment	2.49**	1.2	3.16**	1.19	0.008	0.25
Male	0.02	1.31	0.48	1.31	0.72	0.001
Age	0.11	0.08	0.1	0.07	0.19	0.01
Post_grad	0.64	1.39	0.72	1.4	0.60	0.07
Variance at						
Course Level-2	4.45		4			
Student Level-1	83.89		82.65			
ICC	0.05		0.05			

Note. ***:p<.001;**:p<.01;*:p<.05; ITT=intent to treat; TOT=treatment on the treated

Table 4.7

Results of Non-Passing Group (<60) Multilevel Hierarchy Model

Student (n=252)	IOT *Coeff.*	IOT *SE*	TOT *Coeff.*	TOT *SE*	*p* value	ITT effect size
Intercept	2.43***	0.27	2.32***	0.26	<0.001	
SRLUI treatment	-0.18	0.12	0.11	0.12	0.13	-0.18
Male	0.11	0.14	0.09	0.14	0.48	0.1
Age	-0.002	0.01	-0.002	0.06	0.74	-0.002
Post_grad	-0.15	0.14	-0.18	0.14	0.28	-0.15
Variance at						
Course Level-2	0.07		0.06			
Student Level-1	0.9		0.9			
ICC	0.08		0.07			

Note. ***:p<.001;**:p<.01;*:p<.05

Due to the 0-inflation issue, the non-passing group fails the assumption testing even after transforming it using the log function. However, after refitting the model with robust bootstrap, the standard error is consistent with the fitted model, thus I report the results of using the log function as the fitted model

V – DISCUSSION

Overview of Chapter

The purpose of this study was to create and implement a personalized self-regulated learning user interface (SRLUI) and assess its effects on learning persistence and learning outcomes in MOOCs environments. MOOCs were created to provide quality higher education to the world (Pheatt, 2017), but so far, only learners who excel in self-regulated activities have been successful in MOOC environments (Kizilcec & Halawa, 2015; Perez-Sanagustin et al., 2020). Despite the growing evidence that SRL abilities are a significant contributor to learning success in MOOCs, there have been four major challenges with developing and evaluating educational interventions meant to foster SRL skills in MOOC environments. First, to be cost-effective, historical SRL interventions used pre-course surveys as prompts to nudge learners towards implementing self-regulated learning strategies (Kizilcec et al., 2017a; Yeomans & Reich, 2017). However, these one-time-only, short-term SRL interventions showed limited or no results when reproduced at a larger scale (Kizilcec et al., 2020). The second challenge has arisen from the questionable validity of SRL measurements in prior SRL application studies. Historically, researchers have used self-reported surveys or learners' trace data to infer learners' SRL activities (Joksimovic et al., 2018). Relying on self-reported SRL activities has been criticized for being biased and inaccurate (Azevedo, 2014; Greene & Azevedo, 2007; Terras & Ramsay, 2014; Winne, 2017). Other studies measured SRL activities by making inferences based on learner's trace data[3], which yielded unique yet subjective metrics (Jansen et al., 2020; Kizilcec et al., 2017b; Min & Jingyan, 2017). For example, the act of revisiting an assessment could be labeled as goal setting, strategic planning, or self-evaluation if one uses Kizilcec et al.’s (2017b) framework. Jansen et al., (2020) also questioned whether these indicators can be interpreted as SRL processes based purely on SRL theories. Overall, the metrics used in prior studies were indirect measures of learners’ SRL activities based on surveys and trace data (Jansen et al., 2020; Kizilcec et al., 2017b; Min & Jingyan, 2017). The third major challenge in implementing SRL interventions has been that researchers often find many students do not end up using the interventions. Add to this the already high dropout rates that have always plagued MOOCs, and it becomes hard for researchers to collect data and assess the efficacy of the artifacts (Davis et al., 2018; Jansen et al., 2020). The last challenge hindering studies on SRL interventions is the strict nature of MOOC data privacy which limits researchers’ access to student data, particularly in credential-based MOOCs (Almeda et al., 2018).

SRLUI Design Rationale

The SRLUI was created in order to help students better regulate their coursework and improve their learning outcomes. SRLUI demonstrates an opportunity for course designers to create and implement a personalized user interface on a MOOC platform (see SRLUI architect in Appendix A). Unlike interventions used in other studies, this design is able to support and collect learner’s SRL activity data directly, making for a

---

[3] Trace data is a learner’s log data on the learning management system. The log data includes but not limited to timestamps, duration and the pages learners visit on the course, and the pages a learner clicks on.

more valid and accurate representation of a learner's SRL skills. This section will introduce the SRLUI design rationales from the implementation of SRLUI.

Built upon the framework of Zimmerman's (2000) self-regulated learning (SRL) theory, SRLUI was designed to support SRL behaviors through a longitudinal intervention provided throughout the course with a randomized experimental design. The design rationale of SRLUI accounts for the full cyclical process of Zimmerman's (2000) SRL model from forethought, performance, and self-reflection phases. Each week, learners were provided with recursive SRL support along with a dashboard that featured information on the past learning behaviors, a self-evaluation activity, a section for setting goals and planning tasks, as well as a section for reminders and learning tips. The email reminder function in SRLUI was created based on the findings of the Nudge to the Finish Line (2NFL) project, which successfully reduced at-risk students' attrition and increased completion rate by assisting students through an SMS messaging app (Mabel, Castleman & Bettinger, 2017). Thus, a nudging feature was included in the design of our SRL tool, allowing learners to schedule emails to be sent out reminding them of their planned learning tasks.

Findings

This section will discuss the four main findings of this study based on quantitative analysis: (1) the overall compliance rates of the SRL intervention is 80%, exceeding historical records; (2) no relationship was found between SRLUI and learners' persistence (3) a subgroup of learners was found to benefit from using SRLUI to achieve higher grades; (4) the quantitative tools used in this paper demonstrate rigorous and replicable methods for exploring MOOC learning behavior and performance data.

Improved compliance rates. Evaluating students' uptake of SRLUI is a reasonable way to assess the potential impact of a SRL application. Following the design principles suggested by Zimmerman's (2000) SRL model, SRLUI was embedded within eight MOOCs and used a content-specific and longitudinal intervention. In this study, the compliance rates accounted for the students who used SRLUI in the treatment group. Past literature (Davis et al., 2018; Jansen et al., 2020) suggested that the compliance rates of SRL tools was low (10%-30%). However, in this study, 80% of learners (n=430) provided with SRLUI intervention accessed the tools at least once. A follow-up investigation explored how learners used SRLUI to support their learning. Based on the descriptive data of sample learners, learners used SRLUI to (a) review and catch up with the learning schedule; (b) study content and work on assignments; (c) seek help in the discussion forum; (d) set motivational goals; (e) setup strategies for completing the course (i.e. completing easier assignments before the deadline).

Limited effects on learning persistence. Learning persistence has been reported to be a strong predictor to learning performance (Hus, 2020). To explore learner dropout, a Kaplin-Meier's survival function was used to calculate learner persistence based on their active days during the course period. The findings suggested there was no evidence that SRLUI improved learners' persistence. These results contradicted my expectations based on previous research (Hsu, 2020). I had hypothesized that learners with access to SRL tools would be less likely to drop out, but the data showed no relationship between access to SRL tools and dropout rates. To interpret the findings, other factors were explored based on the literature review to identify what may affect learner persistence. In the initial analysis, the average effect of SRLUI on learner persistence was examined.

However, past studies indicated that learner characteristics in conjunction with the learning context (e.g. formative feedback from the quizzes and projects along the course) may have influenced learner persistence (Chen et al., 2020; Greens et al., 2015). Prior studies also suggested that it was important to consider time-independent (e.g. learners' gender, age, educational level) and time-dependent factors (e.g. formative assessment feedback) on learner's academic performance and persistence (Bowers, 2010a; Chen et al., 2020; Willett & Singer, 1991). Specifically, students were more likely to complete the course if they already had an interest in the subject matter or had prior knowledge related to the course. (Chen et al., 2020; Coffrin et al., 2014).

For future studies, a more complex matrix that can control for other confounding factors is needed to better understand the efficacy of SRL interventions on student persistence (Murnane & Willett, 2010).

Unequal learning effects of SRLUI. Given that the learner's grades had a drastic bimodal distribution, the sample dataset had to be split into two groups, as required by the statistical model. One group were learners with a total grades of 60 or above, and the others were those who scored below 60. Using a hierarchical linear model, the results indicated that learners who achieved passing scores performed 2.5 points higher on average if they had access to SRLUI. In contrast, learners in the non-passing group showed no evidence of improved results even with access to SRLUI.

Turning to the data for further Insights in figure 4.1, learners' behavioral patterns were categorized into four subgroups of learners. Among them, the learner groups with passing grades (i.e. group A and group C) had higher participation rates across the number of videos viewed and averaged quiz grades than their peers with non-passing grades (i.e. group B and group D). It can be inferred that SRLUI may have unequally benefited subgroups of learners. In other words, SRLUI only helps improve grades for a subgroup of learners who are already active within the course. For future SRL interventions, it is suggested that the artifacts could provide more scaffolding to the learners at risk. For example, students who have lower SRL ability, less prior knowledge, or lower self-efficacy might need more direct assistance or training on how to use the intervention to support self-regulated learning (Lee et al., 2008).

Issues and Limitations

It is important to be conservative when interpreting the results of this study because of the following challenges. The first issue was the decision to exclude learners who left the course before week four since SRLUI was implemented from the fourth week until the end of the course. It was reasonable to exclude learners who could not be included in the study; however, that also created a certain level of bias in the sample population. Those first two weeks of classes were still a course-shopping period for MOOC learners (Ferguson & Clow, 2015), so excluding learners who left the course before week four automatically excluded early dropouts. That could be the reason why there were no significant findings that SRLUI could reduce the dropout rates. On the other hand, for design purposes, it also makes sense to only include subjects who have access to the intervention. In short, the study sample excluded early dropouts due to SRLUI being released in the middle of the course, which could create a bias in the sample population. Ideally, SRLUI would have been implemented from the first day of classes so that the entire class population could be involved.

The second challenge is that more data is needed to create a more robust method to measure the efficacy of SRLUI. As noted in the literature review in Chapter II, there were significant correlations between learner characteristics (e.g. learners' motivation and prior knowledge) and learning outcomes (e.g. learner persistence and grades) (Gardner & Brooks, 2018). However, these data were not collected in this study and the analysis couldn't account for these factors and their influence on learner persistence and grades exclusively. Therefore, more robust methods should be used to measure the effectiveness of SRL interventions, especially observational or quasi-experimental studies.

In this study, SRLUI was designed to support learners from a longitudinal perspective, based on the hypothesis that learning occurred progressively from the first day of the course until the end of the 12th week. However, learners in MOOCs might have different learning goals and trajectories. For example, learners who enrolled late might not find SRLUI useful because they tend to focus on participating in the graded assessments while skipping the lecture videos. In addition, there were other SRL tools available, such as planner apps or note-taking software which were not being tracked or considered in this study but may have been used by learners. Thus, not using SRLUI did not necessarily mean that learners were not engaging in SRL activities.

Another issue occurred during the data processing phase. In the dataset, the distribution of the learners' grades was bimodal. Specifically, there was a floor effect at zero on the grading scale. I decided to use grade 60 as a threshold to split learners into two groups for the following reasons: (a) the subgroups met the hypothesis testing; (b) there was no need to drop any data in the samples.

Implications for Future Design and Research

This study provides some insights for researchers and practitioners in terms of SRL intervention design, quantitative tools to analyze MOOC data, and recommendations for future research.

First, to better understand how an SRL artifact manifests SRL behaviors, it would be helpful to include granular data such as a learner's clickstream as a way to triangulate that learner's behaviors. Expanding the scope of analysis in this way could deepen the accuracy of the artifacts measuring SRL behaviors.

Secondly, SRLUI is designed based on the cyclical phase of Zimmerman's self-regulated learning model (2000) and considers each stage equally important. However, SRLUI did not specifically prompt learners to evaluate and revise their next phase learning goals. Instead, it simply presented the goals from the previous week as a reference. From a metacognitive perspective, the evaluation stage should be facilitated with self-reflection, which could lead to improved and better calibrated planning.

Future research could also continue to explore how to effectively support SRL activities in different subgroups. In particular, it would be helpful to consider if scaffolding or training is needed for learners who have low prior knowledge or who have lower self-efficacy. In addition, researchers could also consider using the quantitative research methods utilized in this study for other MOOC studies to see if the findings in this study hold true in other contexts, like in self-paced MOOCs. Specifically, future research could attempt to apply survival analysis with complex methods to confirm, add nuance to, or expand upon the findings in this study.

This dissertation primarily employed quantitative methods to analyze average treatment effects. However, there are many other data collected from learners' interactions with SRLUI, such as learners' engagement with goal setting and task planning, which can be analyzed either quantitatively or qualitatively. I aim to deploy a wider array of analyses, such as text analysis and social dynamic analysis, to examine the data of both verified track and audit track learners. In my future work, I also aim to continue updating SRLUI and employ the interface in future MOOCs to collect more data and gain additional insights. This paper demonstrates for researchers and instructional designers how to construct personalized, self-regulated learning tools on the edX learning management system. More research shall continue in this avenue to bring more insights into how to support the diverse populations in MOOCs.

Conclusion

Previous literature provides strong theoretical evidence for the affordance of self-regulated learning strategies on achieving educational goals. (Green et al., 2010; Newman, 2002; Puustinen and Pulkkinen, 2001; Zimerman, 2000). However, there is still scarce empirical evidence of effective SRL support and appropriate assessments of SRL activities in MOOCs (Jansen et al., 2020). Specifically, prior SRL applications in MOOCs include only one or a couple MOOCs in their studies (Joksimovic et al., 2018). Prior SRL experimental studies employ short-term, one-time only, pre-course surveys with an aim to create cost-effective interventions (Kizilcec & Cohen 2017; Kizilcec et al., 2017a; Yeomans & Reich, 2017); however, these studies do not produce statistically significant results when replicated with larger sample sizes (Kizilcec et al., 2020). Another issue with prior research is that it relies upon pre-course surveys or trace data, which cannot provide strong or clear evidence of cognitive or affective engagement, and therefore it is not representative of a learner's SRL behavior (Joksimovic et al., 2018; Winne, 2019).

This study contributes to MOOC literature by designing a personalized, content-related self-regulated learning user interface (SRLUI) with a randomized experimental design. SRLUI is integrated into the course, featuring a longitudinal, repeated support of the full cycle of SRL phases (i.e. goal setting, task planning, self-reflection, and evaluation) based on Zimmerman's SRL (2000) model. The sample courses include eight credential-based MOOCs. In terms of measurement, this study collected direct SRL behaviors data based on the SRLUI database to provide insights on learners' SRL strategies.

In addition, this study also contributes to MOOC literature by demonstrating that the majority of users with access to SRLUI used it at least once, and that a subset of users performed better when using SRLUI. However, there was no evidence found that access to SRLUI could reduce the number of dropouts.

REFERENCES

Alboukadel K., Marcin K. & Przemyslaw B. (2020).survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.8. https://CRAN.R-project.org/package=survminer

Al-Freih, M. (2017). *Examining the Relationship between Self-Regulated Learning Processes and Persistence to Goals in Massive Open Online Courses*. George Mason University. (Doctoral dissertation).

An, H., Shin, S., & Lim, K. (2009). The effects of different instructor facilitation approaches on students' interactions during asynchronous online discussions. *Computers & Education*, *53*(3), 749-760.

Arkorful, V., & Abaidoo, N. (2015). The role of e-learning, advantages and disadvantages of its adoption in higher education. *International Journal of Instructional Technology and Distance Learning*, *12*(1), 29-42.

Asunka, S. (2008). Online learning in higher education in Sub-Saharan Africa: Ghanaian University students' experiences and perceptions. *The International Review of Research in Open and Distance Learning, 9*(3)

Ayres, P. (2015). State-of-the-art research into multimedia learning: A commentary on Mayer's Handbook of Multimedia Learning. *Applied cognitive psychology, 29*(4), 631-636.

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning, 9*(2), 217–228.

Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools. *In New science of learning* (pp. 225-247). Springer, New York, NY.

Balakrishnan, G., & Coetzee, D. (2013). Predicting student retention in massive open online courses using hidden Markov models. *Electrical Engineering and Computer Sciences University of California at Berkeley, 53*, 57-58.

Barron, B. J., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: Lessons from research on problem-and project-based learning. *Journal of the Learning Sciences, 7*(3-4), 271-311.

Blackwell, L. S., Trześniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*,246–263.

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2019, June). Predict and Intervene: Addressing the Dropout Problem in a MOOC-based Program. *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*, 1-9.

Bowers, A. J. (2010a). Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *The Journal of Educational Research, 103*(3), 191–207.

Bowers, A.J. (2010b) Analyzing the Longitudinal K-12 Grading Histories of Entire Cohorts of Students: Grades, Data Driven Decision Making, Dropping Out and Hierarchical Cluster Analysis. *Practical Assessment, Research & Evaluation, 15*(7), 1-18.

Boyatt, R., Joy, M., Rocks, C., & Sinclair, J. (2014). What (Use) is a MOOC?. *The 2nd international workshop on learning technology for education in cloud* (pp. 133-145). Springer, Dordrecht.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1-13.

Cagiltay, N. E., Cagiltay, K., & Celik, B. (2020). An Analysis of Course Characteristics, Learner Characteristics, and Certification Rates in MITx MOOCs. *International Review of Research in Open and Distributed Learning, 21*(3), 121-139.

Chen, C., Sonnert, G., Sadler, P. M., Sasselov, D., & Fredericks, C. (2020). The impact of student misconceptions on student persistence in a MOOC. *Journal of Research in Science Teaching, 57*(6), 879–910.

Cobos, R., & Ruiz-Garcia, J. C. (2020). Improving learner engagement in MOOCs using a learning intervention system: A research study in engineering education. *Computer Applications in Engineering Education*. https://doi.org/10.1002/cae.22316

Columbia Video Network (n.d.). Columbia Video Network About Page. https://www.cvn.columbia.edu/content/columbia-video-network critical issue for the 21st century. Review of Educational Research, 70,151–179.

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of interactive Media in education, 2012*(3).

Davis, D., Chen, G., Hauff, C., & Houben, G. J. (2018). Activating learning at scale: A review of innovations in online learning strategies. *Computers & Education*, 125, 327-344.

DeBoer, J., Stump, G. S., Seaton, D., Ho, A., Pritchard, D. E., & Breslow, L. (2013). Bringing student backgrounds online: MOOC user demographics, site usage, and online learning. *Educational data mining 2013*.

Emmert-Streib, F., & Dehmer, M. (2019). Introduction to Survival Analysis in Practice. *Machine Learning and Knowledge Extraction*, *1*(3), 1013–1038.

Eriksson, T., Adawi, T., & Stöhr, C. (2017). "Time is the bottleneck": a qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education*, *29*(1), 133–146.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, *6*(4), 50–70.

Fournier, H., & Kop, R. (2010, July). Researching the design and development of a Personal Learning Environment. In *PLE Conference* (pp. 6-8).

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127–203.

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *The American journal of distance education, 19*(3), 133-148.

Greene, D. (2013). Learning analytics for smarter psychological interventions. *MOOC Research Initiative.*

Greene, J. A., & Azevedo, R. (2007). A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. Review of *Educational Research, 77*(3), 334–372.

Greene, J. A., Costa, L.J., Robertson, J., Pan, Y., & Deekens, V. M. (2010). Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment. *Computers & Education, 55*(3), 1027–1043. https://doi.org/10.1016/j.compedu.2010.04.013

Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of Retention and Achievement in a Massive Open Online Course. *American Educational Research Journal, 52*(5), 925–955. https://doi.org/10.3102/0002831215584621

Greene, J. A., Robertson, J., & Costa, L. J. C. (2011). Assessing self-regulated learning using think-aloud methods. *Handbook of self-regulation of learning and performance*, 313-328.

Gunawardena, C. N., & McIsaac, M. S. (2013). Distance education. In Handbook of research on educational communications and technology (pp. 361-401). Routledge.

Harasim, L. (1993). Building a Global Network: The WBSI Experience. *Global networks: Computers and international communication*, *185*.

Harasim, L. (2017). *Learning theory and online technologies*. Routledge.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational research review*, *12*, 45-58.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of educational research*, *70*(2), 151-179.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Hsu, S. Y. (2020) An Exploratory Study of Learner Characteristics, Learning Behavior and Learning Outcomes in MicroMasters MOOC Programs. *Proceedings of Association for Educational Communications & Technology, 1*.

Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in Massive Open Online Courses. *Computers & Education*, *146*, 103771.

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How Do We Model Learning at Scale? A Systematic Review of Research on MOOCs. *Review of Educational Research, 88*(1), 43–86.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), 457-481.

Kentnor, H. E. (2015). Distance education and the evolution of online learning in the United States. *Curriculum and Teaching Dialogue*, *17*(1), 21-34.

Khan, B. (2001). *Web-based training.* Educational Technology Publications.

Kizilcec, R. F., & Cohen, G. L. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences, 114*(17), 4348–4353.

Kizilcec, R. F., Davis, G. M., & Cohen, G. L. (2017a). Towards Equal Opportunities in MOOCs: Affirmation Reduces Gender & Social-Class Achievement Gaps in China. *Proceedings of the Fourth ACM Conference on Learning @ Scale*, 121–130.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *Proceedings of the second ACM conference on learning@ scale* (pp. 57-66). ACM.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017b). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education, 104*, 18–33.

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences, 117*(26), 14900–14905.

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause review*, *48*(3), 62-63.

Kolowich, S. (2013). The professors who make the MOOCs. *The Chronicle of Higher Education, 18.*

Lan, M., & Hew, K. F. (2020). Examining learning engagement in MOOCs: A self-determination theoretical perspective using mixed method. *International Journal of Educational Technology in Higher Education*, *17*(1), 7.

Lan, M., & Lu, J. (2017). Assessing the Effectiveness of Self-Regulated Learning in MOOCs Using Macro-level Behavioural Sequence Data. In *EMOOCs-WIP* (pp. 1-9).

Lee, J. E., Recker, M., Bowers, A., & Yuan, M. (2016, June). Hierarchical Cluster Analysis Heatmaps and Pattern Analysis: An Approach for Visualizing Learning Management System Interaction Data. In *EDM* (pp. 603-604).

Lee, T. H., Shen, P. D., & Tsai, C. W. (2008). Enhancing computing skills of low-achieving students via e-learning: A design experiment of web-based, problem-based learning and self-regulated learning. *Cyberpsychology & Behavior*, *11*(4), 431-436.

Lewin, A., Jöbges, M., & Werheid, K. (2013). The influence of self-efficacy, pre-stroke depression and perceived social support on self-reported depressive symptoms during stroke rehabilitation. *Neuropsychological rehabilitation*, *23*(4), 546-562.

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, *45*, 100727.

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education*, *29*, 40-48.

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning*, *14*(3), 202. https://doi.org/10.19173/irrodl.v14i3.1455

Mayer, R. E. (2005). Introduction to multimedia learning. *The Cambridge handbook of multimedia learning, 2*, 1-24.

McCoach, D. B., & Adelson, J. L. (2010). Dealing With Dependence (Part I): Understanding the Effects of Clustered Data. *Gifted Child Quarterly, 54*(2), 152–155.

Meyer, K. A. (2002). Quality in Distance Education. *ERIC Digest*.

Meyer, K. A. (2003). Face-to-face versus threaded discussions: The role of time and higher-order thinking. *Journal of Asynchronous Learning Networks*, *7*(3), 55-65.

Miller, G., Benke, M., Chaloux, B., Ragan, L. C., Schroeder, R., Smutz, W., & Swan, K. (2013). *Leading the e-learning transformation of higher education: Meeting the challenges of technology and distance education*. Stylus Publishing, LLC.

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same?. *The Internet and Higher Education*, *14*(2), 129-135.

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research.* Oxford University Press.

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the foundations of positive psychology* (pp. 239-263). Springer, Dordrecht.

Newman, R. S. (2002). How Self-Regulated Learners Cope with Academic Difficulty: The Role of Adaptive Help Seeking. *Theory Into Practice, 41*(2), 132–138. https://doi.org/10.1207/s15430421tip4102_10

Ng, A., & Widom, J. (2014). Origins of the modern MOOC (xMOOC). *Full Report,* 34-47. http://www.cs.stanford.edu/people/ang/papers/mooc14 - OriginsOfModernMOOC.pdf

Nichols, M. (2003). A theory for eLearning. *Journal of Educational Technology & Society*, *6*(2), 1-10.

Nipper, S. (1989). Third generation distance learning and computer conferencing. *Mindweave: Communication, computers and distance education*, 63-73.

O'Connell, A. A., & McCoach, D. B. (2008). *Multilevel modeling of educational data*. IAP.

Pérez-Sanagustín, M., Sapunar-Opazo, D., Pérez-Álvarez, R., Hilliger, I., Bey, A., Maldonado-Mahauad, J., & Baier, J. (2020). A MOOC-based flipped experience: Scaffolding SRL strategies improves learners' time management and engagement. *Computer Applications in Engineering Education*.

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students' self-regulated engagements with computerized content. *Educational Psychology Review, 18*(3), 211-228.

Pheatt, L. E. (2017). *The pursuit of profit or prestige: What the diffusion of MOOCs can tell us about disruptive innovation in US higher education.* Doctoral dissertation, Columbia University.

Pintrich, P. R. (2000a). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic.

Pintrich, P. R. (2000b). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*(1), 92-104.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95,*667–686.

Puustinen, M., & Pulkkinen, L. (2001). Models of Self-regulated Learning: A review. *Scandinavian Journal of Educational Research, 45*(3), 269–286. https://doi.org/10.1080/00313830120074206

R Core Team (2020). R: A language and environment for statistical computing.

Reich, J. (2014). MOOC completion and retention in the context of student intent. *EDUCAUSE Review Online, 8*.

Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education* (pp. 45-58). Harvard University Press Review.

Rhode, J. F. (2009). Interaction equivalency in self-paced online learning environments: An exploration of learner preferences. *The International Review of Research in Open and Distance Learning, 10*(1).

Romesburg, H. C. (1984). *Cluster analysis for researchers*. Lifetime Learning Publications: Belmont Calif, 128-135.

Salomon, G. (1984). Television is" easy" and print is" tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*(4), 647.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and learning*, *4*(1), 33-45.

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology*, *73*(1), 93.

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 85-94.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, *2*(1), 3-10.

Simmons, D. E. (2002). The forum report: E-learning adoption rates and barriers. In A. Rossett (Ed.), *The ASTD e-learning handbook* (pp. 19–23). New York: McGraw-Hill.

Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.

Spector, J. M., Merrill, M. D., Elen, J., & Bishop, M. J. (Eds.). (2014). *Handbook of research on educational communications and technology* (pp. 413-424). New York, NY: Springer.

Sumner, J. (2000). Serving the system: A critical history of distance education. *Open Learning: The Journal of Open, Distance and e-Learning*, *15*(3), 267-285.

Terras, M. M., & Ramsay, J. (2014). A Psychological Perspective on the Temporal Dimensions of E-Learning. *E-Learning and Digital Media*, *11*(2), 108–122.

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, *46*(3), 472-487.

Therneau T (2020). A Package for Survival Analysis in R. R package version 3.2-7, <URL: https://CRAN.R-project.org/package=survival>.

Willett, J. B., & Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of educational research*, *61*(4), 407-450.

Winne, P. H. (2013). Learning strategies, study skills, and self-regulated learning in postsecondary education. *In Higher education: Handbook of theory and research* (pp. 377-403). Springer, Dordrecht.

Winne, P. (2017). Leveraging big data to help each learner and accelerate learning science. *Teachers College Record, 119*(3), 1-24.

Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the Internet. *In International handbook of metacognition and learning technologies* (pp. 293-308). Springer, New York, NY.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551-572.

Winne, P. H., Nesbit, J. C., Kumar, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R., & Perry, N. E. (2006). Supporting self-regulated learning with gStudy software: The Learning Kit Project. *Technology Instruction Cognition and Learning, 3*(1/2), 105.

Winne, P. H., Nesbit, J. C., & Popowich, F. (2017). nStudy: A system for researching information problem solving. *Technology, Knowledge and Learning, 22*(3), 369-376.

Winne, P. H., Teng, K., Chang, D., Lin, M. P. C., Marzouk, Z., Nesbit, J. C., & Vytasek, J. (2019). nStudy: Software for learning analytics about processes for self-regulated learning. *Journal of Learning Analytics, 6*(2), 95-106.

Yeomans, M., & Reich, J. (2017). Planning prompts increase and forecast course completion in massive open online courses. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 464–473.

Yousef, A. M. F., Chatti, M. A., Schroeder, U., & Wosnitza, M. (2014). What drives a successful MOOC? An empirical examination of criteria to assure design quality of MOOCs. *Proceedings of 2014 IEEE 14th International Conference on Advanced Learning Technologies* (pp. 44-48). IEEE.

Zhang, Z., & Kenny, R. (2010). Learning in an online distance education course: Experiences of three international students. *The International Review of Research in Open and Distance Learning, 11*(1).

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology, 25*(1), 82-91.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*, 64–70

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling, 53*(1), 141-160.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.

Appendix A: SRLUI Architecture

The following section will explain the architecture of SRLUI. edX utilized XBlocks (Figure 1), a component architecture based on Python, HTML, JavaScript and CSS, to build its learning activities such as videos and quizzes. This feature allows course designers to implement a customized interface using XBlocks and programming logic (such as the SRLUI) through edX's RAW HTML input elements (see Figure 2).

Figure 1: XBlock in edX platform

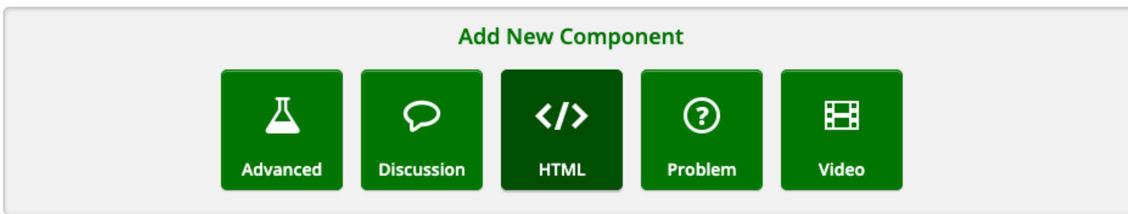
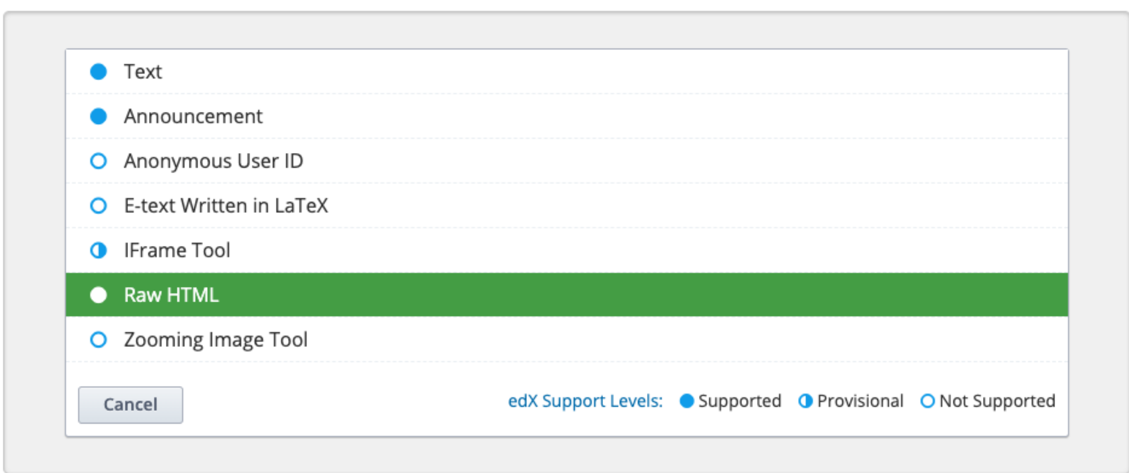


Figure 2: Raw HTML option in XBlock



SRLUI structure can be divided into the front-end and back-end. The front-end, which is built on the edx page, has two major functionalities: (1) a tracking function which monitors learners' interactions with SRLUI as well as learning activities such as video watching and quiz submissions. (2) a display function is designed to show three pages of information, which include (a) a course progress page (b) study planning page and (c) study tips page. The back-end hosts learners' log data of interactions with SRLUI and updates the front-end by using customized and secured APIs.

In the following section, I will describe the front-end and back-end in more detail with illustrations.

Front-end

Tracking functions Javascript is coded as an XBlock on edX to detect and record learners' video watching activity and quiz attempts. edX log data was not used because it has a 24-hour interval and uses different logic. For example, SRLUI counts a video watching event with a minimum of viewing time of 30 seconds (not including pause or stop actions), whereas edX only requires 5 seconds.

Display functions SRLUI builds customized XBlocks with HTML, JavaScript and CSS scripts to display its user interface on edX. Meanwhile, an HTML file was coded to assign learners to either treatment or control group environments. Specifically, a learner with an even ID number is assigned to a control group while a learner with an odd number ID was assigned to an treatment group. SRLUI also uses the following libraries and widgets to enhance its interface:

1. *CanvasJS.chart* utilized to display the line graph on the progress page.
2. *Mailgun*: provides a customized domain from which to send reminder emails.
3. *Agenda*: helps manage the schedule for sending out reminder emails.
4. *Mdtimepicker*: provides a widget for learners to input time for task planning.

Back-end

To store and retrieve learners' interactions with SRLUI, an HTTPs server was built using Node.js in a MongoDB database. The SRLUI server also tracks and stores learners' participation in the discussion forum by using edstem's API. Additionally, customized and secured APIs are developed to allow the following tasks and present the following information:

1. Self-evaluation of learning goals
2. Goal settings
3. Posting and viewing counts of discussion forum activity
4. Video watching and problem attempts
5. Notifications of task reminders