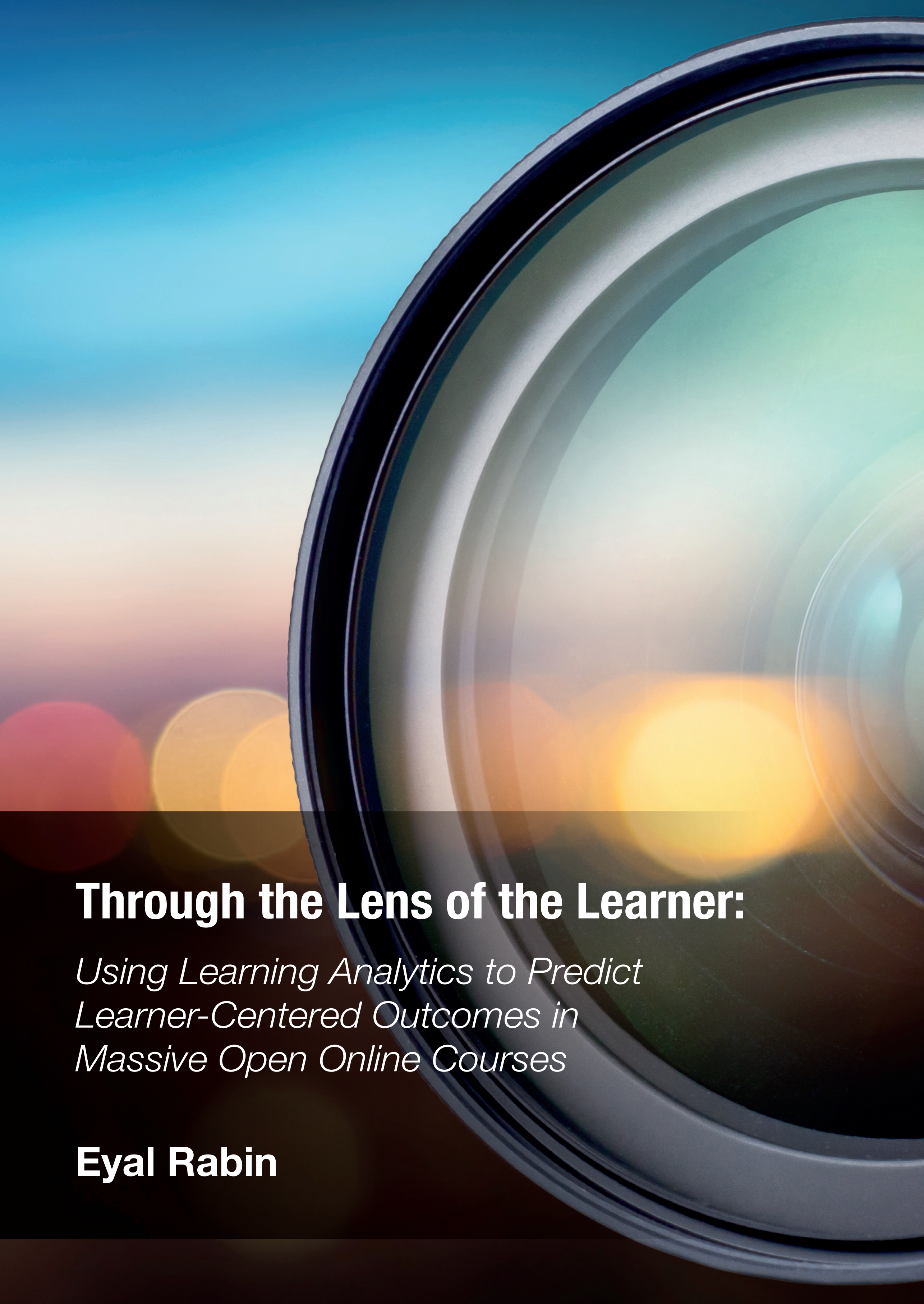
# Through the Lens of the Learner: Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses'



# Through the Lens of the Learner:

*Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses*

**Eyal Rabin**

# Through the Lens of the Learner:

# Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses

Eyal Rabin

The research reported in this thesis was carried out at the UNESCO Chair on Open Education at the Faculty of Management of the Open University in the Netherlands.

Open Universiteit
www.ou.nl



Design and printing by: ProefschriftMaken || proefschriftmaken.nl

ISBN: 978-94-6423-343-8

# Through the Lens of the Learner:
# Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. dr. Th. J. Bastiaens
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 10 september 2021 te Heerlen
om 13:30 uur precies

door
Eyal Rabin
geboren op 7 september 1977 te Petah-Tikva, Israël

**Promotores:**
Prof. dr. M. Kalz, Open Universiteit / Heidelberg University of Education
Prof. dr. Y.M. Kalman, The Open University of Israel

**Leden beoordelingscommissie:**
Prof. dr. C. Delgado-Kloos, Universidad Carlos III Madrid
Prof. dr. Y. Eshet-Alkalai, The Open University of Israel
Prof. dr. M. Scheffel, Ruhr-Universität Bochum
Prof. dr. R. Klemke, Open Universiteit

***“Education is the most powerful weapon which you can use to change the world”***

Nelson Mandela.

# Table of contents

# 1

# Chapter 1

## Introduction

## Through the Lens of the Learner: Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses

## General Introduction

In the digital era, technology is leading to massive changes on multiple fronts – in the economy and professional settings; in the way we communicate and relate to each other; and increasingly, in the way we learn (Bates, 2015). These changes have rapidly intensified during the past few months, due to the Covid-19 pandemic. As a result of the changes in the volume and significance of knowledge in today's digital society, people can no longer rely solely on the knowledge accumulated throughout their primary, secondary, and tertiary education, and the first stages of employment. Full participation in today's knowledge-based society requires people to become lifelong learners. The ability to learn and adapt to new skills is increasingly important in our ever-changing technological universe (OECD, 2007). This means that learning can no longer be divided into specific places and times of knowledge acquisition (school) and places and times of knowledge application (the workplace) (Fischer, 2000).

The changes described above, fueled by digital innovation, have also affected the higher education system, as new categories of educational actors have appeared. Our educational institutions were designed and built mainly in the industrial era rather than in the digital era (Bates, 2015). According to the Europe 2020 strategy report, a fundamental transformation of education and training is needed to address the skills and competencies required for Europe to remain competitive, overcome the current economic crisis, and grasp new opportunities (Commission European, 2020).

The changes in the digital era lead to changes in the learning environment and will produce changes in the methods used by educational institutions to teach. As can be seen, during the Covid-19 pandemic, almost all HEIs had to move to online learning (UNESCO, 2020). One of the markers for these changes has been the rapid rise of MOOCs. The concept of "MOOC" was coined in 2008 by Dave Cormier at the University of Prince Edward Island and Bryan Alexander of the National Institute for Technology in Liberal Education in response to an open online course designed and led by George Siemens at Athabasca University and Stephen Downes at The National Research Council (Canada) (Downes, 2012; Liyanagunawardena, Parslow, & Williams, 2013). The word MOOC is an acronym for Massive Open Online Course, which describes the basic attributes of the concept – an online course designed for unlimited participation and open access via the web (Kaplan & Haenlein, 2016). By the end of 2019, more than 110 million students participated in over 13.5 million courses provided by more than 900 universities around the globe (Shah, 2019). During the first half of 2020, interest in MOOCs had drastically grown (Shah, 2020).

The MOOC phenomenon was preceded by the movement to promote open educational resources (OER) in the 1990s and the publication of teaching materials as open content, initiated with the launch of the OpenCourseWare (OCW) project at the Massachusetts Institute of Technology (MIT) in 1999 (Abelson, 2008). The OER movement, fueled by the Internet expansion, grew in parallel to the evolution of digital distance education (DE), the expansion of Open Universities (OU), and other distance teaching universities (DTU) around the globe.

OERs, as defined by UNESCO (2002) are "teaching, learning, or research materials that are in the public domain or released with an intellectual property license that allows for free use, adaptation, and distribution." In 2001, MIT announced that nearly all its courses would be freely accessible to anyone on the Internet via OpenCourseWare (OCW) (Maria et al., 2014). In 2002, the number of institutions offering free or open courseware increased, and many universities worldwide started to offer open access to their course materials (Vladoiu, 2011). It is still assumed that OER not only makes high quality higher education available at low cost to a large number of users, but that it will also lead to innovation within universities (Mulder, 2015). However, it was still unclear as to how this leap could be accomplished in practical terms. Universities and other enterprises involved in MOOCs had not yet consolidated their business models (Reich & Ruipérez-Valiente, 2019). The emergence and spread of MOOCs brought the OER movement to a new stage. Although MOOCs are not strictly OER, since the resources provided rarely encourage adaptation or re-mix and are not always published under an open license, they epitomize an unprecedented move towards greater accessibility of higher education at no (or low) additional cost (Maria et al., 2014).

Other phenomena that promoted distance education and enabled open access to higher education was the establishment of the Open University in the United Kingdom (OUUK) in 1969. Based on the OUUK model, many open universities were established worldwide. For example, the Open University in the Netherlands and the Open University of Israel. The basic goal of OUs is to provide opportunities for admission to higher education without prerequisites (Tait, 2008). The open universities made significant changes to the landscape of higher education by enabling open access and modular credit accumulation. They reached out to part-time adult students and harnessed innovative technologies to improve their teaching/learning processes. Faculty were provided opportunities to work in teams to develop study materials and to teach high numbers of students (Guri-Rosenblit, 2019).

As mentioned above, the first MOOC was taught by George Siemens and Stephen Downes in 2008. The course, called Connectivism and Connective Knowledge, had over 2,200 participants. It was based on principles from connectivity pedagogy which recommends that learning materials should be aggregated (rather than pre-selected), remixable, repurposable, and targeted toward future learning. The instructional design approach of the course designers focused on connecting learners to each other in order to answer questions or to collaborate on joint projects, which emphasized the collaborative development of the MOOC (Downes, 2012). The success of the connectivist MOOCs led to the second wave of MOOCs, led by Sebastian Thrun and Peter Norvig, who presented their first extension MOOC (xMOOC) in 2011 (Martin, 2012). This course, which had a much more traditional course structure, was characterized by the specified aim of completing the course and obtaining certification in the subject matter. The second wave courses had a specified syllabus of recorded lectures and self-test problems. The course instructor was the expert provider of knowledge, with student interactions usually limited to asking for assistance and advising each other on difficult points (Van den Beemt, Buijs, & Van der Aalst, 2018). To differentiate between these two types of MOOCs, they were then called cMOOCs and xMOOCs respectively (Downes, 2012).

Since most of the MOOCs today are following the xMOOC paradigm, this dissertation focuses on xMOOCs, rather than cMOOCs and will use the general term MOOC to refer to xMOOCs and other variants of MOOCs which do not necessarily fit into the original distinction.

As mentioned previously, MOOCs were originally proposed as a way to bring high-quality tertiary education closer to populations with limited resources, thereby overcoming economic, geographic, or time barriers. MOOCs enabled learners with different academic backgrounds to experience technology-enhanced learning anywhere and anytime, almost free of charge. The courses amplified 21st-century skills (Friedman, 2012) and enabled people to learn throughout their lives and become lifelong learners (Kalz, 2015). MOOCs also have a considerable impact on many different levels of higher education institutions (Little, 2016). In a survey conducted among HEIs in Europe and Canada, it was found that HEIs expect that MOOCs will affect different populations at the institutions, mainly the online/distance students, the academic staff, and part-time students (Jansen, Schuwer, Teixeira, & Aydin, 2015). As the coronavirus pandemic continues, universities have accelerated their use of MOOCs content in their teaching. The leading MOOCs providers, such as Coursera and EdX, announced in March 2020 that any college impacted by the coronavirus, even if not a partner of the company, could request free access to their course catalogs (Young, 2020).

MOOCs served as laboratories for experimentation. They expanded the boundaries of possibility offered by online courses and gained the confidence of users, who saw that the technology infrastructure was in place to support thousands of learners, even within a single course. However, the development of MOOCs has not been without controversy. The high potential of MOOCs has been critiqued on two main grounds. Firstly, most of the students who earn certificates for completing the MOOCs are experienced learners with strong academic backgrounds from developed countries (Christensen et al., 2013; Daily, 2014; Emanuel, 2013; Guo & Reinecke, 2014; Koller, Ng, Do, & Chen, 2013; Laurillard, 2016; Reich & Ruipérez-Valiente, 2019). These participants typically enroll in order to keep their knowledge updated and to develop their professional skills, improve their work performance, or change career (Castaño-Muñoz, Kreijns, Kalz, & Punie, 2017; Greene, Oswald, & Pomerantz, 2015; Liu, Kang, & McKelroy, 2015; Loizzo & Ertmer, 2016; Schmid, Manturuk, Simpkins, Goldwasser, & Whitfield, 2015). However, evidence for employment mobility after participating in MOOCs remains limited (Dillahunt, Ng, Fiesta, & Wang, 2016).

A second criticism of MOOCs is the high dropout rates (Gardner & Brooks, 2018; Reich & Ruipérez-Valiente, 2019) and the vast majority of MOOC learners who never return after their first year as learners (Reich & Ruipérez-Valiente, 2019). These rates are, on average, 93 percent (Chuang & Ho, 2016; Jordan, 2014; Margaryan, Bianco, & Littlejohn, 2015). Although MOOC dropout rates might indeed be high, the question is whether the completion rate is the appropriate measure for evaluating the success of this new form of lifelong learning. Completion rate is a success criterion borrowed from formal education contexts where students enroll in courses with the goal of completing them, and of achieving the learning outcomes defined by the instructor (Henderikx, Kreijns, & Kalz, 2017).

In sharp contrast to these formal educational contexts, participants in open education learning environments such as MOOCs may have diverse goals and expect a variety of different learning outcomes (Kalz et al., 2015). Participants register for MOOCs to explore new ways of learning, experiment with online interaction, seek entertainment, try to meet a personal challenge, or simply for the enjoyment of learning. Only some participants register with the desire to earn a certificate of completion or other formal recognition of their achievements (Hew & Cheung, 2014; Littlejohn, Hood, Milligan, & Mustain, 2016; Liu et al., 2015; Macleod, Haywood, Woodgate, & Alkhatnai, 2015; Onah, Sinclair, & Boyatt, 2014; Wang & Baker, 2018). Some participants might begin a MOOC to figure out whether a particular topic might be worth pursuing and whether they would like to listen to one or more lectures, while others intend on completing all the course material (Koller et al., 2013; Reich, 2014). Henderikx et al. (2017) and Schmid et al. (2015), showed that some participants make selective use of the course materials by using only the content which is of interest to them. Consequently, it has been shown that some participants achieved their learning goals by engaging only in segments of the course (Ho et al., 2015; Liyanagunawardena et al., 2013).

Due to the criticism, it has been proposed that the success of lifelong learning in MOOCs should be evaluated, not through traditional, instructor-focused measures such as completion rates or certificate earning, but rather through more learner-centered outcomes such as learner satisfaction and the fulfillment of learner intentions (Henderikx, Kreijns, & Kalz, 2017; Kalz, 2015; Reich, 2014).

Research on the relationship between the learners' intentions to study MOOCs and their learning outcomes has gained many valuable insights from the participants' digital footprints. Learners in digital environments leave a huge number of digital footprints collected into log-files. The amount of digital trace data that is created in the MOOCs' learning platforms is huge. For example, the 290 courses offered by MIT and Harvard in the first four years of edX produced 2.3 billion logged events from 4.5 million learners (Chuang & Ho, 2016). These increases in available educational data and learning analytics techniques have become a powerful means of informing and supporting learners, teachers, and their institutions in order to better understand and predict personal learning needs and performance (Greller & Drachsler, 2012). The research of learning analytics in MOOCs makes it possible to examine fundamental questions about teaching and learning (Reich, 2015). One such question, which will be dealt with in the next section, focuses on the nature of learner-centered outcomes in MOOCs.

## The Rise of Learning Analytics

Learning analytics (LA) has been defined as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens et al., 2011). Ferguson (2012) included learners' footprints in online learning environments, together with their personal data, interaction data, and academic information as the necessary source of LA. The field of LA is closely related to other fields such as web analytics, educational data mining, academic analytics, and business intelligence (Elias, 2011). The term "learning analytics" attracted the attention of researchers around 2010, with 117 academic publications using it. The term rapidly gained popularity, and in 2018, approximately 6,000 academic publications referred to LA (Winer & Geri, 2019).

The interest in LA stems mainly from the belief that its effective use can lead to improved educational institutional decision-making. Learning analytics advocates also argue that LA leads to advancements in learning outcomes for at-risk students, greater trust in institutions, due to the disclosure of data, and significant improvements in pedagogy (Akçapınar, Altun, & Aşkar, 2019; Siemens & Long, 2011). Although LA has indeed enabled researchers and universities to open the "black box" of education by tracking, aggregating, and analyzing student profiles and digital traces of behaviors captured in information systems in the structure of log-files, these techniques have raised several privacy concerns (Jones, 2019).

The footprints left by MOOC participants enable us to collect the data in log-files, analyze it, and gain significant insights into the learning process using LA (Siemens & Long, 2011). In addition, analyzing data logs of MOOC participants by tracking, measuring, and interpreting learner behavior enables researchers to examine entire populations of learners unobtrusively, rather than taking a sample of the population. The research can be carried out without the limitations of cost, time, authenticity of data, selection bias, or response bias.

However, using learner log-files also has several disadvantages. These methods, used alone, cannot provide information about the psycho-social dispositions of MOOC participants, such as their pre-course motivation and intentions, as well as their learner-centered outcomes; for example, their satisfaction and the fulfillment of their intentions at the end of the course. Pre- and post-course attitudes such as these are usually measured through questionnaires and other reactive measures, which can also have disadvantages when used solely as a basis for research. Measures such as surveys, interviews, focus groups, and observations are time-consuming and limited when conducted at scale (Goggins & Xing, 2016; Xing, Kim, & Goggins, 2015). Besides, data based on self-reports has limited validity, can suffer from social desirability bias, low response rates that lead to sample bias, and do not measure actual learning behavior and the effects of real-time intervention. The triangulation of methods using pre- and post- learner surveys combined with data based on log-file analysis, can provide advantages which benefit researchers and overcome some of the disadvantages of other methods used in isolation (Reich, 2015). The combined use of pre- and post-learner surveys with clickstream data can provide us with valuable educational insights into the correlations between the socio-demographical and psychological characteristics of learners, learning behavior processes and learning outcome variables.

During the past ten years, several socio-demographic and psychological variables have been identified as significantly related to different patterns of learning behavior in online-learning and MOOCs. Those characteristics can be used, directly, indirectly, or in combination with learning behavior, to predict learner-centered outcomes. The next section will review research studies on five of the most important learner pre-course predispositions that have been investigated and identified as significant predictors of learning behavior of students taking MOOCs: (1) Socio-demographic variables (especially age and gender); (2) self-regulated learning skills; (3) motivation; (4) initial intentions; and (5) outcome beliefs. The research studies outlined in the following section form a significant part of the theoretical foundation of this dissertation.

## MOOCs and Learner Pre-Course Predisposition

### *MOOCs and Socio-Demographic variables*

MOOC learners are a more heterogeneous group compared to learners in formal education, comprising male and female participants of all ages worldwide, with different educational, socio-economic, and psychological characteristics. Demographic statistics show that there is a two-to-one male-to-female ratio (67 percent male, 33 percent female) of MOOC students. The median age is 29 years old, and a significant portion of learners are from outside the U.S. (71 percent international, 29 percent from the U.S.) (Chuang & Ho, 2016). Despite the diverse academic backgrounds of the participants, the vast majority of MOOC learners have at least college degrees (Despujol, Turro, Busqueis, & Canero, 2015). Those who complete MOOCs are more likely to have already completed a bachelor's degree or higher (Ho et al., 2014; Semenova & Rudakova, 2016). As mentioned above, this trend represents one of the main grounds for criticism of the potential of MOOCs, as most students who earn certificates for completing MOOCs are experienced learners with a strong academic background (Christensen set al., 2013; Daily, 2014; Emanuel, 2013; Guo & Reinecke, 2014; Koller et al., 2013).

Some earlier studies did not identify gender as an influence on instructor-centered outcomes, such as achievements or completion rates, in MOOCs (Breslow, Pritchard, & DeBoer, 2013; Cisel, 2014; Kizilcec, Piech, & Schneider, 2013; Morris, Hotchkiss, & Swinnerton, 2015). Other studies, found that women are more likely than men to complete a MOOC or obtain certification (Garrido, Koepke, Anderson, & Mena, 2016). In contrast, Semenova and Rudakova (2016), found that, in general, 6%–7% more men than women complete the course.

Furthermore, there are inconsistent findings regarding the association between age and academic achievement. Guo and Reinecke (2014), for example, found a positive correlation between age and grades, while Breslow et al. (2013) did not find a similar correlation. In their examination of completion rates, Morris et al. (2015) found that those who completed courses were, on average, older, while those who dropped out in the first week of the course were on average the youngest group.

## MOOCs and Self-Regulated Learning

One of the central characteristics of MOOCs is that participants are required to make educational choices concerning courses, learning path, and learning schedule (Kizilcec, Perez-Sanagustín, & Maldonado, 2017; Margaryan et al., 2015; Van den Beemt, Buijs, & Van der Aalst, 2018). In a learning environment where participants can choose where, when, and how to study, they must engage in self-regulated learning (SRL) to cope effectively with this autonomy.

Self-regulation is a context-specific process. In the context of learning, SRL is defined as a student's proactive actions aimed at acquiring and applying information, or skills that involve setting goals, self-monitoring, time management and regulating one's efforts towards learning goal fulfillment (Järvelä, Malmberg, & Koivuniemi, 2016; Reimann, Markauskaite, & Bannert, 2014; Tabuenca, Kalz, Drachsler, & Specht, 2015; Zimmerman, 1990). SRL involves different learner dispositions, including metacognition (orientation, goal specification, planning, searching for information, judgment of relevance, evaluation, and monitoring and regulation), cognition (reading, repeating, processing, and elaboration and organization), motivation, and other task-irrelevant aspects (Boekaerts, 1997; Reimann et al., 2014). Through monitoring and control, SRL can significantly influence learner behavior within the MOOC environment, and it plays a major role in determining learning outcomes. Positive relationships between self-regulatory learning skills and academic performance were found in online-learning environments (e.g., Barnard-Brak, Paton, & Lan, 2010; Butler & Winne, 1995; Zimmerman & Martinez-Pons, 1990).

Most online courses today are designed to be self-paced, based on a learner-centric approach which treats the learner as an active agent. This approach provides the freedom to select and control the services and tools that learners use. While this approach allows better opportunities for learners with high SRL, it has negative effects for learners with low SRL, due to the lack of guidance, defined structure, and support (Nussbaumer & Hillemann, 2015). MOOC participants with high levels of goal-setting and strategic planning skills and low levels of help-seeking were able to effectively attain their personal goals (Kizilcec et al., 2017). Several studies found a positive correlation between SRL and satisfaction in online courses (Artino, 2007; Li, 2019; Puzziferro, 2008). Learners who have difficulty regulating their learning process may experience increased dissatisfaction (Sun & Rueda, 2012) and are likely to drop out (Hew & Cheung, 2014; Kizilcec & Halawa, 2015). In summary, as the studies outlined above show, self-regulation is vital for mediating between personal characteristics, contextual features, and actual performance in the learning process.

## MOOCs and Motivation

Studies such as the one by Littlejohn et al. (2016) and Margaryan et al. (2015) found that learners who reported higher levels of SRL also reported higher levels of motivation and commitment to learning. Motivation is widely understood as the process that initiates, guides, and maintains goal-oriented behavior (Ryan & Deci, 2000). This process can be understood as a continuum that ranges between intrinsic and extrinsic motivation. Intrinsic motivation defined as active engagement with tasks because of self-desire to seek out new things and new challenges and to gain knowledge and fun. From the other side of the continuance, extrinsic motivation defined as the regulation of the activity as a function of expectations regarding reward and punishment (Pintrich, Smith, Garcia, & Mckeachie, 1993; Ryan & Deci, 2000).

In a learning context, intrinsic motivation is usually understood as a desire to learn for the sake of understanding (Byrne & Flood, 2005) while an extrinsically motivated learner wants to achieve a goal for the sake of an external reward (Paulsen & Gentry, 1995). Learners typically possess a mix of motivations (Garcia & Pintrich, 1994), however, intrinsic motivation is considered more desirable as it is generally thought to be stronger and more likely to move learners towards success (Dev, 1997; Donald, 1999).

In the context of MOOCs, student objectives range from intrinsic motivation such as curiosity, or the enjoyment of learning and accomplishment, to extrinsic motivation, such as obtaining certification or achieving specific professional purposes, a pay raise, or finding a new job (Castaño-Muñoz et al., 2017).

## MOOCs and initial intentions

According to the theory of planned behavior, intentions are the most important predictor of behavior (Ajzen, 1991). MOOC participants define, either consciously or unconsciously, their intentions towards participating in a course. These intentions have been found to be an important predictor of learning behavior (Littlejohn et al., 2016; Onah et al., 2014; Wang & Baker, 2018), and learning outcomes (Bonafini, Chae, Park, & Jablokow, 2017; Koller, et al., 2013; Konstan, Walker, Brooks, Brown, & Ekstrand, 2015). For example, Koller, et al., (2013) concluded that learners who intended to complete a MOOC achieved higher completion rates compared to those who did not. Wang and Baker (2018) showed that learner intentions to earn a certificate were positively associated with actually earning a certificate in a MOOC, and Bonafini, et al., (2017) found that a students´ desire for certification had an amplifying effect on students' MOOC completion, as well as on the number of videos watched by the students. However, in many cases, the intention is formed, but cannot be realized due to certain barriers which impede performance (Henderikx, Kreijns, & Kalz, 2018). This issue will be discussed in greater detail in the section titled "MOOCs and Learner-Centered Outcomes".

## MOOCs and Outcome Beliefs

Outcome beliefs are generally perceived to play a key role in attitude formation and the creating of intentions. Course outcome beliefs are variables that describe the expectations learners have regarding the outcomes of participating in the course. For example, people may have positive attitudes towards MOOCs if they believe that they are an important factor in providing more employment opportunities (Kalz et al., 2015). On the other hand, registrants may believe that participation in a MOOC will lead to negative outcomes, such as losing leisure time or creating stress, which in turn could lead to negative attitudes towards participation in the MOOC.



In traditional HE courses; learner expectations are usually defined by the course instructor and are largely standardized. However, the diversity of learners participating in a MOOC usually results in a range of motivations for participation (Kizilcec et al., 2013). These widely differing motivations lead to different expectations about the outcomes of participating in the MOOC, which, potentially leads to very different levels of engagement (DeBoer, Ho, Stump, & Breslow, 2014).

To summarize, the previous section presented findings from research studies that examined the effect of participants' predispositions to MOOC learning environments on behavior and learning outcomes. As these studies showed, these predispositions can predict participant online learning behaviors, as well as the learning outcomes of these behaviors. The next section will introduce the behavioral variables examined in this dissertation.

## MOOCs and Learning Behavior

Reich (2015) advocates that we should move from questionnaire-based measures of MOOC learning to a measure of actual learner behavior. Progress in the field of educational research in digital platforms has enabled us to measure, collect, and analyze data unobtrusively. Consequently, we can now "open the black box" as mentioned above, and utilize information, not only on learning outcomes, but also on the learning behavior and learning processes that led to these outcomes.

In MOOC learning environments, learning behavior is manifested mostly through access to different types of learning resources and materials, and through the usage patterns of these resources. This enables us to see, not only learning outcomes, but also proxies of learning behavior and processes. Based on the definition of learning as a process of individual knowledge construction that emerges in a dynamic process of interactions among learners, resources, and instructors (Bransford, Brown, & Cocking, 2000; Siemens, 2004), this dissertation will focus mainly on the interactions between the learners and the course content created by the course instructors. As Kizilcec et al. (2013) have observed, in MOOC-based learning environments, these interactions between learners and course content are shaped by the design of instruction, content, assessment, and platform features, which will be discussed in the next section.

## Patterns of learning behavior

MOOC learning behavior can range from passive observation to active participation and interaction with other users. Learner behavior can vary from selective use of a few resources to the use of all course materials and completion of all learning tasks and assignments. Several researchers have attempted to classify learners according to their behavior. For example, Kizilcec et al. (2013) showed that MOOC learners can be clustered into four distinct groups: *Completing* (finishing all course materials), *Auditing* (watching videos but not taking quizzes), *Disengaging* (decreasing in engagement over time), and *Sampling* (watching only a few videos). Kahan, Soffer, & Nachmias (2017) adopted a more granular approach and clustered MOOC learners into seven clusters of participant behavior: *Tasters*, *Downloaders*, *Disengagers*, *Offline Engagers*, *Online Engagers, Moderately Social Engagers*, and *Social Engagers*.

These differing patterns of interacting with the learning material can lead to very different learning outcomes. Davis, Chen, Hauff, & Houben (2016), for example, showed that learners who successfully passed a MOOC had shown more interest in their quiz scores than learners who did not pass successfully, and had used progress tracking tools more often. Such learning patterns show that learners who make better use of SRL strategies may successfully pass a course more often than learners who do not use such strategies.

Additional learning behaviors include participants choosing to learn in diverse learning sequences and paths (Van den Beemt et al., 2018). MOOCs learning paths can deviate widely from a linear course: Learners can start the courses later than the original launch date and can view and repeat lectures and quizzes several times. Guo & Reinecke (2014), for example, observed that successful MOOC certificate earners viewed only 78% of the course content and skipped the rest. They were more engaged in non-linear navigation behavior than non-certificate earners and tended to “jump” backward to revisit earlier lectures or assessments up to three times more often than learners who ultimately did not earn a certificate.

A variety of behavioral measurements can be used for clustering the participants. These behavioral measurements, which can be extracted unobtrusively from the log-file of the course, include the number of video lectures that participants accessed during the course, the length of the video watched, the number of pauses and replays, and the speed at which the video was viewed.

Although video lectures are the main learning resource for knowledge acquisition in most MOOCs, additional learner behaviors for acquiring knowledge include interactions with learning resources such as text, pictures, and audio files, which also constitute types of learning behaviors that need to be measured. For example, quizzes are indicators of self-evaluating activities that enable MOOC participants to assess their knowledge. In order to measure self-evaluating actions, we can measure the number of quizzes that the participant accessed, the number of attempts to respond to and to pass the quizzes, and the participants’ performance.

Two additional forms of interaction that constitute types of learning behaviors that can be assessed among MOOC participants are social interaction activities and instructor-defined outcomes. Social interactions include the number of discussion forums that the participant accesses during the course and the number of questions, answers, and comments that the participant posts to course forums. Instructor-defined outcomes include participants earning a certificate of course completion and course completion badges. The criteria for receiving the certificate are based on completing a minimum quota of course activities.



Additional dimensions that can be assessed are related to time, which highlights the dynamic and temporal nature of the learning behavior. Examples of these measures include calculating the duration of access time to the MOOC, the amount of time spent on specific activities, the number of times learners log in per week, the total number of MOOC-related activities accessed by the participant, and the number of times the participants repeatedly access the same learning resources.

Temporal elements also contribute to our understanding of the larger picture of learning behavior. These include the time of day and portion of the week during which MOOC content and activities are accessed (Tabuenca et al., 2015), as well as the sequence of accessing learning material (e.g. Davis, et al., 2016; Jovanović et al., 2017; Maldonado-Mahauad et al., 2018; Reich, 2015; Sinha, Jermann, Li, & Dillenbourg, 2014) These factors will be discussed in the next section.

## MOOCs and Learning Sequences

As mentioned above, many research studies on MOOCs have focused on the static counting-based features of learning behavior, while ignoring the sequence (order) of the activities and their temporal nature. In MOOCs, participants can choose the learning path that they will take and their learning sequences. This aspect of choice is one of the many differences between MOOC learning environments and most traditional in-class courses. In most traditional learning settings, course structure is pre-defined and there is only one learning path.

As Li, Wang, & Wang (2017) demonstrated, if we take into consideration the number of activities in which the participant participated and ignore the order or sequences of those activities, we are missing essential information. For example, if we consider three imaginary participants who watched videos (V) and answered quiz questions (Q), the first participant may have watched three videos and then answered three quizzes (V-V-V-Q-Q-Q). In contrast, the second participant may have first tried to answer the quiz questions, and only then watched the video lectures (Q-Q-Q-V-V-V). The third participant may answer a quiz questionnaire every time a video is viewed (V-Q-V-Q-V-Q). Although all the imaginary participants watched three videos and answered three quizzes, as this example illustrates, their learning sequences are very different. However, the majority of studies that use learning analytics and educational data mining have initially concentrated on frequency analysis, while little attention has been paid to the order and sequence of student activity in online courses over time.

To obtain deeper insights into the significance of learning sequences, process mining (PM) has been introduced as a promising technique for the educational research community (Calders & Pechenizkiy, 2012; Van den Beemt et al., 2018). Process mining techniques enable a wide variety of processes to be analyzed using event data recorded from log-files or other kinds of resources (Aalst, Guo, & Gorissen, 2015; Bannert, Reimann, & Sonnenberg, 2014). To identify the sequential nature of the activities, one must extract event logs that include an identifier of the learner, the activity that has been done, and a timestamp (Van den Beemt et al., 2018). The potential promise of PM comes from the many possibilities that this technique offers. For example, PM can be applied to discover underlying processes and patterns, analyze bottlenecks, uncover hidden inefficiencies, check compliance, explain deviations, predict performances, and guide users toward “better” processes (Aalst et al., 2015). In addition, PM methodologies can also uncover underlying business processes, deviations, and in general, usage patterns in an unstructured process in noisy systems with no clear processes, or when processes can occur in many ways, such as in MOOCs (Codish et al., 2019).

PM is typically used to discover frequent sub-sequences of events or to examine relationships between the events (Reimann et al., 2014). Data collected through the learning process can then be correlated with data from survey respondents who completed surveys before and after participating in the course. These could include respondents’ socio-demographic characteristics, educational background, motivations, initial intention, SRL, intention-fulfillment, and post-course satisfaction levels.

This comprehensive approach, which investigates different learning processes and sequences among diverse groups of students, which are then correlated with student achievement, has been adopted by several researchers. For example, Aalst, Guo, & Gorissen (2015) examined differential learning processes among students who passed or failed an online course. These researchers also compared the learning process of male and female students, compared different parts of the course and compared local and international students.

Additional researchers who correlated learning processes and sequences with successful learning outcomes include Van Den Beemt, et al., (2018), who found that successful students exhibited a steadier learning behavior and that this behavior is highly correlated with watching successive videos in batches. Similarly, Kizilcec et al., (2013) showed that the order of media consumption predicts the level of engagement throughout the course. While some learners preferred to focus on trajectories of watching sequences of video lessons from the beginning, others undertook course evaluation and follow-up tasks. The sequence of participation predicted the level of attrition.

Additional studies focused on clustering learners according to their learning sequences and their level of engagement with MOOCs (Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017; Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Munoz-

Gama, 2018). In both studies, all paths taken by learners from the start of a learning session to the end of a session were first analyzed and clustered. This clustering resulted in various possible paths that learners could follow from the beginning of a session to the end. Next, learners were clustered, based on the frequency of these different paths in their learning. Both studies showed that learners engage differently with MOOCs during learning sessions.

In-depth examinations of the sequential nature of learning behaviors have recently been facilitated by techniques borrowed from the natural language processing (NLP) domain. These techniques can be used to identify the sequential nature of learning behaviors and their connection to learning outcomes in MOOCs. Several studies have used NLP features to study dropout and retention levels among MOOC participants by studying the language students use (Crossley, Paquette, Dascalu, McNamara, & Baker, 2016; Elgort, Lundqvist, McDonald, & Moskal, 2018; Kim, Singh Chaplot, & Rhim, 2015; Robinson, Yeomans, Reich, Hulleman, & Gehlbach, 2016). However, relatively few studies (e.g. Li et al., 2017) have applied NLP methods in order to find common sequences of activities and their transitional probabilities.

The studies mentioned above serve as useful examples of research that has investigated the effect of learning sequences on the completion rates of courses. However, until now, to the best of our knowledge, there have been no studies that connect the learning activity sequences of MOOC participants with learning outcomes, such as levels of intention-fulfillment and satisfaction.

Having described the connections between participant predisposition and learning behavior in MOOCs, the next section will describe the effect of participant predisposition and learning behavior on learning outcomes while distinguishing between instructor- and learner-focused outcomes.

## MOOCs and Learning Outcomes

### *MOOCs and Instructor-Focused Outcomes: Grades, Completion rates, and Certified Earning*

An extensive amount of research has focused on MOOC learning outcomes, such as grades, completion rates, and certificate earning (Ho et al., 2014; Kizilcec et al., 2020; Lim, Coetzee, Hartmann, Fox, & Hearst, 2014; Reich & Ruipérez-Valiente, 2019). The aforementioned learning outcomes are perceived as objective measurements, serving as a proxy to the level of student engagement. Learning outcomes such as these are borrowed from formal education contexts where students enroll in courses with the goal of completing them and of satisfying the learning outcomes defined by the instructor (Henderikx et al., 2017; Reich, 2014).

Many research studies have investigated the factors that predict instructor-focused outcomes (Chen et al., 2019; Hong, Wei, & Yang, 2019; Morris et al., 2015; Rosé et al., 2014; Xie, 2020; Zhang, Bonafini, Lockee, Jablokow, & Hu, 2019). It has been found that the participants' level of education (Guo & Reinecke, 2014), experience with MOOCs (Semenova & Rudakova, 2016), prior online learning experience (Morris et al., 2015), gender (Semenova & Rudakova, 2016), age (Breslow et al., 2013; Guo & Reinecke, 2014; Konstan et al., 2015; Morris et al., 2015), geographic location, employment status, and time constraints, (Cisel, 2014), initial intentions (Konstan et al., 2015; Onah et al., 2014), and level of motivation ( Pursel, Zhang, Jablokow, Choi, & Velegol, 2016) are predictors of instructor-focused outcomes.

In addition, instructor-focused outcomes are highly correlated with factors such as number of videos watched, number of posts in forums, attempting mastery quizzes, and SRL behavior (Bonafini et al., 2017; Brooks, Thompson, & Teasley, 2015; Joksimović, Gašević, Kovanović, Riecke, & Hatala, 2015; Kloft, Stiehler, Zheng, & Pinkwart, 2014; Maldonado-Mahauad, Pérez-Sanagustín, Moreno-Marcos, et al., 2018; Moreno-Marcos et al., 2020; Pursel at el., 2016; Rosé et al., 2014).

However, as mentioned above, one key criticism of MOOCs is the high dropout and low retention rates of participants (Gardner & Brooks, 2018; Reich & Ruipérez-Valiente, 2019). The dropout rates for MOOCs are, on average, 93% (Chuang & Ho, 2016; Jordan, 2014; Margaryan et al., 2015). Though MOOC dropout rates are indeed very high, a key question is whether completion rate is the appropriate measure for evaluating the success of this new form of lifelong learning. As stated at the beginning of this section, completion rates, grades, and certificate earnings are success criteria borrowed from formal education contexts in which students enroll in courses with the goal of completing them, and of satisfying the learning outcomes as defined by the instructor (Henderikx et al., 2017; Reich, 2014). The success of lifelong learning in MOOCs should be evaluated, not through traditional instructor-focused measures that characterize formal education contexts, such as dropout rates and earning of completion certificates, but rather, through learner-centered measures that take into account the non-formal nature of MOOC learning. Participants may enroll in MOOCs for a variety of reasons (Littlejohn et al., 2016; Onah et al., 2014; Wang & Baker, 2018), and MOOC participants may have a variety of expected learning outcomes. One of the expected learning outcomes could be absorption of all the course materials and completion of all the course tasks. However, MOOC participants may achieve their individual learning goals by engaging in only a segment of the course (Ho et al., 2015; Liyanagunawardena et al., 2013). Thus, the perception of MOOC success depends on the individual objectives of the learner, and not necessarily on instructor-defined objectives.

In this study, we define MOOC success as the achievement of a pre-defined set of personal objectives, which may or may not be the same as "completion" in the sense of completing all learning activities, tests, and finally receiving a certificate. Therefore, we focus on learner-centered outcomes, i.e. learner intention-fulfillment and satisfaction.

## MOOCs and Learner-Centered Outcomes

### *MOOCs and Intention-Fulfillment*

Intention-fulfillment (IF) is emerging as a promising measure of success for open learning and MOOCs. IF takes into account the personal achievement objectives of learners, rather than external success criteria (Henderikx et al., 2017). Participants, consciously or unconsciously, formulate their intention when accessing the course. The initial intentions affect the learning behavior, learning outcomes, and perception of success. According to the concept of the intention-behavior gap, people do not always do the things that they intend to do (Sheeran & Webb, 2016). previous research, however, identified learner intention for completing a MOOC as a significant estimator of their actual completion (Bonafini et al., 2017; Koller et al., 2013; Konstan et al., 2015). For example, Bonafini, et al., (2017) found that student desire for certification had an amplifying effect on MOOC completion, as well as on the number of videos watched by the students. Thus, participant intention and it`s fulfillment should be taken as a central measure when evaluating the success of a participant in a course.

### *MOOCs and Learner Satisfaction*

Learner satisfaction reflects student perception of the learning experience (Alqurashi, 2019; Kuo, Walker, Schroder, & Belland, 2014; Littlejohn et al., 2016) and is defined as a student's overall positive assessment of his or her learning experience (Keller, 1983). Online learning satisfaction has been emphasized as the most important element defining student online learning success (Horvat, Dobrota, Krsmanovic, & Cudanov, 2015; Naveh, Tubin, & Pliskin, 2012). In addition, online learner satisfaction was included as one of the five elements for the evaluation of the quality of online learning identified by the Online Learning Consortium (Alqurashi, 2019).

Some authors have found positive correlations between student satisfaction and post-secondary student success (Chang & Smith, 2008), intention to use e-learning (Liaw & Huang, 2011; Roca, Chiu, & Martínez, 2006), dropout rates of students and motivation and commitment to complete a degree online (Ali & Ahmad, 2011; Yukselturk & Yildirim, 2008), and the intention to continue using MOOCs (Alraimi, Zo, & Ciganek, 2015; Pozón-López, Higueras-Castillo, Muñoz-Leiva, & Liébana-Cabanillas, 2020).

In contrast, a study of student data from the Open University of the UK by Rienties and Toetenel (2016) found that retention and satisfaction in formal distance education context are not correlated. The authors explain these findings with the fact that learning is not always fun and requires effort. The effort to study via distance learning, combined with the unstructured, self-paced nature of the MOOC environment creates unique types of barriers to the learning process, which can affect learning outcomes, including levels of satisfaction among the participants.

### *MOOCs and Barriers to Satisfaction*

In the informal learning environment of MOOCs, many participants face barriers that prevent them from being satisfied with their course and their learning process. Barriers to learner satisfaction are defined as issues that hinder or prevent learners from reaching their individual intentions and that harm their level of satisfaction (Henderikx et al., 2018). The barriers may be related to the MOOC itself or might be extraneous to the MOOC. Some examples of barriers related to the course itself include bad course content, low quality of course materials, or the absence of the instructor. Alternatively, examples of barriers that are extraneous to the MOOC or the MOOC environment are lack of time, insufficient academic background, family issues, workplace commitments and insufficient technological background (Henderikx, et al., 2018; Khalil & Ebner, 2014; Onah, Sinclair, & Boyatt, 2014).

There are a variety of antecedents to the barriers that participants face while learning online. Characteristics such as age (Henderikx, Kreijns, Castaño Muñoz, & Kalz, 2019), gender (Henderikx et al., 2019; Muilenburg & Berge, 2007), SRL (Kizilcec, Perez-Sanagustín, & Maldonado, 2017; Zalli, Nordin, & Hashim, 2019) and level of self-efficacy (Alqurashi, 2016) were a predictor of barriers to online learning.

After discussing the theoretical background of the dissertation in-depth, the next sections will present the research questions and the structure of the thesis.

## Research Questions

The central research question that this dissertation aims to answer is: How do we evaluate the learner-centered outcomes and their antecedents in open online education? To answer the central research question, the following research questions were investigated:

1. What are the reciprocal relations between traditional HEI and open education, and what are the implications for success measurements in open distance education?

2. How can we define learner-centered outcomes in MOOCs and what are the predictors of those outcomes?

3. What are the barriers to satisfaction that learners face while taking MOOCs and what predicts those barriers?

4. How can we detect user behavior patterns and cluster them based on the learning trajectory of MOOC participants?

5. Can the learning activity sequences of the participants serve as a predictor of the level of intention-fulfillment?

## Overview of this Dissertation

This dissertation consists of seven chapters, comprising five studies, an introduction, and a summary. Together, they aim to answer the central research question: How do we evaluate learner-centered outcomes and their antecedents in open online education?

Figure 1 provides an overview of the studies presented in this dissertation in terms of the research stages, the variables collected (the participants' characteristics and behaviors). The studies are summarized in the following paragraph.

The introduction presented a general overview of open and distance education, the MOOCs phenomena, and the critiques surrounding MOOCs. In response to the criticism, the chapter introduced the concept of learner-centered outcomes, their antecedes, and learning analytics as a methodology to research those phenomena.

Chapter 2 presents a comparative analysis between the business models of traditional HEI and open education. We discuss the impact of digital innovation on the business models of higher education institutions using Raymond's (1999) well-known "Cathedral and Bazaar" metaphor on software engineering methods. The changes promoted by the "bazaar" facilitate the adoption of MOOCs by the mainstream "cathedral", but require, at the same time, the development of new learner-centered outcomes measures, which will be appropriate for the emerging educational ecosystem. This chapter contributes to the evolving literature on the strategic impact of open online education on the HEI landscape.

Chapter 3 introduces two learner-centered MOOC outcomes for non-formal lifelong learning frameworks such as MOOCs: learner satisfaction and learner intention-fulfillment. The study empirically defines them and reveals their predictors in a MOOC. The research results clarify the complex nature of the relationship between learner socio-demographic characteristics, learner behavior, and learner-centered outcomes.

The effects of socio-demographic characteristics on the barriers to satisfaction among MOOCs' participants are discussed in Chapter 4. Identifying these barriers to satisfaction and predicting them provides additional insight into the nature of satisfaction as a learning outcome. The presented research uses a survey which combines pre- and post-questionnaires to gather this data.

Previous studies have shown that clustering participants based on their learning trajectories is more informative and has a higher potential for pedagogical improvement, compared to clustering participants based on static-counting behavioral data (Kizilcec et al., 2013). Chapter 5 seeks to explore a novel approach to detect user behavior patterns by spotting very short user activities and clustering them based on shared variance. This will allow us to construct meaningful behavior patterns in unstructured processes such as MOOCs and other forms of online learning.

Chapter 6 identifies the effect of the learning activity sequences of the participants as a predictor of the level of participant intention-fulfillment. In the study, a novel approach borrowed from the natural language process (NLP) domain had been used to identify different learning activity sequences. The association between the identified learning sequences and the level of IF has been significant and meaningful.

Last, but not least, the general discussion provides an overview of the findings in each chapter and gathers insights from the five studies that have been presented. The chapter concludes with a general discussion and conclusions. Implications, limitations, and future research suggestions are offered.

In conclusion, as outlined in Figure 1, this dissertation studies the nature and antecedents of learner-centered outcomes in massive open online courses, using a combination of research methods, such as self-report surveys, educational data mining, learning analytics, unstructured process-mining, and NLP. The triangulation of methods bridges the disadvantages of each individual method. Self-report surveys have been used to identify both the psychological and educational state of each participant prior to starting a course, and the learning outcomes at the end of the learning period. In parallel, data has been mined from log-files of the behavioral learning activities of the participants. Each learning activity has been identified according to type of activity, identity of participant, and time of action. This enables us to follow the learning trajectories of the participants and to analyze the aggregate counting data. By combining all these research methods, we are able to obtain more accurate and rich insights from the data sets on the learner-centered outcomes and their antecedents.

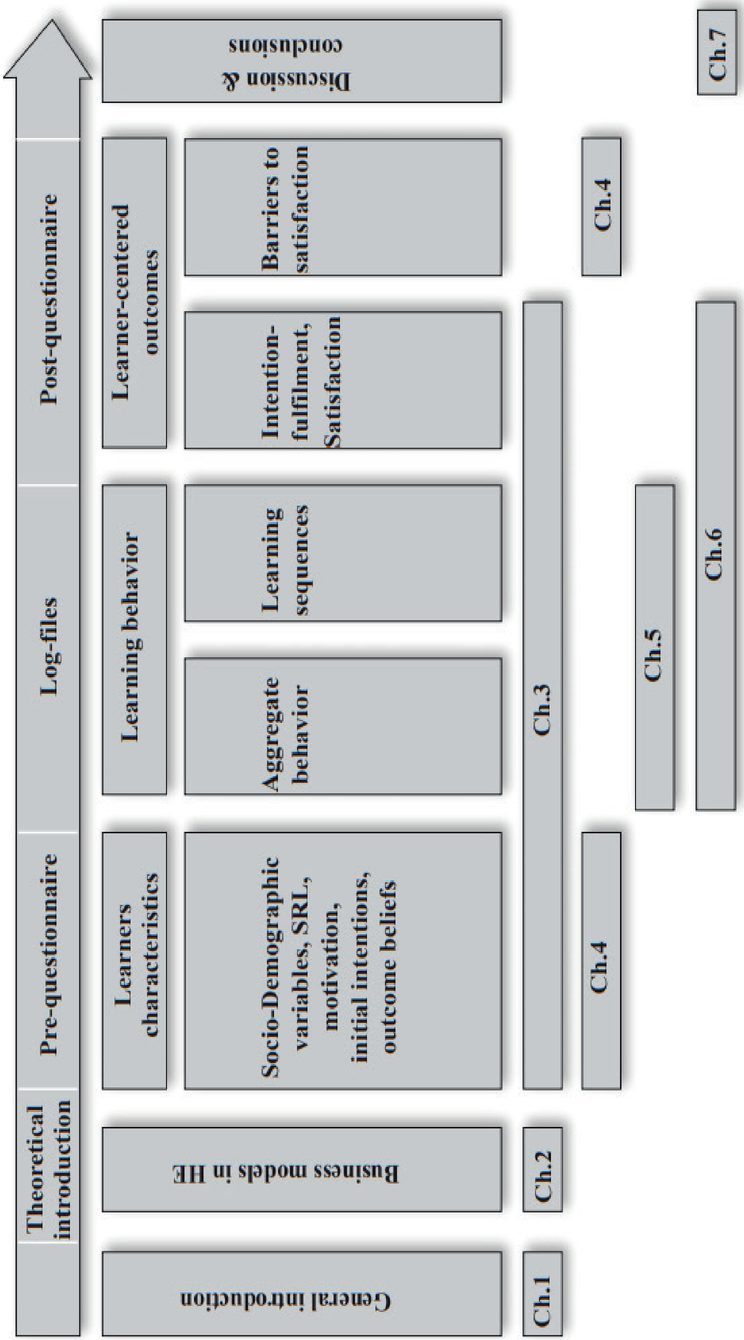



**Figure 1.** An overview of the studies presented in this dissertation in terms of the research stages, the variables collected

2

# Chapter 2

## The Cathedral's Ivory Tower and the Open Education Bazaar – Catalyzing Innovation in the Higher Education Sector

This chapter is based on:

Rabin, E., Kalman, M, Y., & Kalz, M. (2019).
The Cathedral's Ivory Tower and the Open Education Bazaar –
The Impact of Digital Innovation on Higher Education's Business Models.

*Open Learning: The Journal of Open, Distance and e-Learning.*
*https://doi.org/10.1080/02680513.2019.1662285*

## Abstract

Will open education replace traditional higher education, or augment it? Digital innovation in the higher education sector is fuelling speculation about the transformation of higher education and the future role of universities. Much of the speculation makes questionable implicit assumptions about current and future business models in the higher education sector. This conceptual paper applies an innovation management perspective to critically examine the use and misuse of the business model concept in the context of digital innovation in the higher education sector. Using Raymond's metaphor of the cathedral and the bazaar which contrasted traditional commercial software development (the cathedral) with open source software development (the bazaar). We analogise this relationship with the relationship between 'cathedral-type' business models in traditional higher education (e.g. universities) and 'bazaar-type' business models in open education (e.g. open educational resource publishers). Using the historical perspective we now have on the software industry's evolution we critique the ubiquitous replacement narrative of destruction and disruption of the sector, and propose an alternative narrative of interdependence and mutual innovative catalysis. We predict that higher education ecosystems will be based on synergistic relationships between organisations that represent many gradations on the continuum between 'cathedral-type' and 'bazaar-type' organisations.


***Keywords:*** Digital innovation; Business models; Open education; Disruptive innovation; Open innovation

## Introduction

The global higher education sector in general, and higher education institutions (HEIs) in particular, are facing a transformation process triggered by digital innovation (Orr, Weller, & Farrow, 2018). The goal of this paper is to present to decision-makers, policy-makers and researchers in the higher education sector tools to better understand the impact of digital innovation on higher education business models. In order to do that we briefly review digital innovation using the framework proposed by Nambisan, Lyytinen, Majchrzak, and Song (2017). In light of Nambisan et al.'s emphasis on metaphor and narrative as a vehicle for shared sense-making in digital innovation, we present several popular narratives which we characterise as *replacement* – narratives that suggest that digital innovation will lead to the *replacement* of traditional HEIs with novel organisations using open educational resources (OER). This *replacement* narrative is contrasted with an alternative narrative of *interdependence* and mutual innovative catalysis, which suggests that traditional HEIs will not be replaced by organisations fuelled by digital innovation. Instead, we suggest that both traditional HEIs and these novel organisations will form a mutually dependent ecosystem where digital innovation is a major engine of change. We justify our preference for the *interdependence* narrative over the *replacement* narrative by using the 'cathedral and bazaar' metaphor (Raymond, 1999). This metaphor, which originated more than two decades ago in the software industry at the stage that open source software was emerging, allows us to use the history of open source in the software sector as a case study that teaches us about the possible impact of open educational resources on the future history of the higher education sector. The main contribution of this discussion is to present researchers and decision-makers interested in innovation in the higher education sector with a metaphor and a point of view that will help change the narrative. Such a change is conducive to moving beyond outdated concepts of innovation, and to promoting the higher education sector to innovate in a purposeful manner that will allow it to remain relevant and meet its key societal challenges in the coming decades.

In 'What are Universities For?' Collini (2012) describes a paradox in the tension between the growing importance of universities on the one hand and the lack of confidence in the university as an institution on the other hand. Universities have never before seen such a massive growth in numbers of institutions, of students and even of funding, yet they suffer from a lack of confidence and loss of identity. HEIs are expected to cope with the growing global demand for higher education (Economist Intelligence Unit, 2015) and to find answers to scalability of their modes of teaching through innovative digital educational technologies. 'Open education' (Blessinger & Bliss, 2016) is an umbrella term for innovations that could provide answers to the scalability challenges facing HEIs, for example by making digital learning resources openly available for use, reuse and adaptation (Bates, 2015), by enabling the use of open intellectual property licences, and by enabling open curricula, open learning, open assessment and open platforms (Yuan & Powell, 2013).

## Higher Education Innovation: A Replacement Narrative

A common narrative of open education, which was intensified after the great breakout of the massive open online courses (MOOCs) phenomenon, is that the current business model of universities will disappear (Weller, 2015). This *replacement* narrative is supported by concepts such as Schumpeter's *creative destruction* and Christensen's *disruptive innovation* (Bergek, Berggren, Magnusson, & Hobday, 2013).

The use of digital resources for teaching and learning in general, and specifically MOOCs, is often referred to as *disruptive innovations*. The concept of disruptive innovation has been developed and popularised by Clayton Christensen (Christensen, Horn, & Johnson, 2011). The core of the idea is that a product or service takes over a market by disrupting the business model of this market's incumbents. Disruptive innovations often begin as low-quality and low-cost alternatives, and eventually take over a market so swiftly that even powerful incumbents are unable to adjust their business model sufficiently. Eventually, the incumbents end up disappearing from the market. Digital technologies and online education have often been presented as a *disruptive innovation* that could disrupt universities as we know them today (Christensen, Horn, Caldera, & Soares, 2011; Noam, 1996).

Another characteristic of some forms of digital innovation in the higher education sector that supports the replacement narrative is *zero marginal costs* (Rifkin, 2014). This concept describes a condition when costs for products and services reach a ceiling effect, in that producing one more unit, or adding one additional user to a service, has such a low impact on the costs of the service that the added cost is negligible. This phenomenon is common in digital environments, where the costs of communication, storage and processing drop at exponential rates (Benkler, 2006). Rifkin (2014) predicts a future in which zero marginal costs digital educational products dominate the educational landscape and in which organisations that produce these products replace educational institutions as we know them today.

Last but not least, the concept of *unbundling* is repeatedly mentioned in discussions about a future of higher education in which HEIs as we know them today are replaced by radically different institutions. The unbundling of HEIs into separate entities (and/or personnel) who perform the three roles currently assigned to core academic faculty, namely research, teaching and service, is not a new concept (Macfarlane, 2011). The rise of MOOCs reinvigorated the prediction (Craig, 2015) that the business model of universities will become unbundled and result in MOOC-based providers of academic teaching, and other organisations that provide services such as testing and accreditation, academic research, etc. Woolf University is one recent example of this push, applying blockchain, the distributed ledger technology, to reconfigure and reconnect an unbundled educational institution (Fredin, 2018).

These three concepts (disruptive innovation, zero marginal costs and unbundling) have a common theme. This theme – the replacement narrative – predicts the demise of the university business model as we know it today and its replacement by open and flexible business models made possible through digital innovation. The appearance of MOOCs caught the imagination of many decision-makers and policy-makers in the higher education sector, and the consequent hype strengthened the sense that we are about to witness a transformation in the sector. We question these claims about the future of higher education business models and use the history of the impact of open software on the software industry to propose an alternative narrative. This critique is important since the ubiquitous replacement narrative interferes with the way actors in the higher education sector make sense of innovation in higher education (Nambisan et al., 2017). For example, if academic faculty, decision-makers in HEIs, or higher education policy-makers adopt the replacement narrative predicting a demise of the university business model as we know it today, this can lead to confrontational attitudes within these groups and to defensive reactions that will hurt the whole sector and stifle innovation.

Instead, we present an alternative narrative that acknowledges the distributed nature of innovation agency and the fluid boundaries of the innovation space. This narrative is based on the cathedral and bazaar metaphor that was presented two decades ago. The original intention of the cathedral and bazaar metaphor was to suggest the demise of the incumbent business model in the software sector, and in this sense, it is reminiscent of the current situation in the higher education sector. It is now widely understood how open practices can promote innovation for both the incumbents and the (former) newcomers in the software sector, and we wish to apply these insights to the higher education sector.



## Business Models in the Higher Education Sector

In this section we present the concept of business models and demonstrate the use of business models in the context of the two prototypes: the cathedral and the bazaar.

### *Business Models*

A business model is a tool used by researchers and by practitioners to describe and analyse the logic of organisations (Osterwalder & Pigneur, 2010). Although it was developed in the context of for-profit organisations, business modelling is useful not only for businesses but also for any type of organisation, including non-profits in the higher education sector (De Langen, 2013; Kalman, 2014). In this paper we use Kalman's (2014) simple business model, as it was applied to HEIs. This simple business model describes who uses the educational product/s of the HEI and why they use it (the customer value proposition – CVP), what processes and resources make this value proposition possible, and the financial consequences of this activity for the HEI. Table 1 summarises the components of the business model and briefly describes them in the two left-hand columns, and in the two right-hand columns it demonstrates the use of the model at two prototypes of HEIs, the cathedral and the bazaar. The metaphor of the cathedral and the bazaar will be presented in detail in the next section.

**Table 1.** Components of the business model (Adapted from Kalman, 2014)

| **Business model component** | **Description** | **Example taken from cathedral type HEIs** | **Example taken from bazaar type HEIs** |
|---|---|---|---|
| Customer value proposition | The characteristics and needs of the organization`s customers, and the way these needs are met | The students get structured, pre-defined curricula, and study in clearly defined degree awarding programs. | The learners have extensive freedom in choosing their learning materials, based on their preferences goals, and needs. |
| Infrastructure | The resources and processes of the organization | Physical *resources* such as lecture halls, laboratories, and student dormitories. *Processes* such as advising and financial support | Digital *resources*, that allow studying anytime and anywhere. Digital *processes* such as automated grading and peer feedback |
| Financial | The financial principles according to which the organization operates | Income is derived mainly from tuition and government support | Income is derived mainly from learner payments for specific services such as certification, delivery of physical books, and other freemium services |

The most important component of HEI business models is the 'customer value proposition' (CVP). It answers the question 'What are the characteristics of the students?' and 'What is the value that the institution provides to them?' The obvious value is the knowledge the institution provides to the students, but there are many other values (benefits) that students receive from their HEIs, such as access to a social network of like-minded peers, social capital, involvement in diverse cultural and social activities, and the certificate and academic degree that confirm a student's successful completion of all academic requirements. The second component of the business model is the HEI's infrastructure, which includes the resources of the institution, and its processes. Each traditional HEI has dozens of resource categories including real estate (e.g. labs, class-rooms, dormitories, and sports facilities), IT resources, human resources, financial resources, and more. Each HEI also has a large number of processes including teaching, research, and administrative processes (Orr et al., 2018). The third component is financial, and it describes the financial principles according to which the institution operates: the cost structure (i.e. how costs are allocated to various processes and units at the institution), the nature of these costs (fixed or variable), the sources of income from students and from other stakeholders such as government and philanthropists, etc.

No two HEIs are identical in their CVP, infrastructure, or financial profile, and thus each HEI has a different business model. Nevertheless, many institutions can be grouped under a particular kind of business model that characterises the institutions as a group. For example, the business models of top-tier research universities around the world are more similar to each other than to the business models of teaching-focused institutions in their countries, such as (US based) community colleges, or to the business models of distance teaching universities. Business models of non-HEI's in the higher education sector are, again, very different. These organisations include academic book publishers, providers of online learning materials such as MOOCs, providers of software for admin- istrative purposes (e.g. enrolment, financial aid) and for academic processes (e.g.

plagi- arism detection, learning management systems), providers of financial services to students, tutoring services, and more.

Significantly changing a business model is difficult and risky. Attempting such a change could lead to a disruption of the business model (Johnson, Christensen, & Kagermann, 2008) and eventual failure of the organisation. In an organisation that has a good business model, all of the components interlock and complement each other. This interdependency pro- vides robustness and stability, but could also become a hindrance to change (Christensen & Raynor, 2003).

***Cathedrals and Bazaars***

Raymond (1999) famously coined the metaphor of the cathedral and the bazaar to compare commercial and open source software development. At the time Raymond published his ideas, commercial software development was based on centralised design and meticulous execution, while open source software development was based on a noisy emergent process rife with redundancy and imperfection. The main purpose of Raymond's work was to describe how open source software development was different from commercial software development, and to explain the ways in which the apparently inferior process of open software development can lead to better results than the highly structured and closely managed processes of commercial software development. The cathedral metaphor was used to describe the centrally managed software-development process in which a clearly defined development team provides the end-user with a closed software product that can be used right out of the box. The bazaar metaphor was used to describe the more loosely coordinated software development process, carried out by a distributed group that collaborates on an ad-hoc basis, releasing incrementally modified versions of the software, while keeping the product free and the source code open to the community. The end product is constantly evolving, and often requires more technological sophistication from the end user than parallel commercial products.

The cathedral and bazaar metaphor is useful for discussions of business models in the software industry and in other sectors (e.g. Baraniuk, 2008; Bezroukov, 1999a; Fitzgerald, 2006), although the applicability and generalisability of Raymond's claims are controversial due to their oversimplification, utopian nature and inefficiency (Bezroukov, 1999a, 1999b). Furthermore, Raymond's sense that the bazaar will replace the cathedral was proven to be wrong. In fact, the contemporary software sector comprises of a host of companies that combine 'cathedral' and 'bazaar' characteristics. A good example of the complex relationship between the cathedral and the bazaar in the software industry is the 2018 $34 billion acquisition of Red Hat, a company heavily reliant on open source software, by IBM, one of the archetypal 'cathedrals' in the software sector (Lohr, 2018). The main two motivations for this acquisition were investing in open source, and positioning IBM as a cloud computing powerhouse (Vaughan-Nichols, 2018). Thus, without adopting or endorsing Raymond's claims, we adopt his metaphor, and use it and the recent history of the software industry since 1999 in order to gain insights into the higher education sector.

The cathedral is an apt metaphor for the university and for traditional higher education institutions, where most learning takes place in carefully predefined and relatively rigorous tracks (courses, degrees), where most teaching is carried out by paid staff, and where there is a well-developed infrastructure that provides the resources and processes required to support all aspects of a fully developed organisation. In contrast with the cathedral, the bazaar is a metaphor for the open higher education sector, where most teaching is technology-based, where learning is often self-driven by motivated learners, where most exchanges are not based on monetary compensation, and where the infrastructure (resources and processes) is usually fragmentary and distributed. The bazaar approach allows for a greater degree of personal autonomy as a result of more horizontal structures of power and influence due to its decentralised nature (Farrow, 2017). The bazaar has long been evolving in the shadow of its cathedral's proverbial ivory tower, and has recently emerged into the sunlight, at which point its ability to leverage the exponentially declining costs of digital computation, communication and storage, enabled it to offer educational services and products for free or at a significantly reduced cost. The cathedral and the bazaar can thus be used as metaphors for the business models of traditional higher education organisations, and of open higher education initiatives, respectively.

### *Cathedral-Type and Bazaar-Type Business Models in the Higher Education Sector*

How can the cathedral and bazaar metaphors and the concept of business model assist us in the analysis of business models in higher education? We begin by describing Cathedral-type (C-type) business models in the higher education sector. Some well-known archetypical C-type business models in higher education are the research university, major textbook publishers (e.g. Pearson), academic publishers (e.g. Elsevier) and major educational technology providers (e.g. Blackboard Inc.). What are the characteristics of these C-type business models?

The customer value proposition (CVP) of C-type organisations in higher education is to provide the customer with a highly structured path to achieve their goals. In return for the fees paid per customer, the customers (students, faculty members, HEIs) receive services and resources that are outside their area of expertise (or, in the case of organisational customers, which are not a part of their core-competencies), and in return can focus on achieving their overall goals. For example, students pay annual university tuition and receive a clearly defined path to achieving their undergraduate degree, including not only their courses but also a process for selecting the courses, access to advising and other support services, a clearly defined academic calendar, and a host of other services and resources that allow them to focus on achieving their goals as students1. The customers of the textbook publisher are teaching faculty who assign the textbooks to their courses. They receive from the publishing house not only a comprehensive text that provides their students with the knowledge required in the course, but also additional products and services such as online resources for the faculty members (e.g. slides, a teaching guide) and students (e.g. supplementary audio-visual resources, practice questions and quizzes), as well as ongoing updates about future editions of the textbook (Hammond, Danko, & Braswell, 2015). The customers of educational technology companies such as Blackboard are HEIs who purchase a host of software products closely adapted to their particular needs: learning management systems, tools for providing education at a distance, tools for sending mass notifications to students and staff, academic library software tools, etc.

The infrastructure component of the business models of C-type organisations is also easy to characterise. C-type organisations have access to extensive and diverse resources, they rely on a large number of interdependent processes to operate, and most of their resources are provided in exchange for money. Research universities have access to extensive resources (e.g. financial, human, political, scientific and cultural), rely on hundreds of different processes that take place in the context of a large number of departments and units, and are funded through student-paid tuition, various streams of government funding and subsidies, diverse philanthropic sources, and more. Similarly, major textbook publishers and educational software providers are based on an extensive infrastructure, funded by the revenues created through the sale of their products and services (Greco & Wharton, 2008).



Interestingly, the financial component of C-type organisations in the higher education sector is not unique. These organisations can be for-profit or non-profit, and the way they allocate resources to their different activities is unique only in as much as the infrastructure of these organisations is unique (as discussed in the previous paragraph).

Having described the typical business models of C-type organisations, we now describe bazaar-type (B-type) business models in the higher education sector. Some well-known archetypical B-type organisations in higher education are digital course providers (e.g. OERu, Coursera, OpenLearn), open textbook publishers (e.g. FlatWorld and BCcampus) and open educational technology providers (e.g. Moodle). What are the characteristics of these B-type organisations?

The CVP of B-type organisations in higher education is based on providing customers with flexible products that they can adapt and adjust to their needs. The products are usually free, and are usually digital, though they often include 'freemium' options that offer end users the opportunity to purchase additional features or services for a fee (Anderson, 2009). For example, learners can register for a course they want that is offered by course providers such as Coursera, OERu or OpenLearn, without prerequisites or other limitations. These courses are usually flexible, enabling learners to study at their own pace and move between the course components at a sequence that fits their personal preferences. Many courses are offered for free, and learners can often pay for extra services such as personal attention of faculty, certification of course completion credentials, textbooks, unlimited access, etc. Open textbook publishers such as FlatWorld, BCcampus or Openstax usually offer a free or inexpensive online book that is provided 'as is' and can usually be viewed, printed, shared, remixed and reused. Some of these publishers also offer faculty additional resources such as test banks and manuals. The customers of educational technology organisations such as Moodle receive, free of charge, access to a software product that is supported by an open source community. This includes access to the source code so that users can modify and adjust the code to their own needs. Further support requires paying either the organisation or other, independent, suppliers. This CVP of B-type organisations is reminiscent of the CVP of organisations characterised as disruptive innovators (Christensen & Raynor, 2003) which target customers for whom the CVP of the cathedral HEIs overshoots their needs and/or is overpriced.

The infrastructure component of the business models of B-type organisations is quite diverse. Some B-type organisations, especially for-profit ones such as commercial MOOC providers, are similar to C-type organisations in this sector: they have access to extensive and diverse resources and rely on a large number of interdependent processes to operate, most of which are based on the exchange of money. Other B-type organisations, such as Moodle and BCcampus, have limited access to resources, their processes are fewer and simpler, and many of their core activities, such as coding and content development, are based on volunteer work and contributions. Interestingly, some of the core processes of many B-type organisations in higher education rely on C-type institutions, and especially on university faculty who perform tasks such as MOOC development and authoring open textbooks. These tasks are often performed without significant direct financial compensation.

As explained before, the financial component is not a significant differentiator between C- and B-type organisations in the higher education sector, and mostly reflects the different infrastructures of the organisations.

In contrast with Raymond's 1999 conceptualisation of the bazaar and the cathedral as two separate models, we conceptualise C- and B-type organisations as two extremes on a continuum of business models. A similar conceptualisation of business models in higher education was used by Orr et al. (2018) who classified HEI business models by placing them on a continuum between prospector organisations that are more entrepreneurial and defender organisations which are more focused on protection and stability (Miles & Snow, 1978). Similarly, at the one extreme we find C-like organisations that are well established institutions such as traditional universities and commercial textbook publishers. At the other extreme we find organisations characterised by loosely structured product, the absence or minimization of payments, and the stretching of the small amount of resources through the reduction of processes and through unpaid and volunteer work. In fact, some B-like activities can hardly be classified as organisations, and they are so skeletal that it might seem artificial to analyse their business model. For example, an author who writes an open textbook and puts it online has a very rudimentary business model, yet it is a business model nonetheless. Most HEI organisations fall somewhere along this continuum.

In conclusion of this overview, we propose an analytic framework that assists in the analysis of the business model of organisations in the higher education sector, by determining the location of the HEI's business model on the C-like to B-like continuum. In the following example we will demonstrate how this CAB (Comparative Analysis of Business models) framework can be used to compare the business models of three organisations: a traditional university, a MOOC provider such as Coursera, and a provider of open educational resources such as OERu. OERu is a virtual collaboration of HEIs from around the world which allows OER learners to create flexible learning pathways, including pathways that lead to formal academic credit from recognised education institutions. We chose OERu as a candidate for analysis based on a report by Orr et al. (2018) that identified OERu as an organisation with a high level of online, open, flexible and technology-enhanced education (OOFAT) use. This high score on the OOFAT scale suggested that it would also be a good representative of B-type HEIs.

CAB is based on evaluating a series of elements that characterise the components of the business model of the organisation or organisations that are being analysed. Each of the components of the business model being analysed (e.g. the CVP) comprises elements that are evaluated on a quantitative scale, and the results are presented on a radar diagram that helps visualise the relationships between the business models. In this particular example, the centre of the radar diagram represents elements that characterise C-type organisations, and the outer periphery of the diagram elements that characterise B-type organisations. In the case of the three academic teaching organisations we mentioned (university, Coursera, OERu), the analysis will focus on the CVP and infrastructure components of the organisations, since, as we explained above, the financial component of the business model of organisations in the higher education sector does not differentiate between C-type and B-type organisations beyond what is already covered by the other two components. For brevity, the illustrative example uses only five elements (three for the CVP component and two for the infrastructure component) to compare the business models of C-type and B-type higher education teaching institutions:

- Elements that differentiate between the CVP of the HEIs:
    - How structured is the course of study?
    - How flexible are the studies in regards to time and place of study?
    - How high is the value of a credential awarded to students who complete the course of study?

- Elements that differentiate between the infrastructure of the HEIs:
    - What is the extent of resources that enable the organisation to support its teaching services?
    - How high is the number and complexity of processes that support teaching and learning in the organisation?

A radar diagram that illustrates a hypothetical outcome of the analysis would look like that depicted in Figure 1. Lower numbers denote more structure, lower flexibility, higher value of the awarded certificate, more extensive resources, and a higher number of processes and more complex processes.

The CAB diagram in Figure 1 places the university, the archetypal C-type organisation, at the centre of the diagram, places OERu close to the periphery of the diagram, the area that characterises B-type organisations, and places Coursera about midway between the two areas, with processes and resources that are more similar to universities than to

OERu, with a level of structure that is midway between universities and OERu, and with flexibility and certificate value that are more distant from universities.

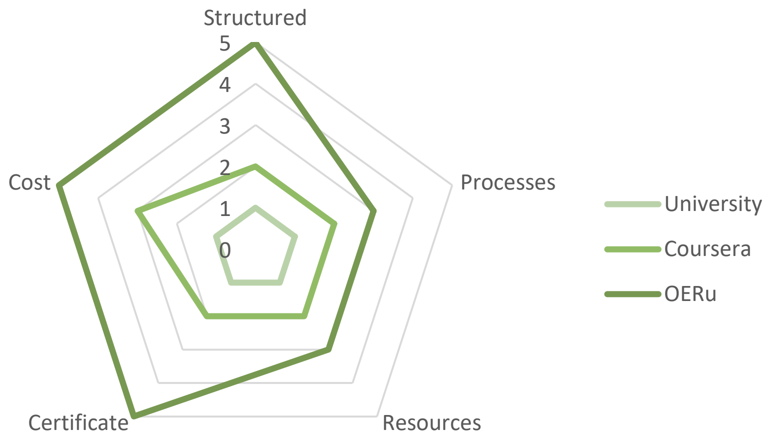



**Figure 1.** An illustrative example of a CAB diagram displaying a simplified comparative analysis of business models of a traditional university, of Coursera and of OERu.

The goal of this illustrative analysis of a hypothetical university and two world-class OER/online education providers is to demonstrate the analytic potential of quantifying the extent to which an HEI`s business model is close to being a C-type or B-type business model. A full empirical analysis would require extensive work to identify and to evaluate many components in the different institutions. Such an analysis is beyond the scope of this conceptual paper.

## Discussion

This discussion section demonstrates how our framework of C-type and B-type business models in the higher education sector can improve our understanding of the implications of digital innovation on this sector. Furthermore, visualising the significant differences in the business models of organisations helps expose false comparisons and analogies, i.e. the proverbial 'comparing apples and oranges'. This analysis is pertinent to this early stage of development of digital innovation in HEI (Orr et al., 2018).

***A Critique of the Replacement of the University Business Model Narrative***

In his book 'MOOCs and a Zero Marginal Cost Education', Rifkin (2014) describes a future in which traditional universities are being replaced by zero marginal cost operations, such as MOOC providers. This claim extrapolates a principle that operates at an organisation that is predominantly a B-type organisation, and which has a very specific CVP and associated infrastructure and financial components, and suggests that this CVP will replace the CVP of C-type organisations. This claim ignores the fact that the components of business models are tightly interlinked, and that changing one significant component in a business model influences other significant components. The fact that the zero-marginal cost can support a B-type organisation does not lead to the conclusion that such an organisation can start providing the significantly more extensive value proposition that is currently offered by C-type organisations. To do so, it will have to add processes and resources, and this will have consequences (e.g. additional costs), that will fundamentally alter the CVP of the organisation (Kalman, 2014) and shift its characteristics much closer to the C-type organisations it is purported to replace.

Similarly, the 'disruptive innovation' theme ignores the fact that many of the resources that the B-type organisations have access to are based at – or originate from – the C-type universities. Most faculty who develop MOOC courses, or the videos, syllabi and other learning materials offered on OERu, are from universities or from other C-type organisations such as museums, major for-profit corporations, government entities, etc. In other words, in the higher education sector, the business models of B-type organisations are highly dependent on resources that originate at C-type organisations. Universities and their faculty still need to invest extensive amounts of time and money to produce courses and other learning materials, and the technology only eliminates the reproduction and dissemination costs (Caswell, Henson, Jensen, & Wiley, 2008; Read, 2011). A report by Duke University estimated that 600 working hours were required in order to build and deliver one MOOC course, including more than 420 hours of effort by the instructor (Belanger & Thornton, 2013). These facts are ignored by those who predict the demise of the university by B-type organisations. The business model of the universities will not be critically disrupted by B-type organisations as long as the B-type organisations' business model continuously relies on resources that can only be provided by a large and diverse number of university-based resources. Coursera and OERu not only fully rely on faculty from C-type organisations to develop the courses and course materials, but also benefit from the fact that their best 'customers' are university graduates, professors and teachers from other C-type institutions (Hansen & Reich, 2015; Koller, Ng, Do, & Chen, 2013).

Finally, the fact that the business model of B-type organisations is heavily dependent on resources from C-type organisations is also one of the reasons that extensive unbundling is less likely to occur in the higher education sector. Many services have been successfully outsourced by universities – the academic textbook industry and academic software industries are just two examples. Nevertheless, few current B-type organisations point to a potential significant unbundling of the three key roles of the current university: teaching, research and accreditation. On the contrary: more and more of the B-type organisations rely on the 'bundled' resources of the university, especially those that arise from the bundling of teaching and research. The category of B-type organisations that is most often discussed in this context is MOOC providers. In particular, the discussion focuses on ways to measure the success or failure of MOOC participants.

### *Measuring Success and Failure in MOOCs*

One of the most extensively debated products of B-type organisations are MOOCs (Kovanović, Joksimović, Gašević, Siemens, & Hatala, 2015). A common critique of MOOCs is that they present an excessively high level of student failure, and, correspondingly, unacceptably low student retention rates. Thus, their product is inferior to that offered by traditional universities (Morris, 2013). This critique, which compares the educational attainment of participants in MOOCs and in traditional universities, makes the classic 'comparing apples and oranges' error, when it measures educational attainment of learners in a B-type organisation but uses criteria that are derived from the C-type business model. This comparison ignores the significant difference between the CVP of MOOC providers versus that of traditional universities. As proposed by Kalz et al. (2015), by Henderikx, Kreijns, and Kalz (2017) and by Reich (2014), criteria for success in MOOCs should be student satisfaction oriented, and reflect the extent to which the MOOC allowed participants to fulfil their intentions. Unlike students who choose C-type organisations and who seek the value proposition associated with them (e.g. a highly structured course of study that leads to a highly valued credential such as a certificate of completion which confirms that the student met all of the requirements defined by the institution), learners come to B-type organisations with goals that are more diverse. Some wish to deepen their understanding of a specific topic. Others wish to master a topic. Others still are looking for personal enrichment and intellectual enrichment, and some teachers and professors wish to improve their own teaching. Research on the outcomes of B-type organisations that uses only criteria that originate in C-type universities limits our ability to understand the outcomes and value these organisations provide. Rather, the success criteria should be as diverse as the reasons participants choose to use the services of a B-type organisation. As Nambisan et al. (2017) suggest, digital innovation requires us to innovate our research methods, for example by using process mining (Van der Aalst, 2011).

### *The Relationships Between B- and C- type Organisations*

One of the interesting insights suggested by the CAB diagram that places the business model of a university in the middle of the graph, Coursera around the university, and OERu at the outer periphery of the diagram (Figure 1) is that the relationship between C- and B-type organisations is that of centre and periphery: The cathedral-type organisations are at the middle, and the bazaar organisations surround it. This suggestive relative placement might appear to be arbitrary in that the scales could have been reversed, placing the B-type organisation at the centre and the C-type at the periphery. Nevertheless, we believe that placing the C-type organisations at the centre and the B-type in the periphery conveys several important ideas. It underlines the fact that we do not yet know how far B-type organisations might yet move away from the centre. The 'B zone' of the graph is a hotbed of experimentation and digital innovation, while the 'C zone' at the centre remains relatively stable and is the reference point for evaluating B-type organisations.

This relative placement also reflects the fact that a business model analysis of B-type organisations reveals extensive dependence on resources that originate in C-type organisations, as well as the blurry boundaries between them (Loeckx, 2016).

### *The Bazaar as a Catalyst of Open Innovation in the Cathedral*

Does the strong dependence of B-type organisations on C-type organisations suggest that B-type organisations feed, parasite-like, on C-type organisations? No. In fact, we claim that B-type organisations have an important role in the higher education sector's ecosystem. Raymond's cathedral and bazaar metaphor for the software industry demonstrates this claim. In the two decades that have passed since Raymond's ideas were published, there have been significant developments in the relationships between the major corporate software developers (e.g. Microsoft, IBM and Apple) and the open source software movement. Open source software is no longer perceived as a threat to the cathedral (i.e. the commercial industry players), but rather as an integral part of the software ecosystem. IBM, known for strong protection of intellectual property (IP) through trade secrets, patents, licensing and other measures, was reported to invest hundreds of millions of dollars in the development of Linux and other open source software projects (Samuelson, 2006), as is demonstrated by the $34 billion acquisition of the open source software company Red Hat (Lohr, 2018). Microsoft too is embracing open source projects (Vaughan-Nichols, 2016), and even Apple, one of the best known examples of the 'walled garden' protective and closed approach to software, is open sourcing some of its products (Finley, 2015).

The best explanation for this shift is open innovation. Open innovation is defined as 'a distributed innovation process based on purposively managed knowledge flows across organizational boundaries, using pecuniary and non-pecuniary mechanisms in line with the organization's business model' (Chesbrough, Vanhaverbeke, & West, 2014, p.27). Extensive research on open innovation in general, and on open innovation in the software industry in particular, reveals that corporations which strategically embrace open software practices can significantly benefit from opening up their developments to the rest of the community, even if that community includes competitors. Furthermore, software companies can benefit from the fact that their employees contribute to external open software projects (Colombo, Piva, & Rossi-Lamastra, 2014). What might have been perceived as a paradox in the past (West & Gallagher, 2006) is now an accepted truth: the bazaar and the cathedral are not mutually exclusive; and innovation, and particularly digital innovation, can be enhanced if it combines IP protection with open practices (Chesbrough et al., 2014; West & Bogers, 2014). Furthermore, according to the theory of disruptive innovation (Christensen & Raynor, 2003), one possible strategy for organisations who wish to avoid being disrupted is through an independent unit which is dedicated to achieving the 'disruptive' goals. Similarly, one possible strategy for HEI who wish to protect themselves from being disrupted would be to use B-type organisations as their autonomous units, which can accelerate innovation. One example of this can be seen in the digital education platform FutureLearn, that is owned by The Open University in Milton Keynes, England (Marszal, 2012). FutureLearn is a controversial project, but the way it influences innovation and policy at the Open University is key to the future survival and success of both organisations (Wilby, 2018).

The analogy between the evolution of the software industry since Raymond's 1999 paper, and the projected evolution of the HEI sector, focuses us on C-type corporations in the software sector who benefitted significantly in the last two decades by opening themselves up to activities and collaborations with B-type organisations. We expect to see such collaborations and interdependencies in the higher education sector too. This concept of open innovation is already prevalent in the development of educational software products such as the Moodle learning management system (Costello, 2014), and we project that much more is to be expected in the near future. We are not aware of research findings on this topic, but there is plenty of anecdotal evidence that faculty who are involved in the development of open educational resources such as MOOCs and open textbooks, are also catalysts of innovation within their own HEI's, integrating nontraditional resources in teaching (Conole, 2012) and experimenting in approaches to using online methods to increase the effectiveness of on-campus teaching (Bates, 2013). Thus, HEIs who wish to successfully deal with the major changes that we are facing as a society in the digital age, should not perceive B-type organisations as existential threats to their future, but rather embrace them, and even nurture and develop them. These B-type organisations enable the C-type organisation to innovate, to bridge boundaries and catalyse cooperation, innovation and creativity.

***Empirical Validation of the Suggested Model***

This conceptual paper is theoretical by its nature. We propose an alternative way to examine the future of higher education ecosystems. Future work could empirically examine the validity of the model and its predictions. In the first stage, the CAB model should be further developed empirically. Content analysis of documents, media coverage, case studies, and interviews will enable validating the model and mapping its different dimensions. In the second stage, a survey can be developed that will analyse and map organisations on the different dimensions of the model. This mapping could, for example, identify those organisations that are more likely to enable synergy between B- and C-type organisations. In the third stage, a longitudinal study will help identify future trends in the relationships between B- and C-type organisations and validate the prediction of the model regarding the synergistic relationships between them.

In conclusion, we propose that despite the persistent tension between B- and C-type organisations in the higher education sector, innovation in this sector will not emerge from B-type organisations disruptively eliminating C-type organisations. Rather, we predict that similar to the software industry, the higher education sector too will develop into an ecosystem populated by interdependent organisations that occupy various niches, and whose business models are characterised by various degrees of 'cathedral-ness' and 'bazaar-ness'.

## Conclusion

This conceptual paper analyzes the impact of digital innovation on business models in the higher education sector. It offers an alternative to the common 'replacement' narrative of the upcoming demise of the incumbent university ('cathedral') business model by open education based ('bazaar') business models. Using an analogy from the software sector, we suggest that the future higher education ecosystems could still be dominated by C-type universities and other HEIs, but that their business models will be based on synergistic relationships with a host of other organisations. These other organisations will represent many gradations on the continuum between C- and B-type organisations. The synergy with B-type organisations will catalyse open innovation in the universities, and keep them better attuned to changing societal needs and preferences. Consequently, universities will not only offer a better CVP to their students, but also turn out university graduates who are better prepared to benefit from the entire higher education ecosystem. These graduates will improve skills (e.g. self-regulation) required to go on to be life-long learners who effectively use open education products such as MOOCs to remain intellectually and professionally up-to-date. Our alternative to the 'demise of the university' replacement narrative acknowledges that the bazaar is an important experimental space that will guide universities to develop and innovate in ways that answer the needs of the students of today and tomorrow.

A key to promoting the healthy growth of this digital innovation-based higher education ecosystem is developing new research tools and novel measures that augment the traditional measures used to research university education. Rather than assuming that there is a single gold standard for successful HEI, the diversity of our research tools should reflect the diversity of organisations, business models and learners in the higher education ecosystem.

### *Note*

1. For simplicity, this paper focuses on the undergraduate educational 'product' of HEIs.

3

# Chapter 3

## An Empirical Investigation of the Antecedents of Learner-Centered Outcome Measures in MOOCs

This chapter is based on:

Rabin, E., Kalman, M, Y., & Kalz, M. (2019).
An empirical investigation of the antecedents of learner-centered outcome measures in MOOCs

*International Journal of Educational Technology in Higher Education. 16(14).*
*https://doi.org/10.1186/s41239-019-0144-3*

## Abstract

This research revealed the antecedes of two learner-centered outcome measures of success in massive open online courses (MOOCs): learner satisfaction and learner intention-fulfillment. Previous studies used success criteria from formal education contexts placing retention and completion rates as the ultimate outcome measures. We argue that the suggested learner-centered outcomes are more appropriate for measuring success in non-formal lifelong learning settings because they are focused on the learner's intentions, rather than the intentions of the course developer. The behavioural measures of 125 MOOC participants who answered a pre- and a post- questionnaire were harvested. The analysis revealed that learner satisfaction was directly affected by: the importance of the MOOC's benefits; online self-regulated learning - goal setting; number of video lectures accessed; and, perceived course usability. Age and the number of quizzes accessed indirectly effected learner satisfaction, through perceived course usability and through number of video lectures accessed. Intention-fulfillment was directly affected by: gender; the importance of the MOOC's benefits; online self-regulated learning - goal setting; the number of quizzes accessed; the duration of participation; and, perceived course usability. Previous experience with MOOCs and the importance of MOOC's benefits, indirectly affected intention-fulfillment through the number of quizzes accessed and perceived course usability.


***Keywords:*** MOOC; Perceived learning outcomes; Structural equation modeling; Student satisfaction; Intention-fulfilment; Learning analytics; Educational data mining

## Introduction

Lifelong learning received extensive support from recent technological developments such as online learning in general, and MOOCs in particular (Kalz, 2015). This development is accompanied by controversy. One key criticism of MOOCs is the high drop-out rates (Gardner & Brooks, 2018; Reich & Ruipérez-Valiente, 2019). These rates are, on average, 93% (Chuang & Ho, 2016; Jordan, 2014; Margaryan, Bianco, & Littlejohn, 2015). Furthermore, most MOOC participants who earn certificates for completing the course are experienced learners with a strong academic background (Christensen et al., 2013; Daily, 2014; Guo & Reinecke, 2014; Hansen & Reich, 2015; Koller, Ng, Chuong, & Zhen-ghao, 2013; Reich & Ruipérez-Valiente, 2019). Though it is true that MOOC dropout rates are very high, the question is whether completion rate is the appropriate measure for evaluating the success of this new form of lifelong learning. Completion rate is a success criterion borrowed from formal education contexts where students enroll in courses with the goal of completing them, and of satisfying the learning outcomes defined by the instructor. Rather, students may enroll in MOOCs for a variety of reasons (Littlejohn, Hood, Milligan, & Mustain, 2016; Onah, Sinclair, & Boyatt, 2014; Wang & Baker, 2018), and MOOC participants may have a variety of expected learning outcomes. For example, MOOC participants may achieve their learning goals by engaging in only a segment of the course (Ho et al., 2015; Liyanagunawardena, Parslow, & Williams, 2013). It has been proposed that the success of lifelong learning in MOOCs should be evaluated not through traditional, instructor-focused measures such as completion rates, but rather through more learner-centered measures such as learner satisfaction and the fulfillment of learner intentions (Henderikx, Kreijns, & Kalz, 2017; Reich, 2014).



## Learner satisfaction, and intention fulfillment

Learner satisfaction reflects students' perception of their learning experience (Kuo, Walker, Schroder, & Belland, 2014; Littlejohn et al., 2016) and is defined as a student's overall positive assessment of his or her learning experience (Keller, 1983). While some authors have found positive correlations of student satisfaction with post-secondary student success (Chang & Smith, 2008), and a positive relationship between learning satisfaction and the intention to use e-learning (Liaw & Huang, 2011; Roca, Chiu, & Martínez, 2006), a recent study of student data of the Open University by Rienties and Toetenel (2016) has found that retention and satisfaction are not correlated. The authors explain these findings from a formal distance education context with the fact that learning is not always fun, and requires effort. While this explanation is relevant in the context of degree-seeking learners in formal education, the role of learner satisfaction in the open learning context of MOOCs should not be underestimated, as these learners participate in the courses for different reasons than degree-seeking students in formal education.

Another success criterion that has been proposed is learner intention-fulfillment. Intention-fulfillment emerges as a promising success measure of open courses and

MOOCs, since it takes into account the personal objectives that the learners intend to achieve, rather than external success criteria (Henderikx et al., 2017). In MOOCs and in other forms of open education, a successful learning experience can take a variety of forms, ranging from viewing a single lecture, attaining a specific skill, or studying a topic of interest, to studying a whole course and fulfilling all of its requirements.

This study focuses on learner satisfaction and learner intention-fulfillment as two learner-centered success measures and examines the factors that impact these subjective success measures in the context of a mid-sized (circa 2000 participants) MOOC on the recent history of the Middle-East. The goal of this study is to identify key factors contributing to MOOC learner satisfaction and intention-fulfillment. We examine how these two dependent variables are predicted by personal learner characteristics (demographic characteristics, previous experience with MOOCs), learner dispositions (self-regulated learning, course outcome beliefs), learner behaviour in the MOOC (e.g. number of video lectures accessed, number of quizzes accessed), and perceived course usability (e.g. ease of navigation, website usability). Understanding the predictors of the two success criteria, learner satisfaction and intention-fulfillment, will contribute to theories of learner motivation and behavior in open online environments, as well as help create more personalized courses and provide lifelong learners with better support in open learning contexts.

The rest of the paper is organized as follows. After a review of related work on possible predictors of the learner-centered outcome variables, we propose research hypotheses. This is followed by the research model, a Method section detailing the research methods, the participants, and the instruments used for data collection, and a Data analysis section. We conclude with the results and a discussion of the findings.

## Predictors of learner satisfaction and of intention-fulfillment

### *Demographic background*

MOOC learners are a heterogeneous group, comprising of male and female participants of all ages, from across the world, with different educational, socio-economic and psychological characteristics (Chuang & Ho, 2016; Koller, Ng, Do, & Chen, 2013). The diversity of MOOC learners has been discussed in several studies. Some earlier studies did not identify an influence of gender on achievement or on completion rates (Breslow, Pritchard, & DeBoer, 2013; Cisel, 2014; Kizilcec, Piech, & Schneider, 2013; Morris, Hotchkiss, & Swinnerton, 2015), while other studies, such as Garrido, Koepke, Anderson, and Mena (2016) found that women are more likely than men to complete a MOOC or obtain certification.

Furthermore, there are inconsistent findings about the association between age and academic achievement. Guo and Reinecke (2014) found a positive correlation between age and grades, while Breslow et al. (2013) did not find such a correlation. In an examination of completion rates, Morris et al. (2015) found that course completers were on average older, while those who dropped out in the first week of the course were on average the youngest group. Based on these findings, we propose the following hypotheses:

> **Hypothesis 1** Gender will be associated with learner behaviour in the course, and with the course outcomes: learner satisfaction and participant intention-fulfillment.

> **Hypothesis 2** Age will be positively associated with a higher level of participant activity in the course, with higher perceived usability and with better course outcomes: learner satisfaction and participant intention-fulfillment.

***Previous experience in MOOCs***

Scholars highlight the high level of previous knowledge and competencies needed to be successful in a MOOC (Santos, Costa, & Aparicio, 2014). Most of the participants who earn certificates for completing MOOCs are experienced learners with a strong academic background (Christensen et al., 2013; Daily, 2014; Guo & Reinecke, 2014; Hansen & Reich, 2015; Koller, Ng, Chuong, & Zhenghao, 2013). Based on these findings we propose the following hypothesis:

> **Hypothesis 3** Previous experience in MOOCs will be positively associated with a higher level of participant activity in the course, and with better course outcomes: learner satisfaction and participant intention-fulfillment.

***Course outcome beliefs***

Course outcome beliefs is a variable that describes the expectations learners have regarding the outcomes of participating in the MOOC. A person may believe that taking a MOOC will result in positive outcomes such as more opportunities in the labor market or negative outcomes such as losing leisure time or creating stress. Those outcome beliefs can affect the learner's behaviour as well as the learner's evaluation of the course.

> **Hypothesis 4** Positive course outcome beliefs will be positively associated with a higher level of participant activity in the course, and with better course outcomes: learner satisfaction and participant intention-fulfillment.

***Self-regulated learning***

MOOCs and other forms of open online education are not only open in access, but also open for participants to choose their learning behavior, their learning path and their learning schedule (Kizilcec, Perez-Sanagustín, & Maldonado, 2017; Margaryan et al., 2015; Van den Beemt, Buijs, & Van der Aalst, 2018). The self-paced nature of online courses treats the learner as an active agent, and provides learners with the freedom to select and control the resources and tools that they are using. Thus, online learning requires a high level of self-regulation. Zimmerman (2000) defines self-regulation as "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" (p. 14). Self-regulation is a context-specific process.

In the context of learning, self-regulated learning (SRL) is defined as students' proactive actions aimed at acquiring and applying information or skills that involve setting goals, self-monitoring, time management and regulating one's efforts towards learning goal fulfillment (Järvelä, Malmberg, & Koivuniemi, 2016; Reimann, Markauskaite, & Bannert, 2014; Tabuenca, Kalz, Drachsler, & Specht, 2015; Zimmerman, 1990). Several studies found a positive correlation between SRL and satisfaction in online courses (Artino, 2007; Li, 2019; Puzziferro, 2008). Learners, who do not regulate their learning process, may experience increased dissatisfaction (Sun & Rueda, 2012). As well, goal setting and strategic planning were positively predict goal attainment in MOOCs and help seeking negatively predicts goal attainment (Kizilcec et al., 2017). These findings lead us to the following hypothesis:

> **Hypothesis 5** Higher self-regulated learning capabilities will be positively associated with a higher level of participant activity in the course, and with better course outcomes: learner satisfaction and participant intention-fulfillment.

***Perceived course usability***
The usability of the course website as perceived by the user (perceived course usability) is influenced by factors such as the user's perception of the course website, and the organization of the course materials into logical and understandable components (Eom, Wen, & Ashill, 2006). Usability refers to whether a system can be used with effectiveness and efficiency to enable users to achieve specified goals in a particular context of use (ISO 9241-11, 1998). Usability affects students' learning effectiveness and overall learning experience, and the level of usability affects the satisfaction level and the learning outcomes of distance learners (Eom et al., 2006). We thus hypothesize that:

> **Hypothesis 6** The level of perceived course-usability will be positively associated with better course outcomes: participant learner satisfaction and intention-fulfillment.

***Learning behaviour in MOOCs***
In open learning environments like MOOCs learners can study when, where and how they wish, alone or with others, and with fewer restrictions on time or space compared to traditional online-courses. Learning behaviour in MOOCs is mostly visible through access and usage patterns of different types of resources. The participants can learn in different learning sequences by watching video lectures and by interacting with different MOOC resources (Van den Beemt et al., 2018). The learning path can deviate from a linear course, learners can start the courses later than the original launch date, can view lectures several times, and do exercises and take quizzes several times. For example, initial findings suggest that successful MOOC certificate earners view only 78% of the course content and skip the rest (Guo & Reinecke, 2014). Successful certificate earners are also more engaged in non-linear navigation behaviour than non-certificate earners. They "jump" backwards to revisit earlier lectures or assessments up to three times more often than non-certificate earners (Guo & Reinecke, 2014). Davis, Chen, Hauff, and Houben (2016) showed that learners who successfully passed the course are more interested in

their quiz scores than learners who did not, and they used progress tracking tools more often. Such learning patterns show that learners who successfully pass the course use better self-regulated learning strategies than those who did not pass the course. Thus, we hypothesize that:

> **Hypothesis 7** The number and variety of course activities performed, and the time spent on the course will be positively associated with the participant's perceived course usability, and with better course outcomes: learner satisfaction and intention-fulfillment.

In summary, the objective of this study was to identify how MOOC participant characteristics and pre-course disposition affect participant learning behaviour in the course, as well as how these predictors affect the perceived course usability, and finally, how all of these variables predict the learning outcomes: learner satisfaction and intention-fulfilment. Figure 1 illustrates the research model of the study.

## Method

### *Participants*

Participants in a nine-week massive open online course (MOOC) of the Open University in Israel were surveyed for this study. The open access course was built on a Moodle platform and dealt with the political and sociological aspects of the "Middle East in our times". A participant was defined as a person who enrolled to the MOOC and who participated in at least one activity in the course. The course was open and free to the Hebrew speaking public without any prerequisites, and did not award academic credit other than a certificate of participation. Each week, a new topic was opened to the participants and the participants were able to watch video lectures, answer multiple choice questions and quizzes, and respond to discussion questions in discussions forums. Participants who fulfilled all of the course assignments received a certificate of participation without charge.

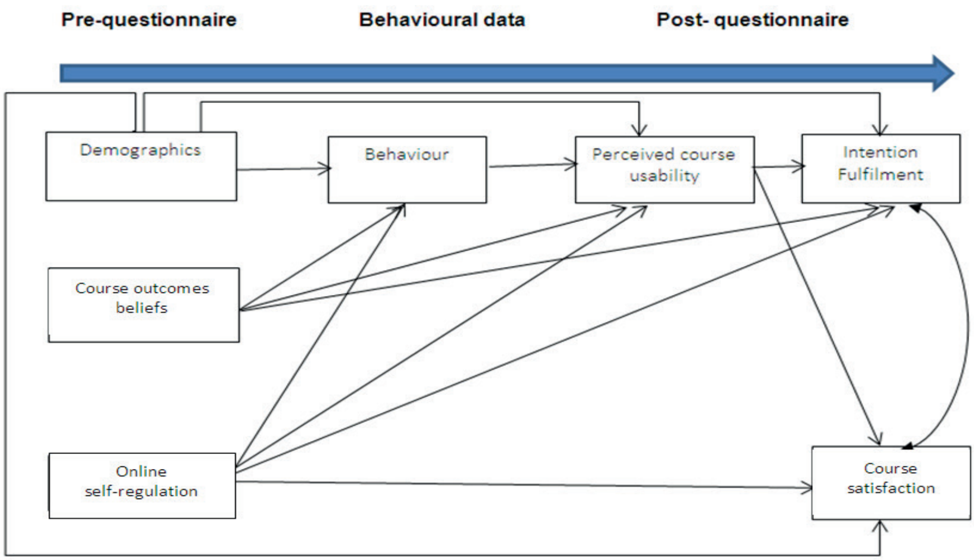



**Figure 1.** A model of key factors influencing learner satisfaction and intention-fulfillment in MOOCs. The row above the model presents the timeline and the three stages of data collection

During the course, participants' online activities were recorded in a log-file. All MOOC participants were invited to respond to two questionnaires: a pre- and a post-course online survey. The survey was not mandatory. Of the 2007 participants who enrolled to the course and participated in at least one activity, 377 (18.7%) participants responded to the pre-questionnaire and 190 (9.5%) participants responded to the post-questionnaire. In total, 125 (6.2%) participants took part in all three stages of the study by answering the pre-questionnaire, participating in at least one MOOC activity, and answering the post-questionnaire. This paper reports findings on this group.

The participants in this study were diverse. Ages ranged from 18 years old to 85 ($M$ = 61, $SD$ = 14.01). Course participants under the age of 18 were not included in the study. 56% of the participants were male and 44% were female. On average, the participants rated themselves as highly skilled internet users on a scale of 1–7 (Very low to very high Internet skills) ($M$ = 6.23, $SD$ = 0.65), though a majority (63.7%) reported that this MOOC was their first online learning experience. The sample of participants included in the study is demographically similar to the population which enrolled in the MOOC: Age and gender of the participants were compared to demographics reported by 457 (23%) course enrollees who responded to a short survey at the beginning of the course that included questions about the gender and the age of the participants. The participant pool was not significantly different from the enrollee pool: A Chi-square test reveals that the gender distribution in the two samples was similar($Chi^2_{(1)}$ = 0.04, $p$ = .84, $Male_{\text{brief survey}}$ = 57%), and t-test for independent samples revealed that there are no differences in the participants age ($t_{(580)}$ = 0.83, $p$ = .41, $M_{\text{brief survey}}$ = 59.97, $SD$ = 16.02). The similarity in gender and age between the large sample and the research sample enables us to generalize the research results beyond the sample of participants who met all of the inclusion criteria.

### *Assessments and measures*

As explained in Participants section, this study comprised three stages of data collection: a pre-course questionnaire, behavioural data collected from log-files, and a post-course questionnaire. All MOOC participants were invited to answer the pre-course questionnaire via email immediately after they had enrolled to the MOOC, and a reminder was sent after 1 week to those who did not yet respond to the questionnaire. In addition, an invitation to participate in the pre-course questionnaire was posted to the MOOC bulletin board. All participants were informed that responding to questionnaire is voluntary, and signed an informed consent before taking the questionnaire. Similarly, on the last week of the MOOC, all enrollees were invited to the post-questionnaire by email, with a reminder after 1 week. A unique identifier connected the survey responses and the behavioural data. An anonymization process had been implemented ahead of the statistical analysis.

### *Pre-questionnaire*

The pre-questionnaire included three sections – demographics, course outcome beliefs, and online SRL.



**Demographics** Participants reported gender, age, and number of MOOCs previously taken. Gender was a two-category variable with male coded as 1 and female coded as 2. Age was reported in years. Previous experience with MOOCs was measured by the number of MOOCs that the participants took up to the time of the survey. The variable was coded as a dummy variable – Took ('1') or did not take ('0') at least one MOOC in the past.

**Course outcome beliefs** Two indices measured course outcome beliefs. (A) Importance of the benefits of participating in the MOOC ('importance of MOOC's benefits'): Eighteen items including statements such as 'will increase my chances in the labour market', and 'will allow me to do my job better' (Cronbach's alpha = .92). (B) Importance of the disadvantages of participating in the MOOC ('importance of MOOC's disadvantages'): Five items including statements such as 'will limit my free time with family and friends', 'will force me to buy a new multimedia computer' (Cronbach's alpha = .74).

**Online SRL** Two indices measuring online self-regulation were adapted from the OLSQ (Barnard, Lan, To, Paton, & Lai, 2009). (A) Goal setting: The ability to set goals for the learning process (e.g. 'I maintain a high standard of learning in my online courses', and 'I set short-term (daily or weekly) goals as well as long-term goals (monthly or for the semester) ') (Cronbach's alpha = .87). (B) Environmental structuring: The ability to arrange the location for studying (e.g. 'I choose the location where I study to avoid too much distraction', 'I find a comfortable place to study') (Cronbach's alpha = .85). Each index included five items. The other four self-regulation scales had not been used in order to decrease the load on the participants. All the indices were on a 7-point Likert scale.

### *Behavioural measurements*
The behavioural measurements were extracted from the log file of the course. The measurements included: (A) The number of video lectures that the participant accessed during the course ('number of videos accessed'). The participants could access the same video lecture more than once. (B) The number of quizzes that the participant accessed during the course ('number of quizzes accessed'). The quizzes were self-evaluation activities that enabled participants to assess their knowledge. The participants could access the same quiz more than once. (C) The number of discussion forums that the participant accessed during the course ('number of forums accessed'). Participants who asked to get a certificate of course completion were asked to post at least two comments to a weekly discussion forum. (D) The duration of time the MOOC was taken ('duration of participation'). This measure was calculated by subtracting the time of the last log-on of the participant in the course, from the time of the first log-on. (E) The total number of MOOC activities the participant participated in, including the number of lectures, quizzes and forums accessed ('number of activities accessed'). This behavioural measurements were log-transformed in order to get a normal distribution of the variable. (F) Whether the participant received a certificate of course completion ('receiving completion certificate'). The criteria for receiving the certificate were based on completing a minimal quota of course activities.

### *The post-questionnaire*
The post-questionnaire included three sections – perceived course usability, learner satisfaction and intention-fulfillment.

**Perceived course usability** Seventeen items on 7-point Likert scale ranging from 1 'totally don't agree' to 7 'strongly agree', including items such as 'It is easy to learn to use this MOOC virtual learning environment', 'I know where to go in this MOOC virtual learning environment' (Cronbach's alpha = .84).

**Learner satisfaction** Single item on a Likert scale ranging from 1 'very unsatisfied' to 7 'extremely satisfied': 'How satisfied have you been with this MOOC?'

**Intention-fulfillment** Four items on 7-point Likert scale ranging from 1 'totally don't agree' to 7 'strongly agree', including items such as: 'I achieved my personal learning goals by participating in this MOOC', 'the MOOC met my expectations' (Cronbach's alpha = .89).

### *Data analysis*
Bivariate correlation analyses (Pearson product moment) were performed in order to identify the predictors of learner satisfaction and intention-fulfillment. The Pearson correlation coefficient ($r$), ranging between – 1 and + 1, indicates the strength and the direction of the relationship of each independent variable with the other independent and dependent variables.

As a preliminary step in preparation for the linear regression analysis, the correlations between the independent variables were evaluated to identify multicollinearity.

Afterwards, stepwise hierarchical linear regression models were performed with learner satisfaction and intention-fulfillment as the dependent variables. Independent variables with higher than bivariate correlation of .60 were entered into the same regression model in a stepwise manner in order to avoid violation of the regression assumptions. Furthermore, in every regression analysis, variance inflation factor (VIF) and tolerance values were checked in order to find evidence of multicollinearity.

Finally, structural equation model-based PLS methodology was employed to examine the shared effect of the independent variables on each other and on the dependent variables: learner satisfaction and intention-fulfillment. PLS is well suited for this research because it is useful for early stages of theory building and testing (Chin, 1998). To reduce the model complexity, only variables that had been identified as significant predictors in the linear regression were entered the model.

## Results

The study's two dependent variables are the outcome measures of success: learner satisfaction and intention-fulfillment. Table 1 presents the results of a Pearson bivariate correlation analyses between the independent variables and the two dependent variables. Appendix includes the Pearson bivariate correlations between *all* the research variables. Learner satisfaction and intention-fulfillment were found to be highly correlated ($r = .78$, $p < .001$).



**Table 1.** Pearson correlations between the predictor variables and the two dependent variables

| Variables | Intention-fulfilment | Learner satisfaction |
|---|---|---|
| **Pre-course** | | |
| Age | .15 | .22* |
| Gender | –.15 | .11 |
| Previous experience with MOOCs | .07 | .01 |
| Importance of MOOC's benefits | .26** | .29** |
| Importance of MOOC's disadvantages | .04 | .05 |
| Online SRL – Environmental structuring | .21* | .21* |
| Online SRL – Goal setting | .33*** | .31** |
| **Behavioural variables** | | |
| Number of videos accessed | .30*** | .43*** |
| Number of quizzes accessed | .37*** | .37*** |
| Number of forums accessed | .24*** | .16* |
| Duration of participation | .24* | .19* |
| Number of activities accessed | .36*** | .34*** |
| Receiving completion certificate | .41*** | .38*** |
| **Post-course variables** | | |
| Perceived course usability | .37*** | .44*** |

Gender - male coded as '1' and female coded as '2', Previous experience with MOOCs – yes coded as '1', Receiving completion certificate - received a certificate coded as '1'.
$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

***Learner satisfaction***

A hierarchical linear regression to predict the level of learner satisfaction was performed in four stages. In the first stage, demographic predictors – *age*, *gender* and *previous experience with MOOCs* entered to the regression in Enter mode. In the following steps the variables were entered in a stepwise method in order to avoid multicollinearity. The second step included predictors from the pre-questionnaire that were found to be in correlation with learner satisfaction (see Table 1). The variables entered were: the *importance of MOOC's benefits*, the *importance of MOOC's disadvantages*, the *level of online SRL in environmental structuring* and the *level of SRL in goal setting*. In the third step, behavioural indices that were in correlation with learner satisfaction (see Table 1) were entered. The variables entered were: the *number of videos accessed*, the *number of forums accessed*, the *number of quizzes accessed*, and the *duration of participation*. In the fourth step, the *perceived level of the course's usability* was entered into the analysis. The results of the four stages are presented in Table 2.

As seen in Table 2, the more the participants set goals for their online learning and the more they perceived the importance of the benefits of taking the MOOC as high, the more they reported higher satisfaction from the course. From the behavioural measurements, the more video lectures the participant accessed, the higher their level of course satisfaction. Lastly, the higher the participants' perceived course usability, the more they reported satisfaction from the course. All the variables together explained 42% of the variance of the learner satisfaction variable.

***Intention-fulfillment***

A hierarchical linear regression was performed in four stages in order to predict the level of participant *intention-fulfillment*. In the first stage, the three demographic predictors *age*, *gender* and *previous experience with MOOCs*, were entered to the regression as control variables. In the following three stages, the variables were entered in a stepwise manner in order to avoid multicollinearity. In the second stage, predictors from the pre-survey that were found to correlate with the level of intention-fulfillment (see Table 1) were entered. These variables were: the *importance MOOC's benefits*, and the *level of online SRL: environmental structuring* and *goal setting*. In the third stage, behavioural predictors that were found to correlate with the level of intention-fulfillment (see Table 1) were entered. The variables entered were: *number of videos accessed*, *number of forums accessed*, *number of quizzes accessed*, and *duration of participation*. In the fourth stage the *perceived course's usability*, was entered. The results of the four stages are presented in Table 3.

**Table 2.** Stepwise linear regression predicting learner satisfaction, performed in four stages

| Stage | Predictor | Beta | T | $R^2$ | $\Delta R^2$ | F(df) |
|---|---|---|---|---|---|---|
| 1 Demographics | | | | .08 | .08 | 2.46 (2,85) |
| | Gender | .17 | 1.62 | | | |
| | Age | .27 | 2.48* | | | |
| | Previous experience with MOOCs | −.05 | 0.47 | | | |
| 2 Pre-questionnaire | | | | .21 | .13** | 4.36*** (5,83) |
| | Gender | .13 | 1.27 | | | |
| | Age | .22 | 2.15* | | | |
| | Previous experience with MOOCs | −.09 | 0.93 | | | |
| | Online SRL – Goal setting | .27 | 2.64* | | | |
| | Importance of MOOC's benefits | .22 | 2.15* | | | |
| 3 Behaviour | | | | .32 | .11*** | 6.50*** (6,82) |
| | Gender | .09 | 0.98 | | | |
| | Age | .14 | 1.44 | | | |
| | Previous experience with MOOCs | −.06 | −0.64 | | | |
| | Online SRL – Goal setting | .26 | 2.72** | | | |
| | Importance of MOOC's benefits | .18 | 1.90^ | | | |
| | Number of videos accessed | .36 | 3.83*** | | | |
| 4 Usability | | | | .42 | .10*** | 8.39*** (7,81) |
| | Gender | .08 | 0.88 | | | |
| | Age | .12 | 1.28 | | | |
| | Previous experience with MOOCs | −.09 | −1.08 | | | |
| | Online SRL – Goal setting | .23 | 2.63** | | | |
| | Importance of MOOC's benefits | .16 | 1.87^ | | | |
| | Number of videos accessed | .30 | 3.36*** | | | |
| | Perceived course usability | .33 | 3.70*** | | | |

Gender - male coded as '1' and female coded as '2', Previous experience with MOOCs – yes coded as '1'
^$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

**Table 3.** Stepwise linear regression predicting intention-fulfillment, performed in four stages

| Stage | Predictor | Beta | T | $R^2$ | $\Delta R^2$ | F(df) |
|---|---|---|---|---|---|---|
| 1 Demographics | | | | .04 | .04 | 1.42 (3,95) |
| | Gender | -.13 | 1.22 | | | |
| | Age | .12 | 1.13 | | | |
| | Previous experience with MOOCs | .07 | 0.69 | | | |
| 2 Pre-questionnaire | | | | .19*** | .15* | 4.42*** (5,93) |
| | Gender | -.17 | -1.79 | | | |
| | Age | .07 | 0.67 | | | |
| | Previous experience with MOOCs | .02 | 0.23 | | | |
| | Online SRL – Goal setting | .29 | 3.05** | | | |
| | Importance of MOOC`s benefits | .22 | 2.35* | | | |
| 3 Behaviour | | | | .32*** | .13*** | 6.34*** (7,96) |
| | Gender | -.16 | -1.84^ | | | |
| | Age | .05 | 0.59 | | | |
| | Previous experience with MOOCs | -.02 | -0.24 | | | |
| | Online SRL – Goal setting | .30 | 3.41*** | | | |
| | Importance of MOOC`s benefits | .19 | 2.10** | | | |
| | Duration of participation | | 2.51* | | | |
| | Number of quizzes accessed | .23 | 2.51* | | | |
| 4 Usability | | | | .37*** | .05** | 7.05*** (8,95) |
| | Gender | .18 | -2.14* | | | |
| | Age | .02 | 0.28 | | | |
| | Previous experience with MOOCs | -.04 | -0.49 | | | |
| | Online SRL – Goal setting | .29 | 3.34*** | | | |
| | Importance of MOOC`s benefits | .18 | 2.11** | | | |
| | Duration of participation | .20 | 2.32* | | | |
| | Number of quizzes accessed | .17 | 1.92* | | | |
| | Perceived course usability | .25 | 2.92** | | | |

*Note:* Gender - male coded as `1` and female coded as `2`, Previous experience with MOOCs – yes coded as `1`.
^$p$<.10,*$p$<.05, **$p$<.01, ***$p$<.001

As seen in Table 3, female participants reported a higher level of intention fulfillment than male participants. The more participants set goals for their online learning and the more they perceived the importance of the benefits of taking a MOOC to be high, the more they reported higher intention-fulfillment.

From the behavioural measurements, the longer the duration of participation, and the higher the number of quizzes accessed during the course, the more they reported higher intention-fulfillment. Lastly, the higher the participants` perceived course usability, the more they reported that their intentions were fulfilled. All the variables together explained 37% of the variance of the intention-fulfillment variable.

### *Prediction of learner satisfaction and intention-fulfilment with SEM analysis*

Structural equation modeling (SEM) - based PLS methodology with maximum-likelihood estimation was conducted using Amos 22 in order to model the relationships between the variables. Missing variables were rare and were imputed using regression estimation (Schreiber, 2008). The variables for the SEM analysis were selected based on the significant correlations identified in Table 1, and the significant coefficients identified in Tables 2 and 3. The results are presented in Figure 2. All paths in the model are significant, and non-significant paths were removed. Sample size had been found sufficient for the number of the variables that had entered into the model (Bentler & Chou, 1987; Tabachnick & Fidell, 2001). The model goodness of fit is satisfactory ($Chi^2_{(33)} = 40.29$, $p = .18$, $CFI = .97$, $TLI = .94$, $NFI = .88$, $RMSEA = .03$). The model explained 36% of the variance of intention-fulfilment and 25% of the variance of the learner satisfaction. The results of the SEM analysis demonstrate several phenomena on the effects of the study's independent variables on learner satisfaction and intention-fulfillment.

The demographic variables influence the dependent variables (DVs) in several ways. Gender had a direct effect on intention-fulfilment, but not on the level of satisfaction. Female participants report that they fulfil their intentions more than males, but there are no significant differences between female and male participants in the level of learner satisfaction. On the other hand, age did not have a direct effect on the DVs, but rather affected the number of videos lectures the participants accessed. There was a positive correlation between participant age and the number of video lectures viewed. Further to that, the number of video lectures viewed was positively correlated with the level of learner satisfaction. In summary, older age predicted viewing more video lectures, which in turn predicted a high level of learner satisfaction. In contrast, age did not predict the level of intention-fulfilment, neither directly or indirectly.

The number of MOOCs previously taken did not directly predict the DVs. Rather, it predicted the number of quizzes the participant accessed. Those who participated in at least one MOOC in the past accessed more quizzes than those who did not participate in a MOOC in the past. The number of quizzes accessed was directly positively correlated with the level of intention-fulfilment. It was also indirectly correlated with the level of intention-fulfillment, as it was mediated by the level of perceived usability of the course.

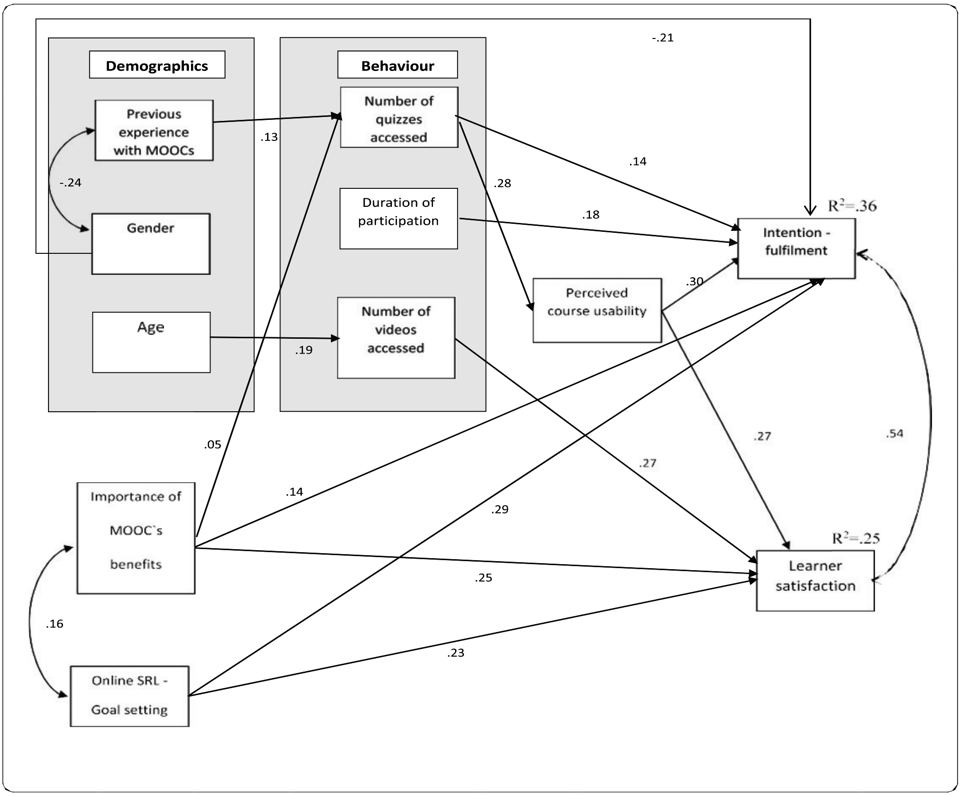



**Figure 2.** SEM analysis presenting the significant factors influencing learner satisfaction and intention-fulfillment. *Note*. All the paths are significant at a $p < .05$ level. The estimations are presented using a standardized coefficient.

The positive outcome beliefs and the goal setting variable affected the DVs directly as well as indirectly. The level of the importance of MOOC's benefits positively affected the level of intention-fulfilment, and thus participants who expected to gain benefits from their participation in the MOOC were more likely to report that they fulfilled their intentions. Another interesting finding is that of the importance of MOOC's bene- fits on the number of weekly quizzes that the participant accessed: Participants who expected a positive outcome from the MOOC, accessed more quizzes than those who didn't expect positive outcomes. The number of quizzes accessed, as was already mentioned, is positively correlated with intention-fulfilment. Note though that the number of quizzes accessed had no significant mediating effect on the connection between the importance of MOOCs' benefits and the level of intention-fulfilment (*Sobel Z* = 1.35, $p$ = .09). Interestingly, the level importance of MOOC's disadvantages did not have an effect on the DVs and for that reason is not shown in the final model.

Participants, who regulated their learning by setting goals, reported higher levels of learner satisfaction and intention-fulfilment without any mediation by the behavioural variables. On the other hand, the ability to self-regulate learning by structuring the learning environment, did not affect the DVs and is thus not shown in the model.

Surprisingly, the duration of participation was not affected by any pre-course variables. On the other hand, duration of participation positively affected the level of intention-fulfilment, but not learner satisfaction. Lastly, the perceived course usability was predicted only by the number of quizzes the participant accessed, and it was positively correlated with both learner satisfaction and intention-fulfilment.

## Discussion

The goal of this research was to better understand the predictors of two important learner-centered outcome measures of success in massive open online courses: learner satisfaction and learner intention-fulfillment. In contrast with previous studies, which focused on the fulfillment of the course developers' intentions and placed retention and completion rates as the ultimate outcome measures, we place learner satisfaction and learner intention-fulfillment as alternative course outcome measures, which are more appropriate for measuring success in the non-formal lifelong learning context of MOOCs.

Participants in a mid-sized MOOC filled out pre-and post-questionnaires and data about their behaviour in the MOOC were collected from the course log files. This study used educational data mining and learning analytic techniques to understand how participants' demographics, their pre-course characteristics when entering the course, their actual behaviour in the course and their perceived course usability predict the two learner-centered outcome variables which describe the learners' level of satisfaction and the extent to which the MOOC enabled them to fulfill their intentions. Furthermore, despite the relatively high correlation between these two outcome variables ($r = .78$), our findings distinguished between two distinct pathways though which the participants achieved these outcomes. These pathways are presented in Figure 2, and elaborated below.

Learner satisfaction was directly and positively affected by four variables: two pre-course variables: the importance of the benefits of taking a MOOC, and online SRL-goal setting; one behavioural variable: number of video lectures accessed; and, one post course variable, perceived course usability. Furthermore, there were two indirect effects on learner satisfaction, through perceived course usability and through number of video lectures accessed. The first path begins with previous experience with MOOCs, which positively influenced the number of quizzes, and which, in turn, positively affects perceived course usability. The second path shows that age positively affected the number of video lectures accessed.

Intention-fulfillment was directly and positively affected by six variables. Gender directly affected the level of intention-fulfilment. The two pre-course variables were: the importance of the benefits of taking a MOOC, and online SRL-goal setting; two

Chapter 3

behavioural variables were the number of quizzes and the duration of time taking the MOOC; and, the post-course variable - perceived course usability. Female participants reported higher levels of intention-fulfillment. Furthermore, previous experience with MOOCs and the importance of the advantages of taking MOOCs, indirectly affected intention-fulfillment through the number of quizzes and the perceived course usability.

Our findings shed new light on the role of the demographic variables on course outcomes. Similarly to the findings of Garrido et al. (2016) who found that women are more likely than men to complete a MOOC or obtain certification, our findings demonstrate that gender had an effect on one of the learner centric outcome variables, by positively affecting the intention-fulfillment variable. Females had a higher level of intention-fulfillment than men did. Further research should explore whether this can be generalized beyond the specific context of this MOOC. In regards to age, our findings are similar to those of Morris et al. (2015) who found that older participants persist in their online studies more then young participants. Similarly, in our study, age was not a direct predictor of course outcomes, but rather predicted a behavioural variable that reflects progress in the course: the number of video lectures that the participants accessed during the course, which in turn predicted learner satisfaction.

As can be seen in Figure 2, the level of importance of the benefits of participating in the MOOC predicted both of the learner-centered outcome variables. It had a direct positive influence on both satisfaction and intention-fulfillment, as well as an indirect positive influence on the number of quizzes taken, which in turn influenced intention-fulfillment directly and satisfaction indirectly. The MOOC did not provide any credit beyond a certificate of completion, and we thus can see how lifelong learners who give the advantages provided by the MOOC a higher value, are likelier to invest more in the course, and to achieve positive outcomes. An applied implication of this finding is the importance of clearly delineating the MOOCs benefits and contributions in a way that allows participants to evaluate the relevance of the MOOC for their personal goals.

A strong and positive impact of goal setting on course outcomes was identified. As Zimmerman (2002) mentioned, the ability to set learning goals is an internal structure that is based on the learner abilities, and can be learned throughout one's life. Interestingly, our findings did not identify that those correlations were mediated by any of the behavioural variables.

Another thought-provoking finding of this study is the difference between the behavioural variables that influenced learner satisfaction and those that influenced intention-fulfillment. The number of video lectures accessed positively predicted learner satisfaction, while the level of intention-fulfillment was directly predicted by the number of weekly quizzes accessed, and by course duration. Accessing video lectures is a passive learning behaviour, while taking self-assessment quizzes, and to a lesser extent persisting in the course, are more active aspects of learner behaviour. A possible insight is that more active course components, such as self-assessment quizzes that provide participants with feedback on their achievements and understanding, assist learners who are focused not only on enjoying the course (i.e. learner satisfaction) but also on using the course to fulfill the personal intentions they had when they set out to study the MOOC (intention-fulfillment).

Finally, perceived course usability was a strong predictor of both course outcomes. This finding reflects the fact that a course with poor usability will delay the learner's progress, and decrease the personal benefits from participating in it (Eom et al., 2006). The only direct predictor of perceived course usability was the number of quizzes taken, which, as discussed in the previous paragraph, is also an important predictor of the key outcome variables.

Several limitations of this study can help drive future research. First, additional factors such as ICT skills and educational background should be examined as predictors of the course outcomes. In our study, those measures showed insufficient variability and could not be included in the analysis. Secondly, participants in our study were a unique sub-group of participants who had chosen to answer the pre- and the post-questionnaire, and not a random sample of the MOOC participants. This limitation is typical for MOOC studies that use self-reported questionnaires (Breslow et al., 2013; Kizilcec & Halawa, 2015). Nevertheless, as mentioned in the Method section, a comparison of the sample's demographic characteristic with the demographics of the course's population did not identify any significant differences. Since the MOOC that had been analyzed was in Hebrew, only Hebrew speaking participants had been able to participate in it. Those limitations reduce the external validity of the results. Future research should develop non-responsive methods to investigate the antecede of the two dependent variables - learner satisfaction and learner intention-fulfillment.



## Conclusions

In conclusion, although the correlation between learner satisfaction and intention-fulfillment is high, the behavioural predictors for the two constructs are different. While the level of learner satisfaction was predicted by the number of video lectures accessed, the learner intention-fulfillment was predicted by the number of quizzes and by the duration of participation in the course. We can see that although these two outcome variables are important, and although they show a high level of correlation, our findings distinguish between the antecedents of these outcomes. The level of satisfaction is determined mainly by the lectures and not by other learning aspects such as evaluation mechanisms, while intention-fulfillment is determined mainly by components that allow participants to self-assess their learning activities.

Finally, following the critiques of Reich (2015) who stated that research to date had little impact on educational practice and the critiques of Pardo, Han, and Ellis (2016) who pointed out that using educational data mining without a theoretical framework reduces the ability of translating the results into a meaningful pedagogical guidance, we would like to suggest that the educational impact of our results is that they propose a deeper, theory-supported, understanding of student perception of the courses and of their outcomes. This emphasis on the student's perspective is essential when discussing lifelong learning. Our findings demonstrate the importance of learners' ability to set goals in order to self-regulate their learning, and the importance of clearly stating the benefits of the MOOC, while providing participants with tools to evaluate their achievements during the course. Course designers and developers should not only develop excellent learning materials, but also assist MOOC participants to set their goals and to evaluate the potential benefits of the course.

***Funding***
The Open University's of Israel Research Authority Grant #101464 supported this work.

# Appendix

**Table 4.** Pearson correlations between all the research variables

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Pre-course** | | | | | | | | | | | | | | | | |
| 1. Age | – | | | | | | | | | | | | | | | |
| 2. Gender | –.23* | – | | | | | | | | | | | | | | |
| 3. Previous experience with MOOCs | .07 | .01 | – | | | | | | | | | | | | | |
| 4. Importance of MOOC's benefits | .09 | .08 | –.04 | – | | | | | | | | | | | | |
| 5. Importance of MOOC's disadvantages | .04 | –.04 | –.02 | .34*** | – | | | | | | | | | | | |
| 6. Online SRL – Environmental structuring | .07 | .07 | .03 | .17 | .27** | – | | | | | | | | | | |
| 7. Online SRL – Goal setting | .08 | .05 | .17 | .15 | .26** | .40*** | – | | | | | | | | | |
| **Behavioural measurements** | | | | | | | | | | | | | | | | |
| 8. Number of activities accessed | .03 | –.03 | .13 | .13 | .15 | .14 | .12 | – | | | | | | | | |
| 9. Receiving completion certificate | –.01 | .01 | .18* | .19* | .18* | .15 | .19* | .53*** | – | | | | | | | |
| 10. Number of videos accessed | .20* | .06 | .08 | .13 | –.15 | .15 | –.05 | .66*** | .37*** | – | | | | | | |
| 11. Number of forums accessed | –.09 | –.05 | .13 | .06 | .18* | .10 | .11 | .88*** | .43*** | .35*** | – | | | | | |
| 12. Number of quizzes accessed | .01 | –.03 | .16* | .20* | .30** | .15 | .16 | .70*** | .52*** | .47*** | .43*** | – | | | | |
| 13. Duration of participation | .09 | –.02 | .03 | –.04 | .06 | –.02 | –.15 | .31*** | .26*** | .24*** | .19* | .24*** | – | | | |
| **Post-course** | | | | | | | | | | | | | | | | |
| 14. Perceived course usability | .12 | .05 | .14 | .09 | .13 | .23* | .12 | .25*** | .31*** | .22*** | .18* | .27*** | .15* | – | | |
| 15. Intention- fulfilment | .15 | –.15 | .07 | .26** | .21* | .21* | .33*** | .36*** | .41*** | .30*** | .24*** | .37*** | .24* | .37*** | – | |
| 16. Learner satisfaction | .22* | .11 | .01 | .29** | .23* | .21* | .31** | .34*** | .38*** | .43*** | .16* | .37*** | .19* | .44*** | .78*** | – |

Gender - male coded as '1' and female coded as '2', Previous experience with MOOCs – yes coded as '1', Receiving completion certificate - received certificate coded as '1'
*$p < .05$, **$p < .01$, ***$p < .001$

# 4

# Chapter 4

## What are the barriers to learners' satisfaction in MOOCs and what predicts them? The role of age, intention, self-regulation, self-efficacy and motivation

This chapter is based on:

Rabin, E., Henderikx, M., Kalman, Y.M., & Kalz, M. (2020). The Influence of Self-Regulation, Self-Efficacy and Motivation as Predictors of Barriers to Satisfaction in MOOCs

*Australasian Journal of Educational Technology 36 (3), 119-131*

## Abstract

Massive open online course (MOOC) participants face diverse barriers that prevent them from feeling satisfied with participating in online courses. This study identified those barriers and their predictors. Using pre- and post-questionnaires, MOOC participants reported several characteristics and their barriers to satisfaction during the course. Exploratory factor analysis identified three kinds of barriers. The effects of participants′ age, gender, level of self-efficacy, motivation, self-regulated learning skills and the intention to complete the course were used as predictors of those barriers to satisfaction. The barrier lack of interestingness/relevance was predicted by the self-regulation indices of self-evaluation, study-strategy and help-seeking. The barrier lack of time/bad planning was predicted by the self-regulation indices of goal setting, time management and study strategy and by the age of the respondent. The barrier lack of knowledge/technical problem was predicted by the level of self-efficacy, extrinsic motivation and the self-regulation index of time management, as well as by the behavioural intention to complete the course. Furthermore, an index averaging the extent of the barriers was predicted by the self-regulation indices of goal setting and study strategy, the level of self-efficacy and the level of extrinsic motivation. Theoretical and practical implications are discussed in order to help MOOC participants, instructors and designers to enhance learner satisfaction.


**Implications for practice or policy:**

- Course developers and online instructors should be aware that participants in MOOCs face a variety of barriers that keep them from being satisfied with the learning process and learning outcomes.
- Practitioners should develop specific interventions for young participants and participants with fewer learning experience in MOOCs.
- MOOC designers and instructors should develop tailored systems and resources that help MOOC participants to self-regulate their learning process and to improve their self-efficacy.

***Keywords:*** MOOCs; Satisfaction; Self-efficacy; Self-regulated learning; Motivation; Intentions; Learners' barriers

## Introduction

Learners who enroll in massive open online courses (MOOCs) selectively engage in parts of the course content, and a small proportion eventually completes the course (Breslow et al., 2013; Ho et al., 2015; Reich & Ruipérez-Valiente, 2019). Several authors have recently suggested adapting the perspective on learning in MOOCs to the realm of non-formal education (Rabin, Kalman, & Kalz, 2019a) and proposed that the success of lifelong learning in MOOCs should be evaluated through learner-centred measures such as learner satisfaction from participating in the course (Rabin et al., 2019a; Reich, 2014). Online learning satisfaction has been emphasised by the American Distance Education Consortium as the most important element defining students' online learning success experience (Horvat, Dobrota, Krsmanovic, & Cudanov, 2015; Naveh, Tubin, & Pliskin, 2012). In addition, online learner satisfaction is one of the five elements for evaluating the quality of online learning identified by the Online Learning Consortium (Alqurashi, 2019). Learner satisfaction reflects students' perception of their learning experience (Alqurashi, 2019; Kuo, Walker, Schroder, & Belland, 2014; Littlejohn, Hood, Milligan, & Mustain, 2016) and is defined as the student's overall positive assessment of the learning experience (Keller, 1983). Research has found positive correlations between student satisfaction and post-secondary student success (Chang & Smith, 2008), the intention to use e-learning (Liaw & Huang, 2011; Roca, Chiu, & Martínez, 2006), retention in an online course (Lee & Choi, 2013; Levy, 2007; Park & Society, 2009), dropout rates of students and motivation and commitment to complete a degree online (Ali & Ahmad, 2011; Yukselturk & Yildirim, 2008).

The unstructured, self-paced nature of the MOOC learning environment creates unique types of barriers in the learning process, which can affect the level of satisfaction of the participants. This research focused on those barriers and on their predictors. The goal was to reveal the barriers to satisfaction that MOOC participants face and the antecedents to these barriers.

In the current study, we adapted the definition of barriers to learning from Henderikx, Kreijns, and Kalz (2018), which resulted in the following definition: barriers to learner satisfaction are issues that hinder or prevent learners from reaching their individual intentions and that harm their level of satisfaction. This type of barrier might differ from barriers in traditional education, due to the online and non-committed nature of MOOC learning and may be related to the MOOC itself, for example, bad course content, low quality of the course materials, or the absence of the instructor. Alternatively, the barriers may be extraneous to the MOOC or to the MOOC environment, for example, lack of time, insufficient academic background, family issues, workplace commitments, and insufficient technological background (Henderikx et al., 2018; Khalil & Ebner, 2014; Onah, Sinclair, & Boyatt, 2014). Predicting the barriers that learners can encounter enables course designers to tailor their design to specific needs, as well as prevent the implementation of unnecessary interventions. Moreover, being aware of the antecedents to the barriers can help learners to anticipate and prevent encountering those barriers or to overcome them with support from personalised, tailor-made assistance tools.

## Barrier antecedents

Participants' age can affect the occurrence of barriers. In a recent study, Henderikx, Kreijns, Muñoz, and Kalz (2019) analysed predictors of barriers to learning in MOOCs. They used the perspective of life stages theory (Stoffelsen & Diehl, 2007) and identified several external barriers that are most prominent in specific age ranges. Their analysis indicated that learners in their early adulthood (20–35 years) and mid-life (36–50 years) most often faced external barriers such as family and work issues. Learners in their mid-life stage (36–50 years) comprise the group most hindered by these issues.

Previous studies have found significant differences in learning, attitudes, motivation, and experiences of online learning, which were associated with the gender of the participant. Some recent studies (e.g., Garrido, Koepke, Anderson, and Mena, 2016) found that women are more likely than men to complete a MOOC or to obtain certification. On the other hand, other studies did not identify the influence of gender on achievement or on completion rates (Breslow et al., 2013; Cisel, 2014; Kizilcec, Piech, & Schneider, 2013; Morris, Hotchkiss, & Swinnerton, 2015). Using learner-centred indices for assessing success in MOOCs, Rabin et al. (2019a) found that gender correlated with the level of intention-fulfillment, but not the level of satisfaction. Female participants reported a higher level of intention fulfillment than males, but there were no differences between female and male participants in the level of learner satisfaction. Regarding barriers to online learning, Muilenburg and Berge (2007) found that men rated barriers of administrative issues, time and support more highly than women. Henderikx et al. (2019) found that female learners faced more barriers related to work-life balance than men, although they did not find significant differences across gender in the number of barriers for pursuing personal learning goals in MOOCs in general. In conclusion, we are not aware of evidence that shows that gender is associated with barriers to satisfaction in MOOCs.

According to the theory of planned behaviour, intentions are the most important predictors of behaviour (Ajzen, 1991). Several studies have shown that pre-course intentions of MOOC participants can predict the actual behaviour of the participants and their post-course evaluation (Koller, Ng, Do, & Chen, 2013; Reich, 2014; Wang & Baker, 2018). For example, Wang and Baker (2018) showed that learner intention to earn a certificate was positively associated with actually earning a certificate in a MOOC. However, in many cases, the intention is formed, but cannot be realised due to certain barriers which impede performance (Henderikx et al., 2018).

Another factor that influences learning outcomes and post-course evaluation is the ability of learners to self-regulate their learning (Kizilcec, Perez-Sanagustín, & Maldonado, 2017; Rabin et al., 2019a; Zalli, Nordin, & Hashim, 2019). Zimmerman (2000) defined self-regulation as self-generated thoughts, feelings, and actions that are planned and cyclically adapted towards the attainment of personal goals. Highly self-regulated learners are characterised by their higher ability to initiate metacognitive, cognitive, affective, motivational and behavioural processes in order to take actions to achieve their learning goals and persevere until they succeed (Zimmerman, 2002). In the absence of support and guidance in self-paced courses, the ability to regulate the learning process is a critical skill for achieving personal learning objectives. Online learners need to determine when, where and how to engage in course content and learning activities. Many learners struggle with self-regulation in online learning environments (Lajoie & Azevedo, 2006). In this study, we used the six dimensions of online self-regulation that were introduced by Barnard, Lan, To, Paton, and Lai (2009): goal setting, environment structuring, task strategies, time management, help-seeking and self-evaluation.

In addition, research has found that learners who indicate higher levels of self-regulated learning (SRL) also report higher levels of motivation and commitment to learning (Littlejohn et al., 2016; Margaryan, Bianco, & Littlejohn, 2015). Motivation is the process that initiates, guides and maintains goal-oriented behaviour (Ryan & Deci, 2000). According to these authors, motivation can be understood as a continuum between intrinsic motivation (defined as an active engagement with tasks because of self-desire to seek out new things and new challenges and to gain knowledge and fun) and extrinsic motivation (defined as the regulation of the activity as a function of expectations regarding reward and punishment).

Another factor which is known to affect learner behaviour and learning outcomes is the level of self-efficacy of the learner. Self-efficacy is defined as the belief that a task is achievable and that the environment in which one works enables one to achieve that task (Brennan, 2013). Students with high self-efficacy do not regard difficult tasks as obstacles to avoid, but rather as a challenge in order to develop their skills. Self-efficacy can enhance learning and performance and lead to higher satisfaction with the achieved results (Alqurashi, 2016). Online learning self-efficacy has been found to be a predictor of student satisfaction in online courses (Artino, 2007; Shen, Cho, Tsai, & Marra, 2013). According to the reasoned action approach (Fishbein & Ajzen, 2011) and self-determination theory (Ryan & Deci, 2000), participants' self-efficacy and level of motivation directly affect their learning behaviour.

As these factors are generally known to affect learning outcomes and learner behaviour, it can be expected that they will also have an impact on experiencing barriers to satisfaction. Although the studies mentioned have focused on factors influencing the barriers MOOC participants experience, or on their level of satisfaction, this study specifically focused on the role of age, gender, learner intention, level of self-efficacy for learning, level of motivation and level of self-regulation in the prediction of barriers to satisfaction in MOOCs. This led to the following two research questions:

- What types of barriers to satisfaction do MOOC participants encounter?
- How do age, gender, learner intention, level of self-efficacy for learning, level of motivation and level of self-regulation affect the barriers to satisfaction that MOOC participants encounter?

## Method

### *Participants*

Participants in this study were 542 English as a second language (ESL) MOOC learners. The participants answered a pre- and post-questionnaire at the beginning and at the end of their learning period. The pre-questionnaire was sent online via email to the MOOC participants when they logged into the course for the first time. After reading about the research goals and signing an informed consent form, participants responded to the pre-questionnaire. Four months after having filled the pre-questionnaire, a post-questionnaire was emailed to all the participants. The data collection ran between July 2016 and February 2018. The course was free of charge, and there were no prerequisites. The MOOC participants were able to join and leave the MOOC whenever it suited them. Most participants (71%) were female. The mean age of the sample was 32.4 years ( = 11.70; age range: 18–81 years).

### *Instruments and procedure*

*Dependent variable*

The MOOC participants were asked at the end of the course (post-questionnaire) to rate 12 barriers to satisfaction that they might have faced during the course. The list of barriers was adapted from Henderikx, Kreijns, and Kalz (2017) and Henderikx et al. (2018), and the items were rated on a Likert scale ranging from 1 (not at all) to 7 )fully). Table 1 present the items in the questionnaire.

To answer the first research question ("What types of barriers to satisfaction do MOOC participants encounter?"), we used an exploratory factor analysis with Varimax rotation due to the orthogonal nature of the factors (Gannon-Cook, 2010).

The exploratory factor analysis of the 12 barriers listed in the survey revealed three factors that accounted for 65.73% of the overall variance. The data set was examined for factor analysis adequacy, which was found to be satisfactory (Kaiser-Meyer-Olkin = .86, Bartlett's Test of Sphericity = 2775, $df = 66$, $p < .001$) (Field, 2005). The three factors that were identified are:

- lack of interestingness/relevance (barriers related to the quality of the learning resources, appeal of the course or of the certification options)
- lack of time/bad planning (barriers related to time planning, capacity for spending time)
- lack of knowledge/technical problems (barriers related to the high amount of information in the course, complexity of the course structure, technical problems or lack of prior knowledge).

These three factors corresponded reasonably well with the classification by Henderikx et al. (2018) – Factor 1 is similar to social context (Component 2), Factor 2 is similar to time, support and motivation (Component 4) and Factor 3 is similar to technical and online-learning related skills (Component 1).

Factor scores were calculated for each of the three factors identified, by averaging the items that make up the factor. Also, an overall index score was calculated by averaging all 12 items. The overall index score of barriers represents the extent to which participants perceived they faced barriers that prevented them from feeling satisfied with their learning experience in the MOOC.

Low to medium correlations were found between the three factors, indicating that although the barriers have some common features, they reflect different obstacles, and their different antecedents should be revealed (Pearson correlation ranging from .28 to .57, $p < .001$). Table 1 shows the mean, standard deviation and the factor loading of the 12 barriers.

***Independent variables***

To assess the characteristics of the participants, an online pre-course questionnaire was administered at the beginning of the course. The questionnaire consisted of the following parts:

- Demographics – Participants reported about their gender and age.

- Intention to complete the course activities – This was a single item asking about the behavioural intention of the participant to complete the course on 3-point nominal scale: *I plan to complete some portion of the course*, *I plan to complete all the course parts*, or *I do not know yet*.

- Self-efficacy for learning and performance – The Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich, 1991). The questionnaire included eight items measuring self-efficacy for learning and performance on a 7-point Likert scale (Cronbach's alpha = .93).

- Motivation – The Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich, 1991) includes four items for measuring internal motivation (Cronbach's alpha = .76) and four items for measuring external motivation (Cronbach's alpha = .70). The items were rated on a 7-point Likert scale.

- Online SRL skills – Six indices measuring online learning self-regulation were adapted from the online learning self-regulation questionnaire (OLSQ, Barnard et al., 2009). The indices were goal setting, environmental structuring, learning strategies, time management, seeking help and self-evaluation. All the items were measured on a 7-point Likert scale (Cronbach's alpha ranged from .71 to .85).

**Table 1.** Results of the exploratory factor analysis for the 12 barrier items

| Constructs and their component items | Mean | SD | Loadings | % of overall variance | Cronbach's alpha |
|---|---|---|---|---|---|
| *Barrier – general score* | 2.62 | 1.17 | | 65.73 | .86 |
| *Factor 1 – Lack of interestingness/relevance* | 2.58 | 1.35 | | 25.83 | .83 |
| The course content did not meet my expectations | 2.82 | 1.78 | .81 | | |
| The course was not interesting | 2.58 | 1.65 | .81 | | |
| The course was not relevant to me | 2.51 | 1.76 | .81 | | |
| The course did not provide any certificate | 2.78 | 2.05 | .58 | | |
| The quality of the course was low | 2.11 | 1.52 | .67 | | |
| *Factor 2 – Lack of time/bad planning* | 3.19 | 1.80 | | 20.22 | .87 |
| I had less time than I expected | 3.25 | 2.02 | .85 | | |
| Other stuff distracted my mind | 3.10 | 1.94 | .81 | | |
| I was not able to plan my time | 3.15 | 2.06 | .89 | | |
| *Factor 3 – Lack of knowledge/technical problem* | 2.18 | 1.24 | | 19.69 | .77 |
| The course was too complicated for me | 2.16 | 1.56 | .79 | | |
| The course overwhelmed me with lots of information | 2.19 | 1.58 | .82 | | |
| I had a technical problem with my computer, Internet connection or the website | 1.92 | 1.56 | .58 | | |
| I lack the skills and knowledge to pass the course | 2.47 | 1.77 | .58 | | |

# Results

To answer the second research question, as to which extent participants' characteristics affect the different barriers, four prediction models were created: one in order to identify the predictors of the overall barrier score index and three to identify each of the predictors of the three indices of barriers that were identified in the answer to the first research question.

In all four models, a stepwise linear regression model assisted in revealing the predictors of each factor. The independent variables were age, gender, level of self-efficacy, level of internal and external motivation, level of the six indices of SRL and level of behavioural intention (learner intention). The level of learner intention was coded into two dummy variables – the first dummy variable compared between those who intended to complete all parts of the course with those who intended to complete only some parts of the course or did not know how many parts they would complete of the course. The second dummy variable compared those who did not know how many parts they would complete of the course with those who intended to complete some or all parts of the course. Table 2 presents the regression coefficients and the summaries of the four prediction models. Only statistically significant results are reported.

The overall score of barriers was predicted negatively by the SRL indices goal setting and study strategy and the level of the participants' self-efficacy. The overall score of barriers was predicted positively by the SRL index of time management and the level of extrinsic motivation. This suggests that from a self-regulation perspective, the lower the goals that participants set for their learning process, the less they plan their learning strategy. In addition, the more participants plan how to manage their time, the more they will face barriers while taking the MOOC. Furthermore, the lower the participants' levels of self-efficacy for learning and the more they are driven to learn by external rewards, the more they will face barriers while taking the MOOC.

Factor 1, lack of interestingness/relevance, was predicted negatively by the SRL indices self-evaluation and study strategy and positively by the SRL index help-seeking. In other words, the less the participants are able to self-evaluate their learning process and to plan their strategy of learning and the more they believe that they will know how to search for help if needed, the more they will find the course not interesting or not relevant to them.

**Table 2.** Results of the stepwise linear regression for the association of age, intention to complete, SRL, self-efficacy and motivation with the participant barriers

| | **General score** | **Factor 1: Lack of interestingness/ relevance** | **Factor 2: Lack of time/bad planning** | **Factor 3: Lack of knowledge/ technical problem** |
|---|---|---|---|---|
| **Predictors** | **Beta** | **Beta** | **Beta** | **Beta** |
| Age | – | – | -.09* | – |
| Intention to complete – all parts vs. do not know yet/parts of the course | | | | -.10* |
| SRL – goal setting | -.16** | – | -.26*** | – |
| SRL – self-evaluation | – | -.20*** | – | – |
| SRL – help seeking | – | .24*** | – | – |
| SRL – study strategy | -.15* | -.13* | -.15** | – |
| SRL – time management | .16** | – | .18** | .12* |
| Self-efficacy | -.12* | – | – | -.23*** |
| Extrinsic motivation | .11* | – | – | .16** |
| **Model summary** | | | | |
| Adjusted $R^2$ | .08 | .04 | .11 | .09 |
| $F$ | 9.96*** | 6.38*** | 12.90*** | 11.08*** |
| $Df$ | 5,404 | 3,404 | 4,400 | 4,401 |

*Note.* Only significant predictors are presented. The predictors gender, SRL-environmental structuring, as well as the dummy variable of intention to complete – do not know yet vs. all and some parts – were not included in the table since they did not significantly predict any dependent variable.
*$p$ <.05; **$p$ <.01; ***$p$ <.001

Factor 2, lack of time/bad planning, was predicted negatively by the SRL index goal setting, study strategy and the age of the respondent and positively by the SRL index time management. The less participants are able to set their learning goals and learning strategies, the more they are able to manage their own learning time, and the younger the participants are, the more they experience lack of time or bad planning.

Factor 3, lack of knowledge/technical problem, was predicted negatively by the level of the participant`s self-efficacy and positively by the level of their extrinsic motivation towards participation and by the SRL index time management, as well as the initial behavioural intention. The more participants reported that they had a low level of self-efficacy and a higher level of external motivation and the more they reported being able to manage their own learning time, the more they faced lack of knowledge and technical problems. Those who intended to complete all the course activities faced less lack of knowledge and technical problems than the two other groups of participants: Those who intended to complete only some parts of the course activities and those who did not know how many parts of the course they would like to complete. In this model, no differences were found between participants who intended to complete only some parts of the course activities and participants who did not know how many parts of the course they would like to complete.

It is interesting to note that the gender of the participants and their level of SRL-environmental structuring did not predict any kind of barriers. Male and female participants face the same barriers while participating in a MOOC. The SRL dimension environment structuring, as opposed to the other SRL factors, did not predict any of the barriers for satisfaction.

## Discussion

The aim of this research was to identify barriers to satisfaction experienced by MOOC participants and the predictors of those barriers. Participants in an ESL MOOC were asked to report their demographic and psychological characteristics and the barriers to their satisfaction with the MOOC.

Three kinds of barriers to satisfaction were identified. The results suggest that, although there are correlations among the barriers that the participants face, the predictors of the indices of the barriers are diverse. This indicates that there are different antecedents to each factor. The level of SRL, self-efficacy, extrinsic motivation, the initial behavioural intention and the age of the participant predicted different indices of the barriers. The results of the regression models suggest that to reduce barriers as a whole, course designers and facilitators should help MOOC participants to self-regulate their learning process and help them to promote their feeling of self-efficacy. Furthermore, they should be aware of the participants' initial behavioural intentions and pay closer attention to young participants.

The overall barrier score index is predicted negatively by the self-regulation indices of goal setting and study strategy and positively by the ability to manage study time. Furthermore, the overall score index of barriers is negatively predicted by the level of self-efficacy of the participants and positively by extrinsic motivation. Goal setting refers to the specification of educational goals or sub-goals in order to exert the effort required to achieve those goals (Schunk, 2005; Zimmerman, 2000). Study strategy refers to activities to improve persistence and effort-regulation in the face of academic challenges (Richardson, Abraham & Bond, 2012). Participants who score low on goal setting and study strategies will face more barriers to satisfaction. Goal setting and strategy planning are related to the ability to plan the learning process and refer to one's decision-making on how to accomplish a learning task (Kitsantas & Zimmerman, 2002). The ability to plan learning is related to feeling satisfied with the outcomes of web-based learning (Whipp & Chiarelli, 2004). On the other hand, participants who allocate, schedule and allot time for learning will face barriers to satisfaction more intensely. This result is counter-intuitive, since we would expect that those who are able to manage their time will also know how to manage the barriers they face. The results, however, suggest the opposite. In the context of consumer research, Townsend and Liu (2012) discussed several conditions and factors for which planning can have a negative effect on self-regulation. If goals are far away or competing implementation intentions are available, planning can actually hinder goal achievement. The authors introduced five studies in which they found that the actual "position" concerning a long-term goal plus an interaction effect between the level of concreteness and a large goal-distance can lead to distress and finally negative effects of planning. More research in relation to the ideal amount of planning needs to be conducted to understand the mechanism that is at the root of this phenomenon. Finally, those participants who feel that they have little ability to handle learning in MOOCs and those who came to study to achieve external goals will face more barriers to satisfaction. These results are in line with the nature of the MOOC that was under investigation, which was developed for self-paced learning and did not reward the participants with external incentives. Therefore, the needs of participants who score low on self-efficacy in handling the learning process and the needs of participants who are motivated by external motivation were not fulfilled.

The three predictors of the first factor that dealt with barriers regarding interest and relevance of the course materials were indices of self-regulation. The predictors help-seeking, self-evaluation and study strategy suggest that we should assist learners to trust themselves, to rely less on help from peers and instructors (Richardson et al., 2012) and to improve their ability to evaluate their learning process by setting quality standards for progress (Boud, 2013). Furthermore, improving study strategies by using activities that promote persistence and effort-regulation in the face of academic challenge (Richardson et al., 2012) is recommended. It seems that the positive correlation between help seeking and the barrier lack of interestingness or relevance is counter-intuitive. The ability to seek help from instructors and friends when experiencing difficulties with academic work is a positive capacity, but in this specific MOOC, which did not provide any external or social support such as learning groups or discussion forums, the need to rely on the help of instructors or peers was experienced as a hindrance. These findings complement those of Kizilcec et al. (2017), who showed that help-seeking is associated with lower goal attainment.

To help learners to overcome the second factor, which dealt with barriers regarding lack of time or bad planning, learners should be encouraged to set educational goals or sub-goals at the beginning of a MOOC (Schunk, 2005; Zimmerman, 2000) and to better plan their study strategy. Yet, it is important to note that learners who manage their time too strictly might also face the barriers lack of time or bad planning. This is an interesting finding since one would expect that the ability to plan the schedule ahead of the course will help learners to overcome this barrier, but it seems that when measuring the subjective feeling of lack of time, participants who manage their study time more closely will also experience this barrier more intensely than those who do not plan as extensively. This again relates to the above-mentioned appropriate level of planning.

The findings of our study also show that the younger the participant, the more explicit this barrier. This finding is complementary to the findings of Henderikx et al. (2019), who focused on barriers to reaching personal learning goals and not on barriers to satisfaction. The authors argued that specific barriers predominantly appear at specific phases of life. Moreover, it can be assumed that the barrier of time management occurs with young learners due to their lack of experience in studying online and their lack of experience in self-regulating their learning. Those results suggest that course designers and instructors should pay more attention to young learners, who are more likely to face those barriers.

The third factor, lack of knowledge/technical problem, is affected by the SRL dimension of time management, the level of the external motivation, the level of self-efficacy and by the initial behavioural intentions of the participants. Participants who scored high on the SRL dimension of time management, high on the level of external motivation and low on self-efficacy were more likely to face those barriers. Participants who intended to complete only some parts of the course activities or did not know how many parts of the course they intended to complete were more likely to face those barriers than those who intended to complete all the course activities.

Time management is the ability to allocate, schedule and distribute time for learning (Yen, Tu, Sujo-Montes, & Sealander, 2016). The inability of participants to manage their learning process relates to the lack of ability to handle barriers related to knowledge and technical issues. The reason for that is possibly that participants who do not know how to manage their study time do not allocte, schedule and distribute enough time to deal with technical issues and with knowledge gaps. Regarding the finding on the effect of self-efficacy, similar results have been found by Bozdoğan and Özen (2014), who showed that the feeling of competency to handle the technical aspects of learning in an online environment are critical for a successful use of information and communications technology for online courses. Bandura (1995) defined self-efficacy as "beliefs in one's capabilities to organise and execute the courses of action required for managing prospective situations." (p. 2). In this sense, lower levels of self-efficacy might hinder the ability of participants to take action on several conditions for a successful learning experience, like the acquisition of knowledge required to follow the course or the ability to handle technical issues. The positive correlation between the level of external motivation and this barrier for satisfaction suggests that learners who are motivated by external rewards will be less tolerant to face a lack of knowledge and a lack of technical abilities. The initial behavioural intentions of the participants also played a role in predicting the barrier for satisfaction. The pre-course intention to complete the full course has been identified in previous research as a predictor of the fulfilling of the course obligations and the earning of a certificate (Ho et al., 2015). Participants in open learning environments can set their own learning goals by defining their individual intention towards participating in the course (Rabin, Kalman, & Kalz, 2019b). In line with that, an interesting finding is that the intention to complete a large portion of the course did not play a role in predicting the other barriers: neither lack of interestingness/relevance and lack of time/bad planning, nor the overall barrier score. Our findings indicate that behavioural intention does not predict the other barriers that learners face while trying to achieve satisfaction from participating in the course. Future research could explore how the intentions of the participants and the barriers they face affect their learning behaviour and their learning outcomes.

The gender of the participant did not play any role in determining the barriers to satisfaction. This finding is interesting since research has shown that there are differences between males and females in learning, attitudes, motivation and experiences of online learning. Muilenburg and Berge (2007) found that men are more likely to rate administrative issues, time and support as barriers to online learning compared to women. However, the results in the current study are reminiscent of the findings of Rabin et al. (2019a), who showed that there are no differences between females and males in the level of learner satisfaction while studying in a MOOC.



## Limitations and conclusions

One limitation of this study is that students were asked to self-report their psychological and educational traits as well as their experience of barriers to course satisfaction. From the self-reported responses, it is hard to evaluate the actual level of barriers that the participants faced. Triangulation with additional sources such as interviews or behavioural indices could be utilised for future research. For example, a mixed-method research set-up would be appropriate to further explore and gain a deeper understanding of quantitative self-report results (Morse, 2016). Secondly, this research focused on a fully online course that did not provide any kind of social support, such as learning groups and discussion forums. It might be interesting to see if research can replicate these findings in online courses that do provide some kind of support from peers or instructors or alternatively in courses in which it is easier for the learner to get help, such as in hybrid or blended courses. Lastly, there might be other factors that affect the level of participants' satisfaction that have not been taken into account in this research. For example, Bornschlegl and Cashman (2019) showed that interaction with other students in an online course was correlated negatively with their level of satisfaction. At the same time, the level of entertainment and the extent to which they perceived the course to contribute to their education was positively correlated with participant satisfaction. Future research could investigate the effect of indices related to student interaction, perceived entertainment and perceived contribution to education and learning on barriers to satisfaction faced by the online course participants.

In conclusion, participants in MOOCs face a variety of barriers that keep them from being satisfied with the learning process and their learning outcomes. Since the open education and MOOC context offers a different set of learning opportunities and social context compared to formal education (Rabin et al., 2019b), the role of satisfaction in the chain of intention-formation, facing barriers, coping with barriers and last but not least the realisation of initial intentions should not be underestimated. Although earlier studies have focused on general factors influencing the appearance of barriers (Henderikx, et al., 2019), this study has specifically focused on the role of learner self-regulation skills, intention, motivation, self-efficacy, age and gender with regard to their influence on the experience of barriers to satisfaction. The ability to identify those barriers to satisfaction and to recognise different groups of participants who are most likely to face different barriers can help to develop human and automated support mechanisms tailored to the needs of the learners. In that way, course designers and instructors will be able to help participants avoid and cope more effectively with those barriers to satisfaction, and subsequently help them realise their individual learning intentions.

More specifically, findings suggest that MOOCs need to make learners aware about which support infrastructures are available within a course to avoid misconceptions and manage expectations. In addition, since for most MOOCs, an increase in support is not likely to be a solution due to limited resources and the need to scale up the number of participants, the implementation of peer-support scenarios should be explored as an alternative solution to serve the needs of learners who need support (Van Rosmalen et al., 2008). Findings also suggest a need for specific interventions for young participants and participants with little learning experience in open learning environments like the MOOC discussed in this study. In addition, it would be helpful if MOOC designers and instructors develop systems and resources that help MOOC participants to self-regulate their learning process and to improve their self-efficacy. We should, though, be cautious when we integrate support systems into these learning environments (Davis, Chen, Hauff, & Houben, 2016) since over-planning can also negatively impact satisfaction depending on individual characteristics and goal distances. This calls for a more tailored approach to planning support in MOOCs.

An important implication for theory development is the confirmation in our findings that satisfaction plays a different role in a non-formal learning context than in a formal learning context. While satisfaction has been identified as a bad predictor for academic achievement in a recent large-scale study (Rienties & Toenetel, 2016), in open education contexts like MOOCs satisfaction is likely to play a more important role for self-regulation and the chain of goal setting, learning behaviour and reflection. Future research should explore these differences between formal and the non-formal educational contexts.

# 5

# Chapter 5

## User behavior pattern detection in unstructured processes – a learning management system case study

This chapter is based on:

Codish, D., Rabin, E., & Ravid, G. (2019). User Behavior Pattern Detection in Unstructured Processes - a Learning Management System Case Study

*Interactive Learning Environments*
*https://doi.org/10.1080/10494820.2019.1610456*

## ABSTRACT

Process mining methodologies are designed to uncover underlying business processes, deviations from them, and in general, usage patterns. One of the key limitations of these methodologies is that they struggle in cases in which there is no structured process, or when a process can be performed in many ways. Learning Management Systems are a classic case of unstructured processes since each learner follows a different learning process. In this paper, we address this limitation by proposing and validating the user behavior pattern detection (UBPD) methodology which is based on detecting very short user activities and clustering them based on shared variance to construct a more meaningful behavior. We develop and validate this methodology by using two datasets of unstructured processes from different implementations of a learning management system. The first dataset uses a gamified course where users have the freedom to choose how to use the system, and the second dataset uses data from a massive online open course, where again, system usage is based on personal learning preferences. The key contribution of the methodology is its ability to discover user-specific usage patterns and cluster users based on them, even in noisy systems with no clear process. It provides great value to course designers and teachers trying to understand how learner interact with their system and sets the foundation for additional research in this class of systems.


***Keywords:*** Learning analytics; learning management systems; process mining; spaghetti processes; pattern detection; gamification

## Introduction

Process mining is a method used to discover underlying business processes, or deviations from such processes, through the analysis of system log files, which represent the actual behavior of users within a system (Van den Beemt, Buijs, & Van der Aalst, 2018; Van der Aalst et al., 2012; Van der Aalst & Weijters, 2004). While process mining has been successful in discovering well-structured processes, it has been less successful in non-structured processes, resulting in spaghetti-like process maps which are hard to interpret and use (Chinces & Salomie, 2015; Li, Bose, & Van der Aalst, 2010). Well-structured processes are processes that are followed by all users, while less structured processes allow users to perform them in different ways. These deviations from the process may, or may not, be acceptable from a designer's point of view.

Structured processes are common and desired in business environments where employees are expected to follow a certain flow of actions to achieve an objective such as the completion of a purchase order, reporting their monthly working hours, or filling a reimbursement form. Despite each of the examples above having deviations in their processes such as in the case of a purchase order that does not match company guidelines or a reimbursement request for a large sum, they can still be considered structured, as even these deviations from the processes are well-defined and structured. Unstructured processes, on the other hand, have no single process to follow, and users can follow any course of action at any point in time. Two such cases are the focus of this article. First, cases where there is no clear process at all, such as in learning management systems (LMS), news consumption sites or a social networking application where there is no point in searching for an overall process since it does not exist. Second, processes which may have existed, but due to a change in the system, such as adding gamification, the process is no longer structured. Gamification is the use of game design elements in a non-gaming environment (Deterding, Dixon, Khaled, & Nacke, 2011) with the intent of increasing user engagement (Kankanhalli, Taher, Cavusoglu, & Kim, 2012; Werbach, 2014), hedonic motivation (Lowry, Gaskin, Twyman, Hammer, & Roberts, 2013; Van der Heijden, 2004), or achieving other business goals (Hamari & Koivisto, 2015). The gamification of information systems involves adding different game elements to existing systems which, as a result, changes the way users interact with them. For example, granting points or badges for specific actions is expected to incentivize these actions, and including user profiles in an application is expected to increase social interaction. Gamification typically involves adding several game elements to a system, and given the voluntary nature of gamification, this means that different users would interact with them differently. As a result, even streamlined processes become less structured, making process mining less beneficial. Gamification of information systems is becoming common within organizations and thus, should receive special interest from system developers and researchers.

Although most process mining methods are not suitable for less-structured processes such as in the case of gamified systems, some methods can still address these limitations. For example, sequence mining (Srikant & Agrawal, 1996), episode mining (Mannila, Toivonen, & Verkamo, 1997), and the apriori and generalized sequential pattern (GSP) methods (Agrawal & Srikant, 1994; Srikant & Agrawal, 1996) are designed to detect recurring patterns, or sub-processes, within an overall noisy process. The sequence hierarchy discovery algorithm (Greco, Guzzo, & Pontieri, 2005) attempts to detect sub-processes and reconstruct them into the full process, assuming it exists. However, these algorithms assume that a process exists and that all users follow it similarly, which is not always true. Our research question is thus: Within a non-structured process or system, can we automatically identify recurring user-level behavior patterns and perform user clustering based on these patterns?

In this paper, we develop and validate the user behavior pattern detection (UBPD) algorithm employing system logs to automatically detects user behavior patterns and cluster users based on these patterns. We define user behavior patterns as usage patterns that certain users perform more, or less than, others. Both case studies used in this paper are based on educational settings, thus from an educational point of view, behavior patterns can easily be interpreted as learner behavior patterns. Our key contributions in this paper are the development of an automated end-to-end process to detect structured behavior patterns within an otherwise non-structured environment.

An additional benefit is the algorithm's ability to detect these sub-processes at the user level, while most existing methods search for sub-processes at the system level. For instance, if half of the users perform task A and then task B and half perform task B and then task A, a methodology seeking for patterns at the system level, would not detect this as a pattern, while UBPD would. The discovered user behavior patterns can be used for additional user clustering or a deeper understanding by system designers as to how their system is being used. Its main applicability is in cases in which there is no structured process, or no process at all, such as LMSs where learners typically log in to perform a specific task and then log out and news websites where users consume news in no particular order. With the advent of digital footprints analysis (Golder & Macy, 2014; Lambiotte & Kosinski, 2014; Williams & Pennington, 2018), where digital records of a person from many sources are combined to create a user profile, such an approach can be useful since data would be unstructured by nature and difficult to analyze.

The algorithm presented is based on a few stages. The first is a data preparation stage in which data are collected from various log files and organized. A sequence mining approach is used to detect the most frequent sequences of actions and organizes them at the user level. The clustering of these sequences per user is done through exploratory factor analysis (EFA), which results in factors representing user behavior patterns. Last, causal nets are used to construct a representation of these factors graphically. Two data sets from different LMSs were used to test the algorithm. The first dataset comes from a traditional, but gamified, academic course, meaning it had no structured processes. A second case study was based on data from a standard massive open online course (MOOC).

The emerging patterns from both cases studies, indicating how different users approached these courses, is presented. Lewis Carroll writes in Alice in wonderworld: "If you do not know where you are going, any road will get you there", therefore we were required to answer the question, how do we know if the results are accurate or random. To validate the results, we generated random user behavior patterns and inserted simulated data representing them into the dataset of the first cases study. The algorithm was executed again – confirming that previous patterns as well as the simulated patterns emerged.

This paper is structured as follows. First, a background on pattern discovery and process mining is provided. A brief background on gamification and the way it can unstructured processes is given, and the limitations of existing process mining methods are outlined. Next, the UBPD methodology is proposed, and relevant considerations are discussed. Two real-life case studies and simulation data are used to demonstrate how the methodology works and how results are achieved. Finally, a discussion of the results, applicability, and limitations of the methodology, as well as future research directions are provided.

## Background

### *Pattern discovery*

Understanding user behavior in online systems helps site developers and designers understand how their system is being used, what works well, and what needs to be improved (Srivastava, Cooley, Deshpande, & Tan, 2000). System log files can partially answer these questions as they provide statistics such as the most accessed page, the frequency of visits per user, and the duration of time on a page. Error log files complement this data by providing information such as broken links, unauthorized access attempts, general errors on the website, and more, depending on the richness of these logs.

Understanding the bigger picture hidden within the log files requires going beyond basic statistics. In systems where users are expected to follow a specific process (i.e. completing an online order or purchase request), analysts might want to know if users are indeed following this process, are there deviations from the process and which users are deviating from it. In systems where there is no process to follow (i.e. news web sites or knowledge management systems), analysts might be interested in questions such as what, if any, sub-processes exist, are all users behaving in the same unstructured manner or are there different classes of users that emerge. As information systems are often a mixture of structured and unstructured processes, in most cases, all the above questions are relevant.

Several advanced methods exist to address these more complex questions. Clustering methods (Ferreira, Zacarias, Malheiros, & Ferreira, 2007; Luengo & Sepúlveda, 2012) are used to group user actions with similar characteristics, classification methods (Pennacchiotti & Popescu, 2011) are used to classify user actions into a given set of classes, and association rules methods (Agrawal & Srikant, 1994; Lau, Ho, Chu, Ho, & Lee, 2009) are used to detect user actions that frequently appear together. Beyond user behaviors, it is sometimes interesting to detect hidden processes or parts of processes. Methods such as process mining (Van der Aalst, 2011b; Van der Aalst et al., 2012; Van der Aalst & Günth, 2007) and sequence analysis (Van Helden, 2003) are used in such cases. Most of these methods use system log files as input and assume a sequential set of activities are recorded in them, indicating there is a process that led to the execution of these sequences of actions, hence, the discovered process.

Sequence mining (Srikant & Agrawal, 1996) and Episode mining (Mannila et al., 1997) examine sequences of events and search for recurring usage patterns based on the most frequent sequences of events. They do not necessarily require that an end-to-end process exists, and rather focus on subsets of processes. The Apriori and generalized sequential pattern (GSP) methods (Agrawal & Srikant, 1994; Srikant & Agrawal, 1996) are commonly used for this task by scanning the entire set of sequences and searching for sequences that meet a minimum frequency threshold but may be time consuming when datasets are large (Han et al., 2001). Episode mining (Leemans & Van der Aalst, 2014; Mannila et al., 1997) uses the notion of a sliding window based on time or number of events and searches for frequent items within this window. Sequence hierarchy discovery is an algorithm that looks at hierarchies of sub-processes (Greco et al., 2005) and tries to combine them into a full process, assuming it exists. Some of the more recent algorithms use stochastic modeling and a Markov chains approach (Balakrishnan & Coetzee, 2013; Faucon, Kidzinski, & Dillenbourg, 2016; Geigle & Zhai, 2017) to address the fact that not all users interact with the system in the same way and describe how users navigate within the system.

Web server log files are good candidates for sequence mining (Mobasher, Cooley, & Srivastava, 2000; Patel & Parmar, 2014; Sisodia & Verma, 2012; Spiliopoulou, 2000; Srivastava et al., 2000) because pages are accessed sequentially, and there are several links a user can select at any given moment. Studies have shown that sequence mining provides good results and is already in use in generating personalized websites (Ferreira et al., 2007). Sequence mining is also commonly used in genome studies to examine DNA sequences (Kaneko et al., 1996).

The aforementioned methods work well for systems with an underlying business process such as in the case of purchasing (Ingvaldsen & Gulla, 2008), audit processes (Jans, Van der Werf, Lybaert, & Vanhoof, 2011), supply chain management (Lau, Ho, Zhao, & Chung, 2009; Trkman, McCormack, De Oliveira, & Ladeira, 2010), and other business processes that have clear start and end points. However, not all systems have an underlying business process. News websites allow users to consume news differently, in Learning Management Systems (LMS) the processes may be extremely short, such as accessing a system to download a presentation, view a video, or submit an assignment, in MOOCs participants can interact with the learning materials in any order and time that they choose, and in social network sites, users can browse content and jump from topic to topic in what may seem like a chaotic behavior.

While process mining methods have shown great success in discovering structured processes, they are less successful with non-structured processes where processes do not have a clear path and any step can follow any step (Rebuge & Ferreira, 2012; Van der Aalst, 2011b). Structured processes are processes in which all activities are repeatable and have a well-defined input and output, while unstructured processes are processes where activities have no pre- or post-activity and are determined based on experience, intuition, trail-and-error, and rules-of-thumb (Van der Aalst, 2011a). Discovering specific usage patterns in non-streamlined and non-structured processes is a promising research direction (Celino & Dell'Aglio, 2015). Even in cases in which there is a significant underlying process, it may have so many deviations, that the ratio between the deviations and main process is too large, and the existing algorithms would struggle to fully understand what the intended process is and what are the deviations. In such cases, sequence mining methods are typically used to identify sub-processes that may or may not add up to a full process. When there is no clear process, the focus is switched from examining how a system is being used, to how different users are using it, also referred to as user behavior patterns. User behavior patterns are sequences of actions that are performed by a user sequentially (Tseng & Lin, 2006) or almost sequentially. There is no definition to the amount of actions that constitute a pattern, and in some cases, even two activities qualify as a pattern (Kang, Liu, & Qu, 2017).

For the detection of user behavior patterns to be useful, the process of detecting and analyzing behavior patterns must be fully automated, which is missing in current research. In some studies (Davis, Chen, Hauff, & Houben, 2016; Hou, 2015; Huang, Chen, & Lin, 2019) the analysis process is indeed automated using sequence and clustering methods, but the data collected and the pattern detection processes are based on manual observations and interpretations, or on a set of predefined expected behaviors. The limitations of these methods are both in the manual classification step and in their need for a predefined set of behavior classes. Another issue with many of the existing processes is that they work at the system level and not at the user level. They seek to understand the overall process or sub-processes performed by users, ignoring the inherent differences between users. The above leads to the following research question: Within a non-structured process or system, is it possible to automatically identify recurring user-level behavior patterns, and perform user clustering based on these patterns?

***The case of gamification – when a process becomes unstructured***

Gamified systems are good examples of loosely-structured processes. Gamification is the use of game design elements in a non-gaming environment (Deterding et al., 2011) with the intent of increasing user engagement (Kankanhalli et al., 2012; Werbach, 2014), hedonic motivation (Lowry et al., 2013; Van der Heijden, 2004), or achieving other business goals (Hamari & Koivisto, 2015). In recent years, gamification is commonly included into LMS (Buckley & Doyle, 2016) as a means to increase motivation. The inclusion of game elements, into a utilitarian environment, such as LMS, is likely to change the way users interact with the system due to the additional options and affordances provided, reducing the structure of existing business processes. Due to the unstructured nature of gamified systems, using process or sequence mining to discover an underlying process would be challenging and can become even more challenging if the system was initially unstructured.

The most common approach to studying the effects of game elements on users is to examine the isolated effects of specific game elements and assess their contribution to the overall objectives of the gamification implementation. The most common game elements studies are points (Mekler, Brühlmann, Opwis, & Tuch, 2013), badges (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2013; Antin & Churchill, 2011; Hakulinen, Auvinen, & Korhonen, 2013), leaderboards (Butler, 2013; Costa, Wehbe, Robb, & Nacke, 2013; Landers & Landers, 2015; Mekler et al., 2013), and levels. The majority of studies focus on effects of a single game element on gamification success (Hamari & Koivisto, 2013; Li, Grossman, & Fitzmaurice, 2012), providing insights at the game element level. In reality, gamified systems do not include just a single game element, and the ability to understand user behavior patterns provides the ability to study the interaction between game elements and their influence on gamification success, which is a line of research only a few scholars pursue (Codish & Ravid, 2014a, 2014b).

The goal in gamification is to trigger user behaviors that support business objectives. Designers may intentionally try to trigger a specific behavior through gamification (e.g. create a cooperative environment or a sharing culture), however, they might also add game elements without fully understanding of how users would relate to them. In any case, even with proper design, it is hard to predict precisely how users would interact with game elements. Due to the unexpected behaviors that may arise (Callan, Bauer, & Landers, 2015; Werbach, 2014), measuring the outcomes of gamification is an important activity that should be performed throughout the implementation phase.

One option for measuring success of gamified systems is to measure the desired business objectives before and after gamification implementation. While such an approach has its benefits, it lacks the ability to provide insight into how individual users are influenced. This latter point is important since not all users would be influenced in the same way, and while some users may be extremely engaged, others may be negatively affected. Understanding how users interact with a system, be it an expected behavior or not, requires systematic detection of these user behavior patterns, which, as mentioned, is not trivial. To date, few authors (Ašeriškis & Damaševičius, 2014; Codish & Ravid, 2015; Sisodia & Verma, 2012) have proposed going beyond the analysis of trivial user behavior patterns in gamified environments and seek emerging patterns through log analysis. However, these studies do not provide an automated method to perform these tasks and focus on the theoretical conceptual steps that should be taken.

Systems and gamification implementations differ from each other, thus, any methodology for detecting user behavior patterns must be completely automated and system independent. We propose the User Behavior Pattern Detection (UBPD) methodology, which is based on sequence analysis methods, as an automated process for detecting differences in behavior patterns between users. We consider a user behavior pattern as a pattern that is common to several users but not to all users, which is the essential difference between a user behavior pattern and a system level usage pattern. To demonstrate and validate the methodology, we use a learning management system, which has no streamlined processes, and include gamification to make it even less structured.

# Methodology

### *Terminology*

Extracting user behavior patterns from a system requires examining sets of common usage patterns and looking for user-specific repeating patterns. Unlike methods such as episode mining (Mannila et al., 1997) and sequence analysis (Van Helden, 2003), where the objective is to find frequently recurring patterns, in this case the objective is to find patterns that are frequent for only some of the users. Having such patterns is an essential phase in the ability to cluster users based on their behavior patterns.

Using process mining terminology (Van der Aalst et al., 2012), the following terms are defined as summarized in Table 1. An event is an archetype action that can be recorded by the system. Events are determined by the system's capability to generate them. Examples of an event are opening a file, visiting a page, or viewing a video. An activity is a single event performed by a user and recorded by the system. If a user performs an event many times, each occurrence of performing the event will be recorded as an activity. Not all events need to be analyzed, such as system-generated events, time-based events, or error messages. These can be considered irrelevant to user behavior analysis, and at a certain point during the cleanup phase, they should be removed. However, it is important to note that in some cases, these supposedly non-relevant events may trigger events by the user and should perhaps not be ignored.

Systems often record many types of events that practically represent the same action. For example, suppose there are different events called opening link A, opening link B, and opening link C. If these events represent opening a link with no need to distinguish between them, we should represent the three events as a single action called "open link". This means that an action is a superset of events that, for analysis purposes, represent similar events.

A session includes all activities performed by the user between the timeframe of logging into the system and logging out of the system. Thus, there is a need to identify these sessions. In cases where a user logs in and logs out, this is straightforward, but in many cases, such as when systems remember user authentication, the login is automated and is not recorded as an event. Logging out of a system depends on users' habits and awareness of privacy issues. In some cases, users close the system without logging out, and in cases in which a personal device is used, a logout may never happen. To overcome this limitation, it is common to use a threshold of 30 minutes of inactivity to indicate the start of a new session (Clark, Ting, Kimble, Wright, & Kudenko, 2006).

**Table 1.** Behavior patterns methodology terminology

| Term | Definition |
|---|---|
| Event | An archetype action that can be recorded by the system |
| Activity | A single event performed by a user and recorded by the system |
| Action | A superset of events, that for analysis purposes represent similar events |
| Motif | Recurring sequences of actions that appear in a network more frequently than expected in a random network |
| Session | All activities performed by the user between the timeframe of logging into the system and logging out of the system. If a user is not active for more than 30 minutes, a log out activity is automatically defined |

Searching for user behavior patterns requires the identification of cases in which a specific sequence of actions reoccurs more frequently for some users than it does for others. Most process mining methods do not focus on user behavior differences, and thus seek frequently performed sequences of actions regardless of who performed them. The focus on user-specific behavior patterns is the key difference between UBPD and existing process and sequence mining methods. Searching for frequent sub-sequences of actions within a given sequence is the focus of several algorithms, such as the Apriori (Agrawal & Srikant, 1994), the GSP algorithm (Srikant & Agrawal, 1996) that expands the Apriori algorithm, and episodes finding (Mannila et al., 1997), in which episodes are defined as "a collections of events that occur relatively close to each other in a given partial order" (Mannila et al., 1997, p. 259). These algorithms are good at finding overall frequent sequences of actions. They do not, however, directly address our need for detecting user-specific behavior patterns.

Borrowing a term from genetics research, where sequence mining is commonly used, a motif is defined as a "recurring pattern that appears in a network more frequently than expected in a random network" (Alon, 2007; Milo et al., 2002). Motif research originally focuses on detecting how proteins regulate genes, but it is used in different domains as well, gaming among them, where they are used to understand how specific actions regulate behavior (Ghoneim, Abbass, & Barlow, 2008). In terms of behavior patterns, motifs are the recurring sequences that appear in user sessions. Figure 1 shows how all the terms defined above relate to each other.

Algorithms dealing with finding frequent subsets of actions, i.e. motifs, differ in how they achieve this. In our case, we seek to find user specific usage patterns we can relate to a user behavior. The most predominant question that needs to be addressed is what qualifies as a frequent motif. Algorithms address this by defining threshold values determining that any value above the threshold is frequent, but how this threshold is calculated has not yet been determined.

### *User behavior pattern detection process*

The following section outlines the UBPD methodology. A graphical overview of the methodology is presented in Figure 2. The methodology is broken into four main parts: Extract transform and load (ETL), sequence mining, clustering, and interpretation phases.

As with all process mining methodologies, the first stage of the methodology is an extract, transform, and load (ETL) process where the data to be analyzed are collected from the various data sources and combined, cleaned, and organized in a format to which an algorithm can be applied.

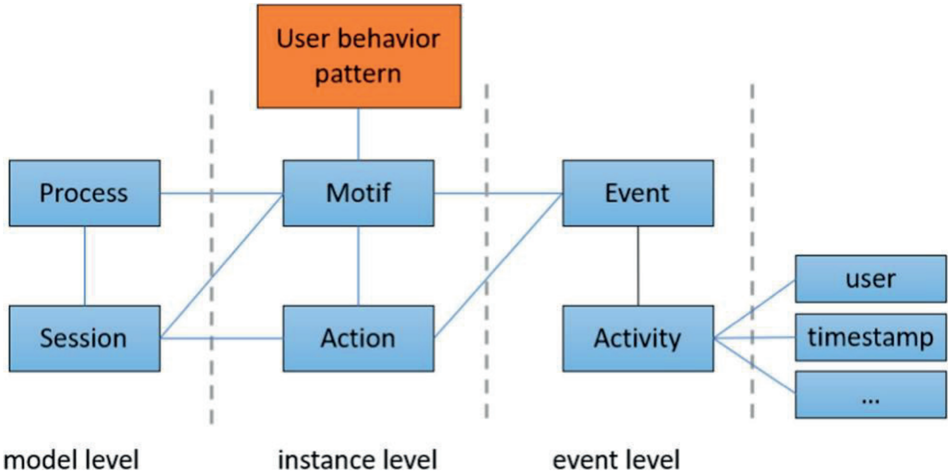



**Figure 1.** Visual representation and links between methodology terms

The ETL stage is unique for each system because data is stored and organized differently in each system, but the results need to be in a single dataset that includes, at a minimum, the user id, activity, and time of event. Activities may or may not include additional information allowing for further data analysis, but our methodology does not require it. Each activity represents an event that a user performed, however, not all logged events need to be analyzed as they might represent time-based events, error messages, or administrative tasks, that are not relevant to the understanding of user behavior. As part of the ETL configuration, designers should consider which event to include in the analysis dataset. It should be noted that in cases where a user behavior may be triggered by an event, it should not be deleted.

Designers should determine which events should be clustered together using the same action, and the ETL phase should then rename the activities dataset to include at a minimum, the user id, action, and time of action. For data processing efficiency reasons, it is useful to enumerate each action with a unique identifier to allow for faster data analysis and simplified results presentation. If there is no need to cluster events into actions, this step is not necessary, but in many cases, different events do have similar meanings.

In the second phase, the actions dataset is broken into user sessions. Each user session is prefixed with a login action and postfixed with a logout action, if they did not already exist. The output of this stage is a list of sessions that include a user identification and a time-ordered sequence of user actions within each session (Figure 3[a]). Consecutive identical actions are ignored in this process since we seek to understand the transition behavior between actions. If a user spends a long time doing something, we consider this to be a single action. For instance, if a user is reading content on a web page, and continues to read content, this is considered a single activity that does not transition from reading content to reading content.

A sliding window of size W is used to define sequences of actions with a length of W. The size of W can vary from as low as two actions and up to the size of the longest session. Smaller window sizes (e.g. shorter sequences) have an advantage because they can detect short behavior patterns that are masked when looking at wider window sizes. Due to the long tail effect, smaller window sizes also guarantee that the motifs selected are those who are more frequent. Wider window sizes are more likely to represent the true meaning of a sequence of actions, but they also reduce the number of sequences that are extracted from each session, up to the point where the window size is longer than the session length and nothing is extracted. Balancing between shorter window sizes and more meaningful sequences, it is recommended to set the upper limit of the window size to the first quartile of the session length, which means that up to 25% of the sessions are ignored. Allowing larger window sizes would result in loss of information to analyze which can harm the analysis. Analyzing the ratio between the number of unique motifs and total number of motifs, against the window size, would allow to determine the optimal window size which beyond it, increasing the window size would have a minor effect on the ratio. The output of this stage is a list of motifs of length W performed by each user. Figure 3(b) shows the output of this stage for a window size of three using the example in Figure 3(a).

A single motif represents a very short sequence of actions. In systems where users can easily navigate between different actions, we would like to understand which sequence of actions (i.e. motifs) lead to which sequence of actions most frequently. A set of motifs which are frequently performed together by some users more than others, represent a user behavior pattern. Detecting these groups of user behavior patterns is done through clustering groups of similar behaviors using an exploratory factor analysis (EFA) with the most frequent motifs as input. Each of the most frequent motifs is assigned to a dummy variables and a count of the number of occurrences of that motif for each user is done. The matrix of users and the number of occurrences for each motif (i.e. the dummy variable) by user is used as the input to the EFA. The output of the EFA is a set of constructs that represent user behavior patterns as they cluster motifs which load high on some users and low on others. The selection of EFA as the clustering method was done after using different clustering methods such as hierarchical clustering (Murtagh & Contreras, 2017), Dendrograms, and K-means. All algorithms produced similar results but the EFA was the most efficient in terms of performance and the number of configuration parameters.

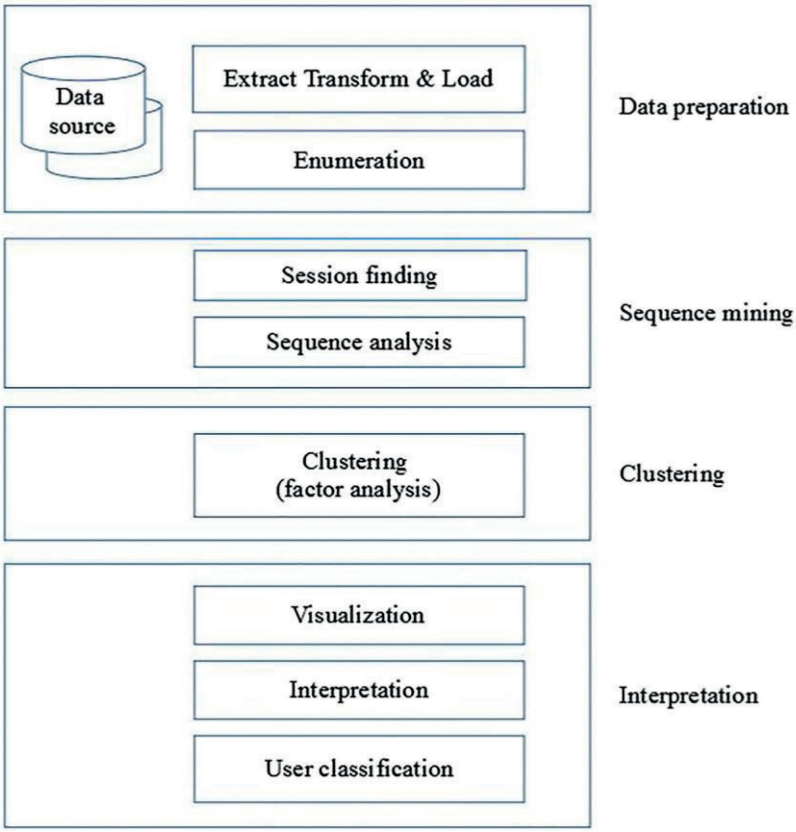



**Figure 2.** An overview of the UBPD methodology



The details of running a factor analysis are beyond the scope of this paper – for a detailed analysis see Cattell (2012); however, the result of this process is a set of constructs that includes motifs that users perform together. The exact number of constructs to expect depends upon the complexity of the system analyzed. The standard cut-off criteria of eigenvalues smaller than one can be used, unless it is possible to clearly define the number of expected behavior patterns. Since each construct includes a set of motifs (e.g. sequences of activities), the best visual representation of a construct is a causal net. Causal nets are directed networks showing the flow of activities from node to node Figure 4 shows how drawing the relations between all motifs in a construct provides a view to the user behavior pattern.

a.

| User ID | Session Data |
|---|---|
| nnn1 | A B C E G B D F Z |
| nnn1 | A B E C B G Z |
| nnn2 | A E C B D Z |
| nnn2 | A D F B G E C B Z |

b.

| W=3 | |
|---|---|
| User ID | Subsequence |
| nnn1 | ABC |
| nnn1 | BCE |
| nnn1 | CEG |
| nnn1 | EGB |
| nnn1 | GBD |
| nnn1 | BDF |
| nnn1 | DFZ |
| nnn1 | ABE |
| nnn1 | BEC |
| nnn1 | ECB |
| nnn1 | CBG |
| nnn1 | BGZ |

**Figure 3.** Schematic output of the session identification stage: (a) session data and (b) motifs for a given user.

Factor analysis provides a score for each subject on each construct. A high score on a specific construct means that the behavior represented by the construct is more salient for that user. The combination of scores given to each user on each construct represent the users' overall behavior classification. For instance, if a system has two constructs being interpreted as competitiveness and curiosity, and we can define a high-medium-low scale for each construct, nine different classes of users can be drawn from these two constructs.

The last phase in the process is interpreting the meaning of the construct. Factor analysis effectively detects when there are commonalities between the behaviors in a construct but cannot interpret their meaning, which is something that system designers and analysts should determine. System designers should also be the ones to determine the course of action to take as a result of these findings.

The methodology presented so far is based on a myriad of existing methods in process and sequence mining that are combined to interpret usage logs and detect specific recurring user behavior patterns. Executing this methodology requires the extraction of sequences of activities, which is typically a system-specific manual process, and a standard statistical software package to perform the factor analysis. While these methods are all grounded in theory, combining them to identify user behavior patterns is a novel approach. In the next section, we demonstrate the use of this methodology using two different real-life examples.

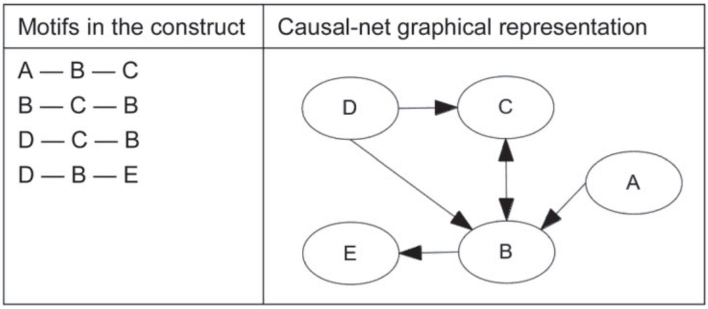



**Figure 4.** A sample representation of motifs of size three belonging to the same construct

## ***Case studies and simulation***

Both case studies presented in this paper are based on the Moodle LMS but represent different learning scenarios. The first case study is based on a standard academic course where various gamification elements were added causing the usage of the LMS to be more chaotic. The second case study is based on a MOOC with users mostly viewing videos and submitting assignments. The behaviors expected in both case studies are different. In the MOOC case study, we expect to discover users with different learning strategies, while in the gamified course we expect to find behaviors that are impacted by the gamification. Existing research already uses behavior patterns to

Figure 5 provides a visual representation, using a Petri-net structure, of the two case studies showing their actual data, along with a standard academic course with no modifications. This representation highlights the differences between courses and the inability of producing meaningful insights based on such a representation.

LMSs carry a major promise for adaptive learning and enriched learning experiences (Costa, Alvelos, & Teixeira, 2012); however, in many cases, student interactions with them are centered around downloading class material, handing in assignments, and reading announcements (Costa et al., 2012). Such tasks are atomic, or very short processes that are less interesting from a process mining lens because each task is only two or three steps long (see Figure 5-II).

### *Case study A – gamified academic course*

This first case study is based on an existing learning environment which was gamified by adding different game elements. The data used for the analysis are from four consecutive semesters in which the course was offered in the same format. Students participating in the course were undergraduate students in their third year out of four with more than 95% of the students majoring in industrial engineering and management.

*Course setting*
The main objective of the gamified course was to increase student engagement with course materials by encouraging more frequent and meaningful interactions. The main functionalities of the standard LMS were kept, and game mechanics were added. First, a discussion board was added where students and staff could discuss items relevant to the course material. Discussion boards include good design principles for the incorporation of games in education (Aviv, Erlich, & Ravid, 2005; Li et al., 2012; Lieberoth, 2015) providing interaction opportunities between students and staff, allowing students to create content, build online identities, explore ideas, and take risks (Gee, 2005a, 2005b). For each contribution to the discussion board, students received a default value of 10 credit points, and for more meaningful contributions, participants received up to 50 points. Meaningless contributions, such as "I agree with the comment above", did not grant points. Each post was graded automatically and in real-time using software developed for this purpose. The number of points each participant had was visible to all students through a leaderboard. Contribution to the discussion board was partially mandatory, as students were required to reach 600 points over the semester. However, there were other mechanisms of earning points available to those who did not feel comfortable posting their thoughts online. The average number of points achieved by students ($n$ = 303) was 792, with a standard deviation of 502, and a median of 700. The minimal amount of points was 300, and the maximum was 4418, indicating that some of the participants were extremely engaged while others were not. Many of the students continued discussions way after having reached the mandatory 600 points. Students were granted badges for completing certain activities in the discussion boards, such as contributing posts (1, 5, 10, 20, 50, or 100), responding to questions, and participating in various activities online.

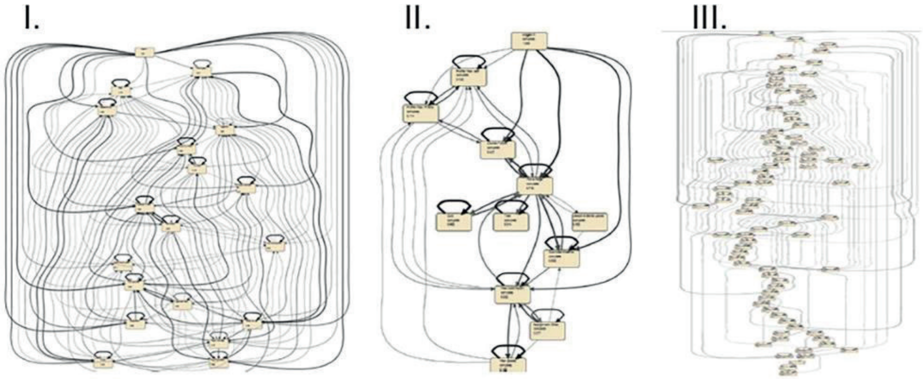



**Figure 5.** Network representation based on actual data of three types of courses. (I) Gamified course – Case study A, (II) Reference structure – Standard academic course, and (III) MOOC – Case study B

Additional game mechanics aimed to increase engagement included voluntary weekly quizzes about the material taught that week. The weekly quiz scores were summed and presented in a dedicated leaderboard that ranked students. Logic riddles or small game-theory experiments in which students could voluntarily participate were made available at certain points throughout the course.

The use of points, badges, and leaderboard game mechanics is often criticized by gamification scholars, who claim that they are trivial implementations that harm long-term intrinsic motivation (Barata, Gama, Jorge, & Goncalves, 2013; Hanus & Fox, 2015; Mekler et al., 2013). While this may be true in some cases, for students whose intrinsic motivation is weak to begin with, these mechanics have been found to be successful for short-term tasks (Anderson et al., 2013; Butler, 2013; Hakulinen et al., 2013; Landers & Landers, 2015; Mekler et al., 2013) and were thus used in this study.

***Data preparation***

The log file used for analysis included 504,040 activities performed by 381 students participating in the course. The number of unique activities was 127 out of which 57 were deemed as system events such as emails sent and password reset requests or other redundant activities, leaving 70 activities in the analysis. These activities were mapped to 29 distinct actions – combining, where appropriate, similar activities into a single action.

A Perl program developed for this purpose takes the base dataset and processes it, separating the base dataset into sets of sessions. Using the sessions dataset, a separate dataset is created for different window sizes, which will later assist in the selection of the appropriate window size for the specific case. The window size selection is a key factor that must be determined at the beginning of the analysis. Analyzing the effect of increasing the window size on the average number of motifs per unique motif is shown in Figure 6. We would like to increase the window size up to the point where increasing it further, simply creates many unique motifs with very few instances in each. Based on the knee demonstrated in Figure 6 it is possible to determine that the right window size is three and that beyond that window size, the number of motifs per user does not change much.

Table 2 summarizes the impact of the window size on the number of motifs extracted and the number of unique motifs extracted. As window size grows, fewer motifs are extracted, and more of them are unique making them harder to analyze. A smaller window size means fewer actions are included, making the results less robust.

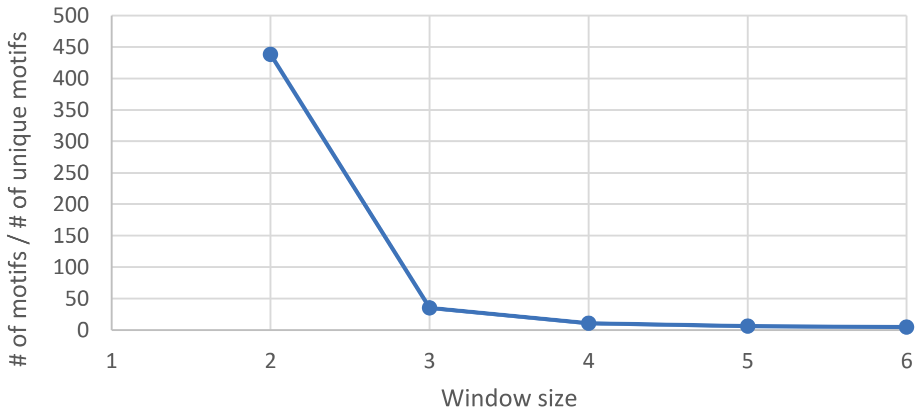



**Figure 6.** The ratio between the number of motifs and unique motifs – Case Study A

***Pattern detection***

Next, the motif dataset for a window size of three was processed by an R program developed for this purpose using the psych package and the embedded factanal procedure. The program summarizes the different motifs per user and performs an EFA based on the most frequent motifs using a varimax rotation. Since there is no prior assumption as to the number of factors to extract, the eigenvalue lower or equal to one criterion (Kaiser, 1960) was used. While additional methods exist for making this decision, such as parallel analysis (Horn, 1965), the method we use examines many different combinations of motifs and factors, allowing us to determine the optimal number for this problem. Eigenvalue was selected due to it being computationally simple and commonly used in research.

**Table 2.** Window size calculations for case study A

| Window size | # of motifs | # of unique motifs | # of motifs / # of unique motifs |
|---|---|---|---|
| 2 | 119662 | 273 | 438.32 |
| 3 | 68187 | 1931 | 35.31 |
| 4 | 56534 | 5203 | 10.87 |
| 5 | 47953 | 7581 | 6.33 |
| 6 | 41683 | 8801 | 4.74 |

The results of this analysis are Petri nets representing user behavior patterns. Petri nets in this context, are used as a graphical tool similar to flowcharts, block diagrams, and networks (Murata, 1989) and are commonly used to represent processes (De Medeiros & Weijters, 2005). Defining what counts as most frequent is not straightforward. Ideally, the entire population of motifs would be included in the analysis, but since there are significantly more motifs than users, there is a limit on the ratio between motifs and users. A high ratio of 1:100 would result in fewer factors that do not explain variability, while a low ratio of 1:3 may result in an unreliable model since EFA is sensitive to such cases (MacCallum et al., 1999). The model was executed several times with different ratios, to assess the optimal ratio. As more motifs are included in the analysis, it is expected that the number of factors discovered will increase, and this is indeed what happened. However, more factors do not necessarily mean a better result, as factors may either be meaningless or repeat themselves with slight variations if the model is overfitted.

The frequency and variability of motif occurrences may also influence the ratio selection. As shown in Figure 7, there is a significant long tail effect, and the top 20 motifs account for nearly 65% of all motifs. However, the ratio between the frequency of appearance and variability is noisy, meaning that some of the less-frequent motifs create more variability, indicating that a higher number of motifs should be used to include more variability in the analysis.

Determining the right number of motifs to include in the analysis was done by running the analysis several times with different numbers of motifs and optimizing between the explained variance of the model and the number of motifs used. The results of this analysis are summarized in Figure 8. The x-axis shows the number of motifs introduced into the model. Left y-axis shows the number of factors discovered by the model, and the right y-axis shows the actual ratio used by the model after removing motifs that do not significantly load on any factor. The right y-axis also show the explained variance of the model. Ideally, a parsimonious model is preferred allowing for a minimal number of motifs and factors, explaining the maximum variance in the data. Taking this into account, a model using 36 motifs representing a 1:16 ratio was selected, explaining 75% of the variance, generating five distinct usage behavior patterns.

The model using 36 motifs was finally executed resulting in five factors. Patterns are presented as Petri nets, making them easier to understand visually. While EFA provides the understanding that a certain behavior is salient, the reason for the pattern being salient is a matter of interpretation. Table 3 shows the emerging patterns and a subjective interpretation based on our understanding of the environment in case study A.

While the results of case study A are plausible, we wanted to test the validity of the results by supplementing the actual data with simulated data of patterns that do not exist in the original dataset. If the methodology can detect these new patterns, our confidence in the correctness of the results is higher. In addition, if the results, can reproduce the same patterns as the data prior to simulation, our confidence in the validity of results is higher.

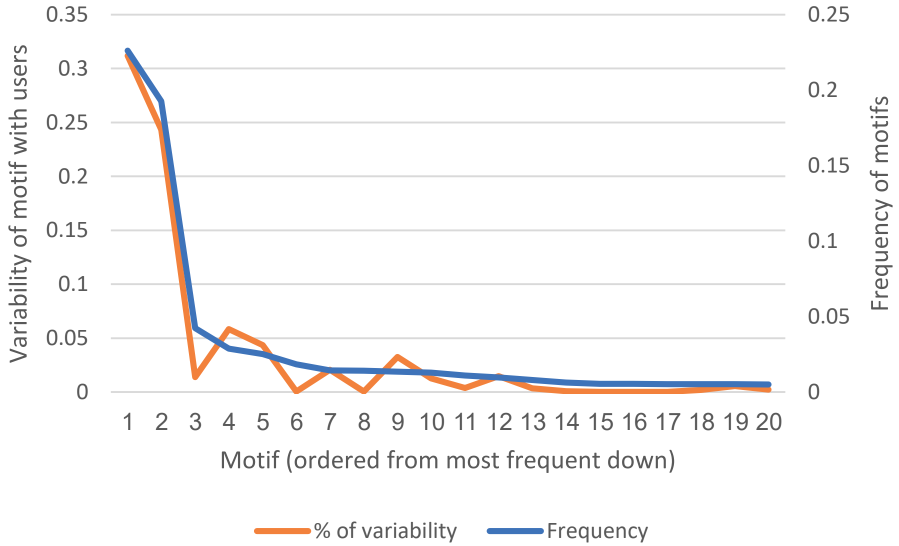



**Figure 7.** Variability and frequency of top 20 motifs

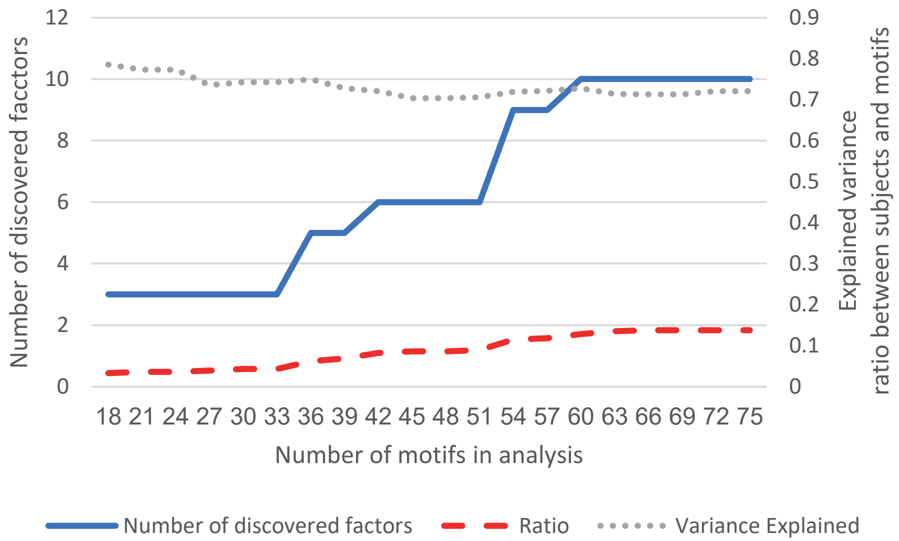


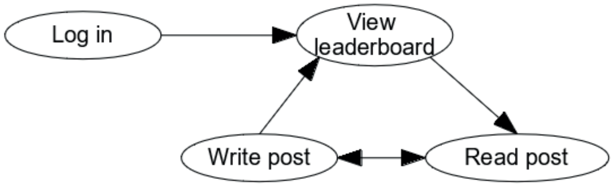
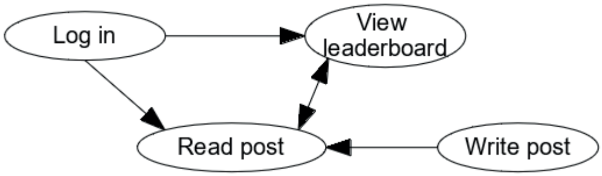
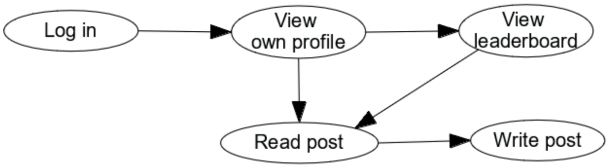
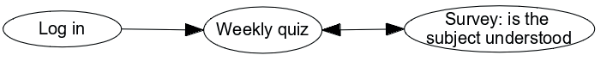

**Figure 8.** Summary of executing the model several times using different ratios – case study A

The data generated through the simulation process included the two patterns shown in Figure 9. The procedure for generating the data for pattern A was such that for each user, a random number of motifs representing actions that appear in the new patterns was generated, using a normal distribution. To include some variability, 30% of the motifs were set to be positive-false, i.e. represent a sequence of actions that involve the additional actions but do not match the pattern. Pattern B was simulated such that 40 motifs that match the patterns were randomly generated for every third user, ensuring significant variation between users. While adding variability to the patterns is necessary as the methodology is based on detecting variability, the value of 30% was arbitrarily chosen. As the variability increases, there would be no pattern to detect while on the other hand, with very low variability clustering method based on variability would not detect these patterns.

A window size of three was used for both the simulated model and the actual model, allowing better comparison between them. The simulated data included 80,943 motifs, out of which 2001 were unique motifs. These values are comparable with those found in Table 2 for the non-simulated data. A descriptive view of the data is shown in Figure 10 showing comparable results to Figure 7.


**Figure 9.** Simulated patterns

**Table 3.** Usage patterns - case study A

| Behavior pattern | Pattern | Possible interpretation |
|---|---|---|
| A1 |  | Content contribution. The user logs in and is curious about his leaderboard position. He contributes and reads posts checking its influence on his position compared to others. |
| A2 |  | Content reading. The main focus of this behavior is viewing content. It may include viewing the leaderboard or checking the status of the user's or other users status. |
| A3 |  | Badge collection pattern. Badges were given for contributing data and were presented on the user's profile page. The key reason for a user to visit his profile page was to view their badges. In this behavior, the user logs in and looks existing or newly received badges, which leads him to explore additional status items such as the leaderboard, and to contribute more content. |
| A4 |  | Knowledge points collection pattern. Two mechanisms were available for collecting knowledge points and in this pattern, users performed both sequentially. Knowledge points were the second type of points available for collection. |
| A5 |  | Social networking pattern. Users reading content that other users posted would be curious about the users' postings and visit their profile pages to read about them and view their badges. |

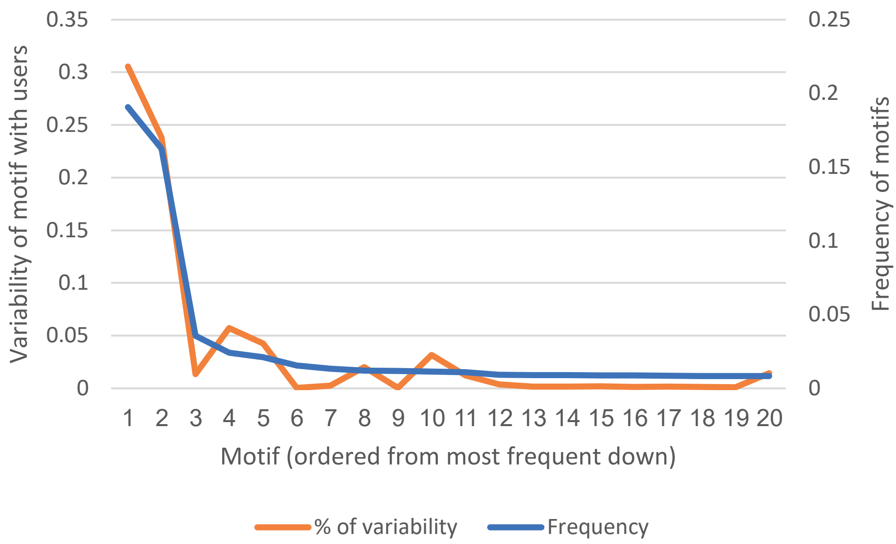



**Figure 10.** Frequency and variability of top 20 motifs - simulated data

Next, the model was executed several times using different numbers of motifs as input to the EFA to determine the correct number of motifs to include in the analysis. The selection criteria were as before: fewer motifs, higher explained variability, and fewer factors. While Figure 11 indicates that a simple model of 18 motifs can be used, we selected a model with 36 motifs, which provides close results to that of 18 motifs but richer behavior patterns. As expected, the simulated model successfully identified the simulated patterns and behaviors A1, A2, and A4, as shown in Table 3. Increasing the number of motifs above 51 resulted in identifying behaviors A3 and A5 as well.

To summarize case study A, the UBPD algorithm detected five key behaviors performed by students in a gamified academic course using an LMS. The detected behaviors were related to the gamification of the course and how different students interacted with them. Unlike existing algorithms, there was no prior knowledge required about the existence of these behaviors, and their discovery and relating them to students was fully automated. The discovered pattern supports prior research indicating that different people are engaged differently by gamification (Codish & Ravid, 2014b; Hamari, Koivisto, & Sarsa, 2014).

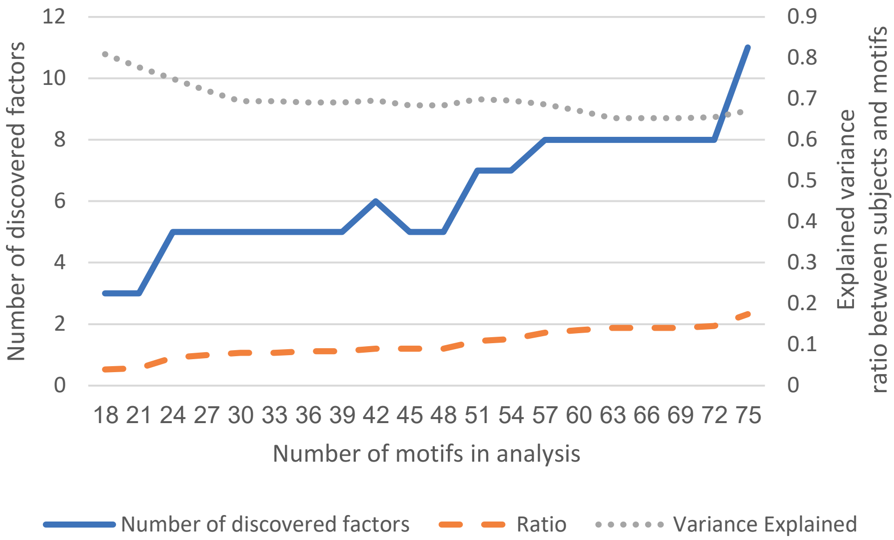



**Figure 11.** Summary of executions using different input variables - simulated Data

Including simulated data into the original data makes it possible to examine the validity of the algorithm. Original patterns were reproducible but required the inclusion of a larger number of motifs in the model, which is reasonable considering that instead of generating the original five behavior patterns, the simulation data were required to generate at least seven patterns. The simulated patterns appeared as they were expected to appear, despite the inclusion of positive-false motifs to the data indicating the algorithms ability to deal with noise.

***Case study B – MOOC***

In the second case study, data derived from system logs of a mid-sized MOOC on the recent history of the Middle East delivered in Hebrew were examined. The MOOC was offered by the Open University of Israel between 4 April 2015, and 7 July 2015. Students considered in this analysis were those who enrolled in the MOOC to get access to all the course materials and teachers (Kalz et al., 2015) and did at least one activity in the course. The course was freely available to the public without any prerequisites on knowledge or any other obligation and did not offer an academic recognition for completion of the course. During the course, participants' activities were recorded in a log-file.

MOOCs have specific characteristics that make them excellent candidates for learning analytics (Clow, 2013; Coffrin, Corrin, de Barba, & Kennedy, 2014; Kizilcec, Piech, & Schneider, 2013). They typically include many participants, have detailed log files, a good diversity of participants, and a process which is loosely defined. In most MOOCs, learners are expected to follow a standard process of watching video lectures in a specific order, answer quizzes and participate in online discussions. The key benefit of a MOOC is that it allows users to follow different paths that suit their learning styles, objectives from the course, time constraints, and other factors influencing their decisions. Therefore, while a main process does exist, learners will often deviate from it. Figure 5(c) shows a process map for a standard MOOC where it is clear there is an overall process, but various deviations are apparent.

*Data preparation*
The data file included data from 367 out of 1942 participants in the course, who agreed to have their data included in this analysis. Participants age ranged between 18 and 85 years ($M$ = 61, $SD$ = 14.01). Fifty-six percent were males. For most (63.7%), this MOOC was their first online learning experience, and they indicated themselves as having high Internet skills ($M$ = 6.23, $SD$ = .65, in a scale range from 1 “Has very low Internet skills” to 7 “Has very high Internet skills”).

The data file was clean of non-relevant data and included 93,942 log entries with 86 unique activities. As done in the first case study, an analysis to determine the best window size was executed. The results of this analysis appear in Figure 12 and show that as before, beyond a window size of three, the ratio between motifs and unique motifs becomes very low, which would result in low variability, making EFA less effective.

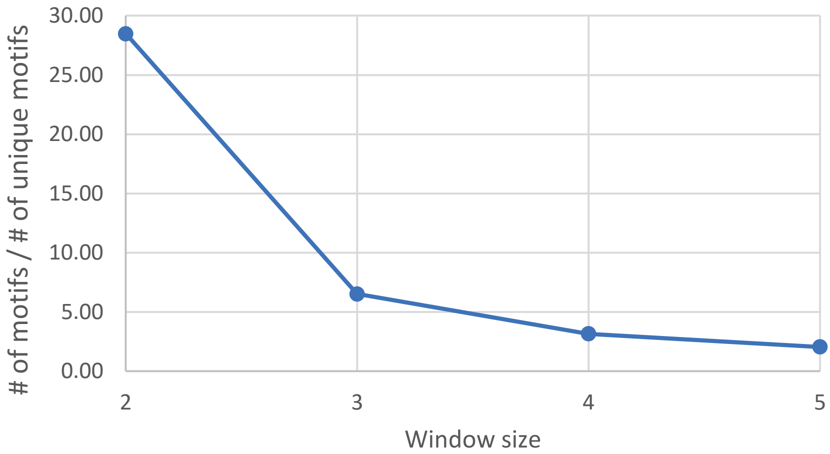



**Figure 12.** The ratio between the number of motifs and unique motifs - case study B

*Pattern detection*
Based on the window size analysis, motifs of window size three have been included in the pattern detection algorithm, and the model was executed 20 times with a different number of motifs each time to determine the best model. The results of this analysis can be viewed in Figure 13. Forty-two motifs were included in final analysis based on

Chapter 5

the observation that at this number, the explained variance was almost the highest while keeping a low ratio and fewer factors. Finally, patterns were extracted through the EFA process, and interpretations of the factors are shown in Table 4. The visualization of patterns through Petri nets are shown in Appendix A.

To assess the impact of selecting more motifs into the analysis, the same model was executed with 57 motifs, which as shown in Figure 13, provide a similar level of explained variance while producing two additional behavior structures. For the analysis to be sound, it is expected that adding more motifs into the analysis will produce a similar set of behaviors, with richer data, which indeed happened. All behaviors detected with 42 motifs. The additional motifs detected appear in Table 4 as behaviors C8 and C9.

Case study B demonstrated the ability to extract the behavior patterns of students participating in a MOOC. A total of seven behaviors were extracted using a minimal set of motifs, and an additional two behaviors were extracted when using a larger number of motifs. While some of the behaviors were expected, such as in the case of B4 in Table 4, others were more surprising, such as in C8 where there are users who focus mostly on the first video lectures for every week.

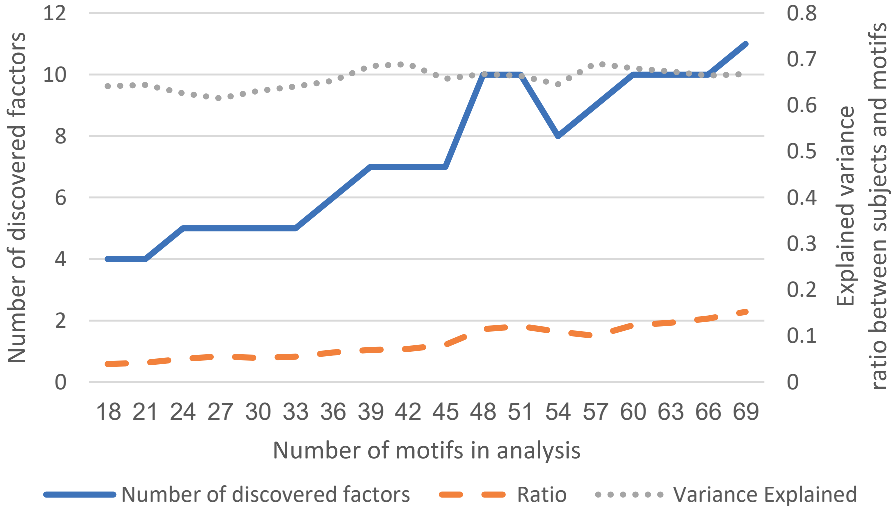



**Figure 13.** Summary of executions using different input variables - case study B

**Table 4.** Usage patterns - case study B

| Behavior Pattern | Possible interpretation |
|---|---|
| B1 | Users were motivated to complete all weekly quizzes. The weekly quiz is a self-evaluated activity that enables learners to evaluate their knowledge base on materials covered in the previous week. |
| B2 | Sporadic first-week behavior. Users expressing this behavior viewed the first videos of the course one at a time and not sequentially. |
| B3 | Users who mostly viewed the first videos of weeks 2–4 non-sequentially. This is a sporadic behavior that can be interpreted as an exploration behavior of merely checking on each week's topic, but not completing it. |
| B4 | Users who viewed each week's lectures in sequential order. This is the expected behavior of a learner. |
| B5 | Users who viewed lectures 1.4 and 1.5 not sequentially. Unlike B2 where users viewed lectures 1.1, 1.2, and 1.3 – Users who are strong on this behavior also viewed lectures 1.4 and 1.5 in a non-sequential manner. Users low on this behavior are those who did not continue to view the remaining lectures of the week. |
| B6 | Users who accessed the site to view announcements in the general discussion forum. The general discussion forum was used as a social tool enabling learners to receive updates about the course progress and to introduce themselves to the learners` community. |
| B7 | Users accessing the site to view week four forum. This behavior received no plausible explanation from the course staff. |
| C8 | Users viewing the first lectures for each week. People with this behavior viewed the first and sometimes also the second lecture of each week non-sequentially. These might be people who are interested in the introduction to each topic without going into more detail. |
| C9 | Users who viewed all of the first weeks' lectures sequentially. These would be people who were fully engaged only at the beginning. |

## Discussion and conclusion

Process mining is typically used to uncover underlying business processes and deviations from them by discovering actual user behavior and comparing it with the expected behavior (Van der Aalst et al.,2012; Van der Aalst & Weijters, 2004). While successful at discovering well-structured processes, it is less successful in less structured processes where users have the freedom to execute the process in different ways. The challenge in the latter case is to detect these differences and understand if there is a reason for different users to behave differently. Our research question in this paper is: Within an unstructured process or system, can we automatically identify recurring user-level behavior patterns and perform user clustering based on these patterns? Specifically, as we focused on learning environments, these user behavior patterns can be viewed as learning processes.

This paper presents the user behavior pattern detection (UBPD) methodology along with two case studies based on LMS implementations, demonstrating its usage, and thus, answering this research question. Simulation data were included to present the effectiveness of the methodology in discovering patterns that were injected into the data. In the first case study, a simple academic course was used, but after adding several game elements into it, it has become a complex, unstructured system. The second case study was based on a MOOC, where users have the freedom to decide what to do and how to do it. The differences between these two cases are evident when looking at Figure 3.

UBPD is unique in its focus on finding user behavior patterns that exist for only some users. It uses EFA to detect groups of activities performed together that explain the variability in the system. However, in processes with no variability in which all users perform a process in the same way, UBPD would not be of use. In systems where some of the processes are structured, and some are not, UBPD would detect the unstructured processes, ignoring the structured processes. In such cases, UBPD does not replace existing methods but rather complements them. The user clustering, which has been described above, is another key benefit of the methodology, as it provides insight into different user behavior patterns.

Several parameters and decisions were included in the methodology and are discussed in the order they appear within the methodology. The selection of actions to include in the analysis has a direct influence on the resulting patterns. Grouping activities into actions is often a straightforward task since it should be clear which activities should be grouped; however, it is important to ensure that the grouped activities represent a clear action. For instance, in the educational setting used in this study, all activities related to the submission of an assignment were grouped into an assignment submission action since they all have the same meaning. In both cases studies, activities such as resetting a password or downloading a presentation were not included in the analysis, however, this does not always have to be the case. Resetting a password is an administrative task, and thus not included, but if it is included, and UBPD detects it as a user behavior pattern (i.e. enough variability exists between users with regards to that activity), perhaps it indicates that some users are more forgetful than others. If actions only have a few occurrences, they will be removed later as part of the EFA process since they would not be considered frequent motifs.

In both case studies and the simulation data, a window size of three was selected. Figure 6 and Figure 12 show that beyond this size, the number of unique motifs grows significantly, resulting in many motifs with only a few occurrences per user. This window size might differ in other systems, and it is recommended validate this number for different situations and dataset sizes. In case study A, the dataset was larger and included fewer users and fewer actions. This resulted in a stronger tail effect than in case study B, which had a smaller dataset, significantly more users, and more actions analyzed. An additional reason for keeping a smaller window size is that using a large window size carries the risk of missing short usage patterns of two or three actions.

It can be assumed there is no known number of factors to expect during the EFA stage. Typically, EFA tries to maximize the explained variance, which in both case studies resulted in a minimal number of motifs to include as variables, and as a result, extracted factors. Including too many motifs into the analysis can result in overfitting and extracting meaningless patterns. Also, in cases in which there are few subjects, as in case study A, there is a limit on the ratio between subjects and motifs that must be kept (MacCallum et al., 1999). Since we are interested in extracting rich behavior patterns, we executed the model several times with a different number of motifs and selected a point that balanced these limitations. In case studies A and B, we demonstrated how adding motifs to the analysis does not change the discovered factors and can only result in additional factors. While this step was executed manually, it is possible to automate this step to determine the right number of motifs to include.

The validity of the resulting factors has been tested in several ways. First, simulation data have shown that when known patterns were injected into the existing dataset, the methodology was able to detect them correctly without impacting the existing patterns. This ability provides the confidence that the detected patterns are correct. Additionally, while increasing the number of motifs in the analysis increased the number of factors, only new patterns were added without impacting existing patterns, emphasizing the stability of the discovered patterns. The objective of process mining is to discover an underlying process, but the meaning or reasons for a discovered process are left in the hands of system analysts to explain. In both of our case studies, the resulting patterns were presented to analysts and their interpretation of the results is included in Tables 3 and 4. Case study B, however, includes a pattern that was repeated in the two executions of the UBPD that had no plausible explanation by designers. It is possible that such a pattern indeed exists, but designers are unaware of it. It is also possible that it is a factor that should have been removed since it is based on a single motif (Streiner, 1994). Even if we ignore the unexplained patterns, UBPD was capable of automatically detecting user behavior patterns within unstructured processes, which is a task with which existing methodologies struggle (Rebuge & Ferreira, 2012; Van der Aalst, 2011b).

This paper presents three key contributions to the world of process mining, as well as several contributions to the development and analysis of interactive learning environments. From a process mining perspective, it provides the ability to discover different usage patterns of different users. While existing methodologies focus on the detection of the processes or sub-processes of a system, UBPD seeks to find the variance in how users interact with the system. To demonstrate this point, assume that all learner in a LMS perform a specific task similarly, such as reading an essay and immediately answering some questions about it. Methodologies such as episode finding, Apriori, or GSP would easily detect this pattern; however, UBPD would not since it would be performed by all users similarly. On the other hand, if different users performed that process differently (e.g. some read and answer questions immediately while others read part of the essay, answer a question, leave, and then come back to complete the task), the algorithms above might not detect any process at all, whereas UBPD would detect the process and the different ways people performed it. UBPD will even provide insight into which users are doing what. This was evident in the simulation we performed in case study A, where UBPD did not detect a pattern that was included to all users. However, when adding a pattern to only a few users, it was immediately detected.

The second contribution is the ability to deal with noise even within a sub-process. Existing methodologies seek stable processes or, such as in the case of association rules, stable relations between activities. UBPD detect similar motifs and through EFA, groups them into meaningful patterns represented as Petri-nets in Table 3. Finally, the methodology produces factor scores from the EFA to each user for each pattern, indicating how salient this behavior is for each user. A user can receive a high score on several behavior patterns, indicating those are the behaviors they perform most, or a low score on all behaviors meaning the discovered patterns do not represent their behavior. Using these scores to produce on-the-fly user clustering, is a unique capability that UBPD introduces and can be further explored.

From an educational point of view, UBPD detects how different learners interact in a learning environment. When designing a learning environment, educators often have a specific course of action that learners would follow, such as view all lectures sequentially, yet many do not follow that path. Being able to understand learner preferences, can help designers ensure that their design addresses these different preferences. In this study, we examined learner behaviors across a full semester, but it is possible to use shorter time frames such as a week or a month, and understand how learning preferences evolve.

Being able to provide close to real-time feedback on individual learning processes and comparing these processes with other learners and learning objectives carries a great potential for future developments in the field of personalized learning and adaptive learning. Detecting the learning processes currently being used and giving each learner a score on them can be used in many ways. Learners can see their learning process compared to other, which can be further used to modify or enhance certain behaviors. Teachers can use this data to assist specific learners and adapt their teaching styles, system designers can use this data to redesign or improve learning environments, and last, adaptive systems can automatically modify themselves based on actual usage data to encourage required changes in learning behaviors.

## Limitations and next steps

Although simulation has been used to demonstrate the ability of UBPD to detect processes successfully, additional simulations should be done to determine the sensitivity of the methodology to variability. If there is no variability, processes would not be detected, and if the process is too variable processes would not be discovered since EFA would remove the actions from the analysis. This additional analysis was not included in this study to keep the focus on the paper on the methodology and should be further examined.

The process of selecting activities for analysis and combining activities into actions requires additional analysis. In the proposed methodology, this is part of a manual ETL process, but ideally, it can be automated using clustering methods. Additional manual steps, such as determining the correct number of motifs to include in the EFA, should be automated.

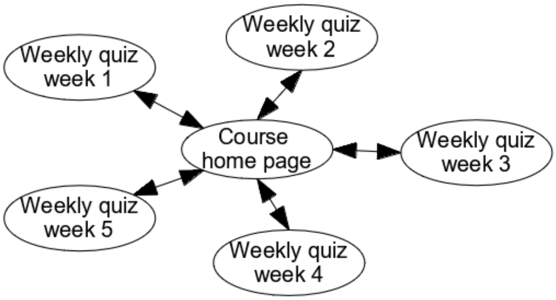
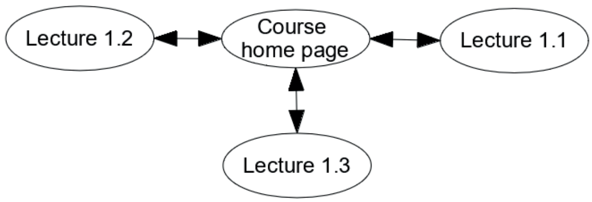
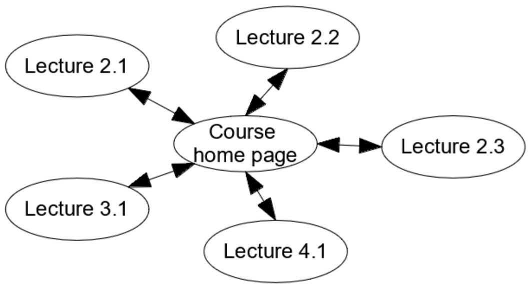
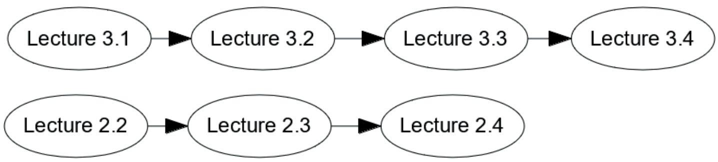
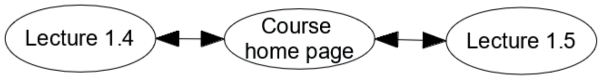
The clustering method used in this study was EFA which loads most of the variance on the first cluster. While different clustering methods have been examined throughout the study, a more in-depth comparison of different methods should be done, acknowledging that for different domains, different clustering methods might be more suitable. In addition, once a clustering is validated, different machine learning methods can be applied to further improve the clustering.

The two case studies came from a similar domain of LMS. Data from other types of systems should be analyzed to ensure the external validity of the methodology. Finally, in LMS and MOOCs specifically, user behavior changes over time. In future studies, a temporal model should be included checking user behaviors over time and providing meaningful data to system analysts as to what is happening right now in the system, not merely an overall of how the system is being used. The stability of behaviors can be tested as well over time since some behaviors might be salient at the beginning of a course and not at the end.

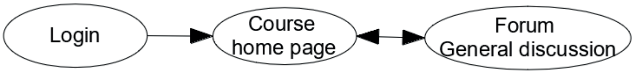
## Appendix A. Factor analysis results for different window sizes.

Behavior patterns B1–B7 shown in Table 5 are patterns that appeared when using 42 motifs in the EFA phase. These behaviors occurred again when using 57 motifs, mostly with richer patterns.

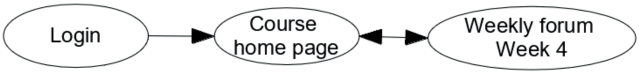
**Table 5.** Usage patterns - case study B

| Behavior Pattern | Pattern |
|---|---|
| B1 |  |
| B2 |  |
| B3 |  |
| B4 |  |
| B5 |  |

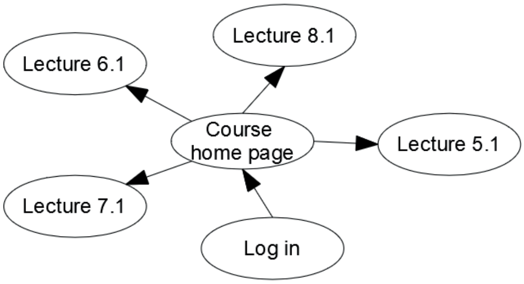
B6


Login
Course
home page
Forum
General discussion

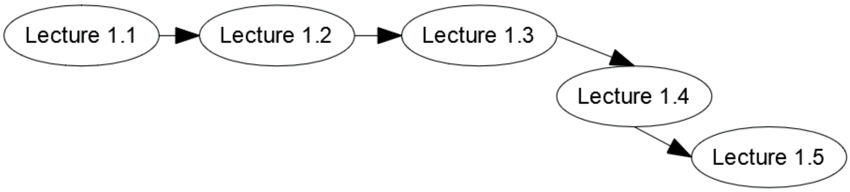
B7


Login
Course
home page
Weekly forum
Week 4

C8


Lecture 8.1
Lecture 6.1
Course
home page
Lecture 5.1
Lecture 7.1
Log in

C9


Lecture 1.1
Lecture 1.2
Lecture 1.3
Lecture 1.4
Lecture 1.5

6

# Chapter 6

## Identifying Learning Activity sequences that are Associated with High Intention-Fulfillment in MOOCs

This chapter is based on:

Rabin, E., Silber-Varod, V., Kalman, Y.M., & Kalz, M. (2019). Identifying Learning Activity Sequences that are Associated with High Intention-Fulfillment in MOOCs.

*The 14th European Conference on Technology Enhanced Learning (http://www.ec-tel.eu/). Delft, the Netherlands. September, 2019*

## Abstract

Learners join MOOCs (Massive Open Online Courses) with a variety of intentions. The fulfillment of these initial intentions is an important success criterion in self-paced and open courses. Using post course self-reported data enabled us to divide the participants to those who fulfilled the initial intentions (high-IF) and those who did not fulfill their initial intentions (low-IF). We used methods adapted from natural language processing (NLP) to analyze the learning paths of 462 MOOC participants and to identify activities and activity sequences of participants in the two groups. Specifically, we used n-gram analysis to identify learning activity sequences and keyness analysis to identify prominent learning activities. These measures enable us to identify the differences between the two groups. Differences can be seen at the level of single activities, but major differences were found when longer n-grams were used. The high-IF group showed more consistency and less divergent learning behavior. High-IF was associated, among other things, with study patterns of sequentially watching video lectures. Theoretical and practical suggestions are introduced in order to help MOOC developers and participants to fulfill the participants' learning intentions.

***Keywords:*** Massive Open Online Courses, Intention-fulfilment, Keyness, N-gram, Learning Activity Sequences.

# Introduction

### *Participants Retention and Completion in MOOCs*

Massive Open Online Courses (MOOCs) demonstrate the potential of scaling higher education by means of digital media and the Internet. More than 100 million participants signed up to 11,400 courses from 900 universities around the globe (Shah, 2018). MOOCs enable participants of different academic backgrounds to study at any time and in any place, to enhance their learning experience and to gain important 21st-century skills free or at significantly lower costs. The high potential of MOOCs has been criticized due to low retention and completion rates (Gardner & Brooks, 2018; Reich & Ruipérez-Valiente, 2019) that often drop below 10% of the participants who registered to the course (Chuang & Ho, 2016; Jordan, 2014; Margaryan et al., 2015).

### *Intention - fulfillment*

Some researchers have questioned whether completion rates and completion certificates are the appropriate measures for evaluating the success of this new form of life-long learning (Henderikx et al., 2017; Rabin et al., 2019). Their basic claim was that the success of lifelong learning in MOOCs should be evaluated not through traditional instructor-focused measures such as dropout rates and earning of completion certificates, but rather through learner-centered measures that take into account the informal nature of MOOC learning. One such measure is intention-fulfillment (IF) which measures the extent to which the learners fulfilled the initial intentions they had when accessing the course. This measure takes into account the personal objectives that the learners intend to achieve, rather than external success criteria (Henderikx et al., 2017). In MOOCs and in other forms of open education, students may enroll with different intentions that effect their learning behavior (Littlejohn et al., 2016; Onah et al., 2014; Wang & Baker, 2018). From that point of view, a successful learning experience can take a variety of forms ranging from viewing a single lecture, attaining a specific skill, or studying a topic of interest, to studying a whole course and fulfilling all of its formal requirements. Thus, the participants´ intentions and their fulfillment should take center stage when evaluating the participants´ success in the course.

### *Learning activity sequences*

Learning behavior in MOOCs is mostly visible through logs, which record access and usage patterns of the different course resources (e.g. video lecture, quiz, etc.). Many MOOC studies are based on simple access logs, counting each time the learner accessed or used a course resource, but ignored the order of the activities and their sequential nature (Li et al., 2017). Taking into consideration only the number of activities that the participants performed and ignoring the sequence of activities, provides only a partial picture. For example, as demonstrated by Li, et al. (2017), if we consider three imaginary participants who watched videos (V) and answered quiz questions (Q), one of them can watch all the videos and then answer the quizzes (V-V-V-Q-Q-Q) while another participant might first try to answer the quiz questions and only then watch the video lectures (Q-Q-Q-V-V-V). A third participant might follow each video by a quiz (V-Q-V-Q-V-Q). Although all three fictional participants watched three videos and answered three quizzes, their learning paths, or sequences, are fundamentally different.

Several researchers attempted to understand differences between the learning paths of MOOC participants who passed or failed a course. It was found that learners who passed the course followed a path that had different characteristics than those who did not pass the course (Davis et al., 2016; Guo & Reinecke, 2014). For example, replaying videos more than once, and watching a relatively high percentage of the course videos, were positively correlated with finishing the MOOC (Sinha et al., 2014) On the other hand, Van den Beemt, Buijs and Van der Aalst (2018) found that successful students exhibit a more steady learning behavior and that this behavior is highly related to regularly watching course successive videos in batches.

Several studies used natural language processing (NLP) features in order to study MOOC participants' dropout and retention mainly by studying the language students use (Crossley et al., 2016; Kim et al., 2015; Robinson et al., 2016). However, we found only few studies that applied NLP methods such as n-gram analysis, to study learner activity sequences (Li et al., 2017). None of those studies had used NLP methods in order to predict subjective success outcomes in MOOCs such as intention-fulfilment. In this study, we apply methods that originate from the NLP realm, to analyze learning activities and learning activity sequences and to compare those activities and activities sequences between participants who report high-IF and participants who report low-IF.

## Method

### *Sample*

In the current study, we used clickstream data gathered from log files of 462 participants in a MOOC teaching the subject English as a Second Language (ESL) to identify the learning process of the participants. The data collection for the current study was carried out between July 2016 to February 2018. During this period, the participants were able to join and leave the offered MOOC whenever they liked to.

### *Course activities and their annotations*

MOOCs usually comprise of modules such as video lectures, quizzes and other resources (Lackner et al., 2014). The manner in which students interact with these course resources are considered conceptualizations of their higher-order thinking, which lead to knowledge construction (Chi, M, 2000). In this ESL-MOOC, the participants were able to choose ten different types of activities in any order, place and time. The course was arranged by units. Each unit contained an introductory page (I). This page pointed participants to several additional resources: a list of learning strategy videos (S), a PDF reading comparison text that is used throughout the unit (P), a recommended learning track (T), several lessons (L) quizzes (Q) and a final exam (E). Each of the lessons comprises of a single video (V) and links to specific learning strategy videos (S). Participants who watched videos could click the video play/pause button according to their personal progress during the video lecture. Although the course does not provide academic credit, the participants could get a participation badge (B) if they answered all the questions in the quizzes and achieved a predefined minimum score. The participants were also able to watch the list of rights (R) (credits) of the course materials. In total, we harvested 61,713 activities. It is important to note that the logs only recorded the clicks, and did not record other activities (e.g. reading text, feedback on quizzes). Table 1 summarized the courses' activities, their codes, and a short description of each.

**Table 1.** Course activities – codes and description

| Activity | Code | Description |
|---|---|---|
| Badge | B | A page that enables the participant to see their achievements during the course |
| Exam | E | Self-administered final exam that summarizes the entire course |
| Introductory page | I | The participant accessed an introductory page of the course |
| Lesson | L | The participant entered a page that includes a video lecture, a list of skills that will be taught in the unit and relevant learning strategies (S) |
| Pdf text | P | The participant accessed a reading comprehension PDF text that was used in the lesson |
| Quiz | Q | Closed questions with immediate feedback. The participant had been able to answer the same quiz more than one time |
| Rights | R | A page that includes the credits and rights to course materials |
| Learning strategy | S | The participant watched short and focused videos dealing with learning strategies |
| Track | T | The participant accessed the page that provides the recommended learning track of a lesson. |
| Video play/pause | V | Each time a participant pressed the play/pause button in a video lecture |

### *Computational tool kit for sequence analysis*

Preprocessing: In order to use the NLP tools to analyze learning sequences, each participant`s sequence of learning activities was coded as mentioned above in Table 1.

For the sequence analysis, we used Antconc 3.5.7, a multiplatform toolkit developed for carrying out corpus linguistics research and data-driven learning (Anthony, 2018b, 2018a). Specifically, we used two NLP methods: n-gram tool, and keyness tool.

The n-gram tool allows us to find common "expressions", i.e., common sequences of activities, and their transitional probabilities. In the current study, the n-gram analysis consisted of uni- bi-, tri-, and four-grams calculations by Antconc. For each group separately (high-IF or low-IF), we sorted the $n_i$-gram lists according to their probability values. We then excluded activities with probability below 0.1, and calculated two measures:

1. The *relative frequency* of each $n_i$-gram sequence was calculated by dividing the ab- solute frequency of that $n_i$-gram sequence of activities by the total number of $n_i$-grams in that group. For example, the bi-gram sequence V-V occurred 6,767 times in the low-IF group, which was divided by 25,742 (total number of bi-grams in that group), resulting in a relative frequency of 26%.

2. *Participation range* was calculated by dividing the number of participants that performed each $n_i$-gram sequence of activities by the total number of participants in that group. Thus, the participation range is the relative distribution (entropy) of each $n_i$-gram sequence. For example, 186 participants out of the 231 participants in the low-IF group performed the V-V sequence. Therefore, the relative distribution of this sequence is 81%.

The keyness analysis was carried out in order to identify the activities that are unusually frequent (or infrequent) in one group in comparison with the activities in the other group. The keyness analysis provides an indication of a keyword's importance as a content descriptor in a given corpus relative to a reference corpus (Biber et al., 2007). "A word is said to be "key" if [...] its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a p-value specified by the user" (Scott, 2011). The statistical significance of keyness is calculated by using the value of log likelihood (Anthony, 2018a; Scott & Tribble, 2006) and the size of the differences is calculated by effect size (Gabrielatos & Marchi, 2012).

***Dependent variable***
The fulfilment of the initial intention (IF) was measured by 4 items on 7-point Likert scale ranging from 1 'totally don't agree' to 7 'strongly agree' (e.g. 'I achieved my personal learning goals by participating in this MOOC', 'the MOOC met my expectations'; Cronbach's alpha = .89). The participants were split into two groups according to their post-course IF level divided by the sample median (*med*=4.75). Two hundred and twenty participants had been identified as high-IF and 242 participants had been identified as a low-IF. Participants that carried out less than four activities were not included in the sample, leaving a total of 445 participants – 214 with high-IF and 231 with low-IF. Due to the anonymization process, no demographic information was available about the participants.

## Results

In the following section, we first present the differences between the two groups in total activities per participant – high and low IF. We then present the learning sequences findings using the n-gram and keyness measurements.

Table 2 shows the descriptive statistics of the number of activities per participant in each group. In total, 61,713 activities were analyzed (high-IF = 35,790; low-IF = 25,973). The non-parametric Mann-Whitney U test indicated that the number of activities per participant was significantly higher for the high-IF group compare to the low-IF group ($U$ = 17223.5, $p$ < .001). In order to check if there are differences between the two groups in their level of heterogeneity, we checked whether the standard deviations in the number of activities are significantly different between the low and the high IF groups. Levene's test of the homogeneity of group variances showed significant difference ($F_{(1,443)}$ = 1.46, $p$ < .05). Although on average the number of activities in the high-IF is higher compared to the low-IF group, the standard deviation of the number of activities and the maximum activities per participant are both higher in the low-IF group compared to the high-IF group (see table 2).

**Table 2.** Descriptive statistics of the number of activities per participant and the activity frequencies in the high and low IF groups

| | Low-IF group | High-IF group |
|---|---|---|
| Num. of participants | 231 | 214 |
| Mean num. of activities | 112.44 | 167.24 |
| Mean rank of activities | 190.56 | 258.02 |
| Median num. of activities | 50.00 | 122.50 |
| S.D. of activities | 192.35 | 159.16 |
| Maximum activities | 1776 | 857 |
| V | 12,426 (47.84%) | 19,344 (54.05%) |
| T | 4,255 (16.38%) | 5,127 (14.33%) |
| Q | 3,170 (12.20%) | 3,535 (9.88%) |
| P | 2,222 (8.56%) | 2,795 (7.81%) |
| I | 1,687 (6.50%) | 1,911 (5.34%) |
| L | 1,276 (4.91%) | 1,640 (4.58%) |
| E | 567 (2.18%) | 857 (2.39%) |
| S | 305 (1.17%) | 493 (1.38%) |
| R | 53 (0.02%) | 70 (0.20%) |
| B | 12 (0.05%) | 18 (0.05%) |

### *N-gram analysis*

In order to identify the learning sequences of the two groups, we used n-gram analysis to compare sequences of activities (activities' relative frequency analysis) and their distribution among the participants (range analysis). The two analyses are complementary to each other. While the activities' relative frequency analysis answers the question of what is the relative prevalence of an activity or sequence of activities in a specific group of participants, the range analysis answers the question, what is the percentage of participants that participated in an activity or sequence of activities?

The number of the unique tokens in the unigram analysis is 10 (representing the 10 codes of activities), the bigrams – 95, the trigrams – 682 and the four-grams – 3,134.

Figures 1a-d present the results of the activities' relative frequency n-gram analysis and Figures 2a-d present the results of the range n-gram analysis. In both cases, only activities with probability above 0.1 were included.

Figure 1a presents the comparison of the unique unigrams in both groups (the figure represents the information in Table 2). The video activity (V) is more salient in the high-

IF group compared to the low-IF one. On the other hand, the track (T), lessons (L), quiz (Q) and exam (E) activities have higher occurrences in the low-IF group compared to the high-IF group.

Figure 1b presents a difference in the V-V bigram between the low-IF and high-IF groups that is larger than the differences in the other bigrams. The participants in the high-IF group sequentially press the video play/pause button more than the participants in the low-IF group. Interestingly, five of the bigrams (Q-Q, P-Q, S-L, V-L, and T-Q) are unique to the low-IF group.

Figure 1c presents the trigrams activities that show a similar pattern to the bigrams, with more participants in the high-IF group that sequentially press the play/pause button video (V-V-V). While looking at the sequences that are unique to one of the groups, it can be seen that in the low-IF group, there is a unique sequence of practicing the final exam (E-E-E), a sequence that does not exist in the high-IF group.

The four-gram figure (Figure 1d) presents a prominent presence of the high-IF group compared to a minor presence of the low-IF group. The participants in the high-IF group made more four-gram sequences of video watching (V-V-V-V), and sequences of video watching after watching the recommended learning track (T-V-V-V), accessing the lessons (L-V-V-V), answering a quiz (Q-V-V-V) accessing the reading comprehension text (P-V-V-V), self-practicing the final exam (E-V-V-V), etc.

The results of the range n-gram analysis show similar trends. The range shows the percentage of participants who actually did each activity (or sequence of activities) out of the overall activities (or sequence of activities) in each group. The calculation of the range enables us to calculate the relative distribution (entropy) of each activity. Figure 2a shows that, in the high-IF group, four activities have been performed by above 80% of participants, while in the low-IF group only two activities were carried out by 80% or more of participants. Two activities in the high-IF group were performed by 50% to 79% of the participants compared to five activities in this range of participation in the low-IF group. In both groups, the three activities - S, R, and B - were carried out by less than 40%. A higher percentage of participants in the high-IF group pressed the play/pause video button (V), accessed the quizzes (Q), accessed the reading comprehension PDF text (P), accessed the introductory page of the course (I), and accessed to the video lessons dealing with learning strategies (S). No differences were found between the two groups in the range of participants who accessed the recommended learning track (T), the self-practice exam (E), the right of use (R), and the achievements page (B).

The differences in the range parameters between the two groups increase when we look at the bi-, tri- and four-grams (Fig. 2b-d). This is evident by the fact that the longer the n-gram, the higher the participation range in the high-IF group compared to the low-IF group. The low-IF participants, on the other hand, performed five unique bi-gram sequences, one unique tri-gram sequence, and no unique four-gram sequence of activities. The decrease in unique sequences and the fact that we only analyzed n-grams with relatively high probability (>0.1), means that the low-IF participants use more varied sequences by less and less participants. This also means that in the range parameter, the high-IF group behaves more consistently and that more participants behave similarly (lower entropy).

***Keyness results***

Video play/pause activity (V) was identified as a key activity in the high-IF group compared to the low-IF group. Participants in the high-IF group pressed the play/pause video (V) button 1.28 more than the participants in the low-IF group ($log(.25) = 232.11$, $p < .001$, *Effect Size* = 1.28).

In the low-IF group, we found that lessons (L), track (T), exam (E) and quiz (Q) activities are key activities compared to the high-IF group. Participants at the low-IF accessed to more lessons ($log(.25) = 84.28$, $p < .001$, *Effect Size* = 1.27), followed more recommended learning track ($log(.25) = 49.44$, $p < .001$, *Effect Size* = 1.71), accessed more exams ($log(.25) = 36.64$, $p < .001$, *Effect Size* = 1.23) and participated in more quizzes ($log(.25) = 11.21$, $p < .001$, *Effect Size* = 1.10) compared to the high-IF group. These results are reflected in the relative frequency unigram analysis mentioned above.

## Discussion

The purpose of the current study was to compare behavioral patterns and learning sequences between participants with high and low IF in a MOOC. The comparison was conducted in order to identify behavioral differences between activities and activity sequences of these two groups using NLP techniques, namely n-gram and keyness.

In order to achieve those aims, we compared the differences in the relative frequencies of learning behavior sequences and in the participation range (participation entropy) by using n-gram analyses and keyness analysis.

As might be expected, participants with high-IF are more active in the course compared to participants with low-IF. Furthermore, the unigram analysis and the keyness analysis revealed that participants in the high-IF group pressed the play/pause video button more often than the participants in the low-IF group did. On the other hand, participants in the low-IF group more frequently accessed lessons, recommended learning tracks, and took exams and quizzes. These results suggest that the participants in the high-IF group were more focused on acquiring knowledge, as evidenced by watching the video lectures, which contained the course content. On the other hand, the participants in the low-IF group showed a more diverse and less orderly ("messy") learning behavior. Our interpretation of these patterns is that the participants in the low-IF group were less sure what to do in the course. They spent more attention on understanding what and how to learn, and on quizzes and final exams, and less on knowledge acquisition. These results are similar to the results by Mukala, Buijs, & Van Der Aalst (2015), who showed that students who passed a Coursera MOOC followed a more structured process in submitting their weekly quizzes until the final quiz and in watching video, when compared to students who did not pass the course. It is important to note that our conceptual replication of the results uses a broader perspective about success and failure in MOOCs. We see that the activities of the participants in the current MOOC can predict more subjective success outcomes, namely intention-fulfilment.

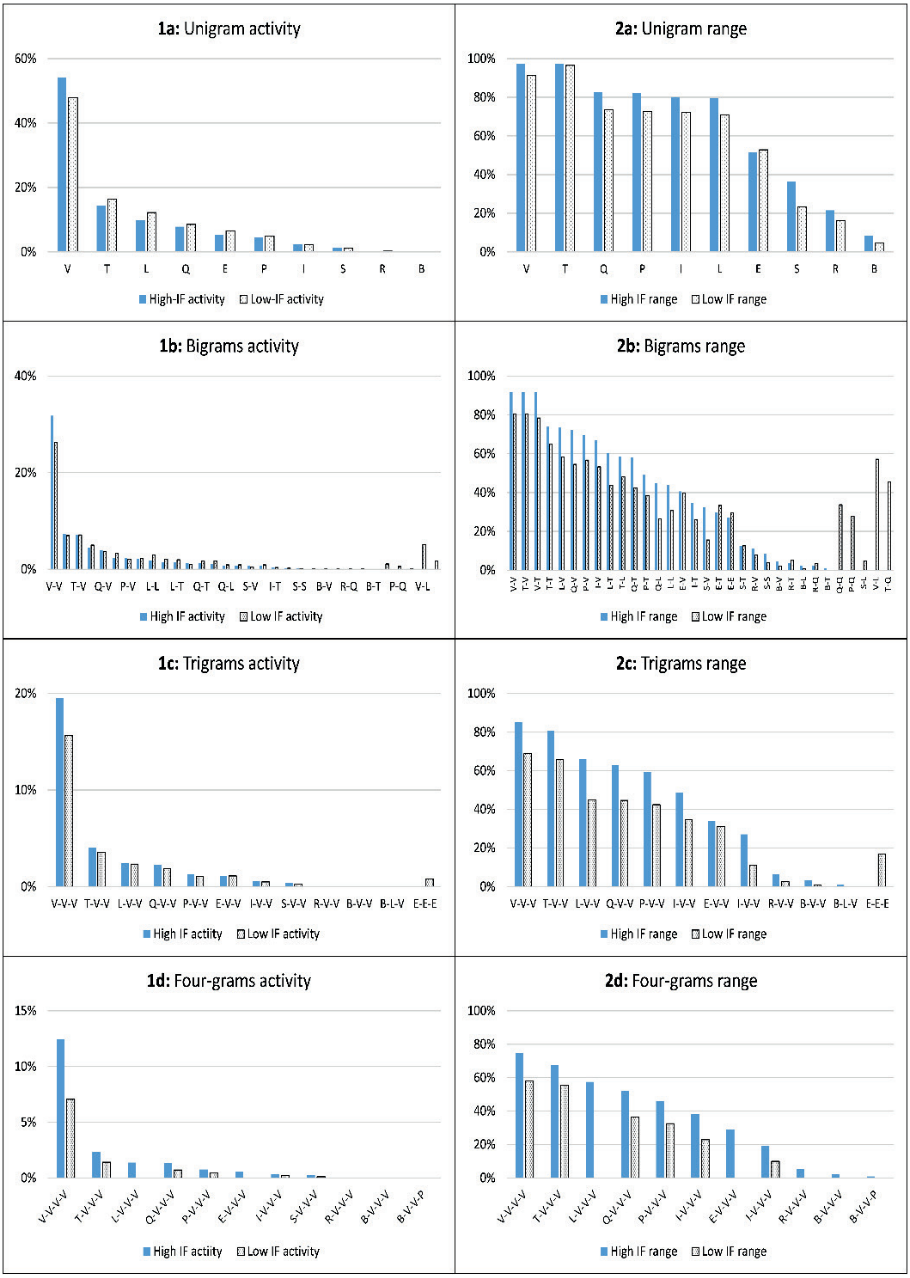



**Figure 1a-d and 2a-d.** Relative frequency of activities and relative range distribution among the two groups in uni- bi- tri- and four- grams

The n-gram analysis enabled us to compare the most probable sequences of activities and their distribution among the participants. Although Li, et, al. (2017), showed that the most effective n-gram for predicting students' activity in MOOCs is the trigram, our analysis suggests that we can differentiate between the groups even with a shorter string of annotation, meaning a bi-gram. The bigram analysis reveals that the high-IF group was characterized mostly by a two-step sequence of the knowledge acquisition activity of watching video lectures sequentially (V-V), while the low-IF group was characterized by diverse bigram activities such as repeating the assessment tasks (Q-Q), moving from the reading comprehension to the quizzes without watching the video lecture (P-Q), moving from the short and focused videos dealing with learning strategies to the lesson (S-L), moving from the video lecture to the lesson (V-L), and moving from the recommended learning track to the quizzes (T-Q). These results are similar to the findings of Van den Beemt, et al. (2018) who used other success criteria such as passing rates. The researchers showed that regularly watching successive videos in batches leads to high passing rates.

Nevertheless, for the two parameters – activity frequency and participation range – we found that looking at longer n-gram sequences is beneficial in predicting the level of IF. The longer the n-gram, the higher the divergence between the two groups. Moreover, the longer the n-gram, the more prominent are the participants from high-IF group. The results showed that the activities of the high-IF group are more predictable, suggesting that this group behaves more consistently and similarly. When we analyze longer sequences, it is clearer that the participants in the high-IF group are following the designed path, i.e. the learning path suggested by the course designers in this particular MOOC.

Several limitations should be considered. First, we used median splits in order to distinguish between participants with high and low IF. This technique helped us to simplify our analyses and discussion. Recording continuous variables into categorical variables is often criticized due to the rough segmentation of the continuous variable (DeCoster et al., 2011), but this simplification was useful in our case. The results showed that we could easily differentiate between, and predict the learning sequences of the different participants. Future work could use a more sensitive segmentation and a larger amount of clusters. Another simplification that was used in this research is the use of only one learner-centered success measure, namely IF. Future research should use additional subjective success measures such as learner satisfaction (Rabin et al., 2019) and perceived achievement (Rabin et al., 2019; Ross, 2006; Yoon et al., 2018).

Future research could also look at additional kinds of knowledge acquisition with video lectures. The MOOC studied here offered two kinds of video lectures – content-based lectures (V) and learning strategy lectures (S). As shown in Figures 2a and 2b, in the high-IF group, a wider range of participants accessed the learning strategy videos (S) and learning strategy videos following by video lectures watching (S-V) compared to the low-IF group. Further investigation of the effect of using those learning strategy lectures on the level of IF is outside the scope of this study, but could be productive.

## Conclusions

To conclude, the purpose of the current research was to distinguish between the low and the high IF groups based on their learning behaviors. The results suggest that the single activity and sequential behavior of the participants enable us to identify their affiliation group. As has been shown by the keyness analysis, the two groups are different in the pattern of single activities, and bigger differences become apparent in the longer n-grams, both in terms of the relative prevalence of the activity and in terms of the number of participants who performed it. The high-IF group showed more homogeneous behavior. One of the contributions of our study is the feasibility of developing automatic intervention systems, which will analyze learning sequences in real time and identify inconsistent participant behavior, to support the participants in real time. For example, such system could propose a different learning track for learners, depending on their behavioural pattern. Alternatively, learning strategies could be proposed for specific sub-groups supporting their self-regulated learning.

7

# Chapter 7

## General Discussion

The central research question that this dissertation aimed to answer was: How to evaluate learner-centered outcomes and their antecedents in open online education? To address this question, two learner-centered outcomes, namely, learner satisfaction and learner intention-fulfillment were identified as alternative measures of course outcomes. These alternative measures, as this study has demonstrated, are more appropriate for measuring success in the unique context of non-formal lifelong learning that characterizes MOOCs, in contrast to outcome measures such as grades, retention, and completion rates which previous studies viewed as the ultimate outcome measures. As shown throughout this dissertation, the use of different learning analytics approaches can reveal the nature of these two learner-centered outcomes, namely, learner satisfaction and learner intention-fulfillment and their antecedents.

To guide the research project, five studies were conducted. These five studies can be summarized from a theoretical perspective, as well as from a methodological perspective. Broader discussion about the theoretical and methodological implications of these studies will be presented in the "Implications for researchers and practitioners" section of this chapter.

From the theoretical perspective, these studies investigated, all together and each one separately, the theoretical construct "learner-centered outcomes". The studies empirically define this theoretical construct and reveal its antecedents and the barriers to reaching these outcomes. The construct learner-centered outcomes is defined as success measures that are subjectively assessed by the learner based on his or her impression of the course. These learner-centered outcomes are contrasted with learning outcomes that are pre-defined by the course designers and instructors. All five studies demonstrated the usefulness – from both the organizational perspective and the psycho-pedagogical perspective – of using learner satisfaction and learner intention-fulfillment among learners who use open educational resources (OER) such as in the case of MOOCs.

From the methodological perspective, the studies applied several methods to investigate the nature of learner-centered outcomes and their antecedents. All these methods used learning analytics as the methodological framework to measure, collect, analyze and report data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs (Siemens et al., 2011). The studies used different statistical tools such as correlations, linear regressions, structural equation models (SEM), user behavior pattern detection (UBPD), and natural language process (NLP) techniques.

The first study (Ch 2.) "The cathedral's ivory tower and the open education bazaar - catalyzing innovation in the higher education sector" focused on the theoretical framework that laid the foundations for the dissertation. This study explored the effect of digital innovation on higher education (HE) sectors and compared the business model of traditional universities with the business model of open education. Using Raymond's metaphor (1999; 2001), these changes can be seen as two opposite ends on a continuum: On one end, the 'cathedrals': traditional higher education institutions (HEI), such as the campus-based university, and on the opposite end, 'bazaars': a host of innovators in

HE, such as publishers and users of open educational resources. Several authors argue that MOOCs, as a disruptive innovation, will replace the traditional HE system by offering education for free (Rifkin, 2014), but our thesis is that open education will not replace the traditional HEI, but rather create a reciprocal relationship between the cathedrals and the bazaars. It is argued that, for example, traditional HEIs will be able to use the infrastructure created by the MOOC providers, while open education will be able to use the knowledge and expertise that has been developed by faculty members in HEIs as content developers. Additional proof of the validity of this model was recently demonstrated during the COVID-19 pandemic when universities used courses provide by the MOOC platform as a substitute for campus-based teaching (Ma & Mendez, 2020).

Despite the potential of MOOCs, some critics have suggested that MOOCs may have failed to achieve their promises mainly due to low success rates among students who start MOOCs. This critique is based mainly on misconceptions about the business models of the Bazaar-type institutions: The comparison between the business models showed that the customer value proposition (CVP) of MOOC providers is very different from the CVP of traditional universities. Following the differences in CVP between these two types of education providers, our central recommendation in this first study is that criteria for success in MOOCs should be learner satisfaction oriented and should reflect the extent to which the MOOC allowed participants to fulfill their initial intentions. This recommendation reinforces findings by researchers such as Henderikx et al. (2017), Kalz (2015), and Reich (2014).

The second study (Ch. 3) "An empirical investigation of the antecedents of learner-centered outcome measures in MOOCs" aimed to empirically characterize the two learner-centered outcomes, namely learner satisfaction, and learner intention-fulfillment, and identify their antecedents. This study used educational data mining and learning analytics techniques to understand how participants' demographics, their pre-course characteristics when entering the course, their actual behavior in the course, and their perceived course usability predict the two learner-centered outcome variables. Despite the relatively high correlations between the two learner-centered outcomes, learner satisfaction, and learner intention-fulfillment, the results of the second study showed two distinct pathways through which the participants achieved these outcomes.

The analysis in the second study (Ch. 3) revealed that several factors directly affected learner satisfaction in the model developed in the study, including the perception of the importance of the MOOC's benefits, level of online self-regulated learning - goal setting, number of video lectures accessed, and perceived course usability. Factors that indirectly affected learning satisfaction included age and the number of quizzes accessed. Those variables indirectly affected learner satisfaction through perceived course usability and the number of video lectures accessed. Intention-fulfillment was directly affected by gender, the importance of the MOOC's benefits, online self-regulated learning - goal setting, the number of quizzes accessed, the duration of participation, and perceived course usability. Previous experience with MOOCs and the importance of MOOC's benefits indirectly affected intention-fulfillment through the number of quizzes accessed and perceived course usability.

The findings in this study shed new light on the role of the demographic variables on learners-centered outcomes. Gender affected only the intention-fulfillment variable. Female learners had a higher level of intention-fulfillment than male learners. On the other hand, gender did not predict the level of satisfaction. Further research should explore whether these results can be generalized beyond the specific context of this MOOC.

In this study, age was not a direct predictor of course outcomes, but rather predicted a behavioral variable that reflects progress in the course, i.e. the number of video lectures that the participants accessed during the course, which in turn predicted learner satisfaction. In contrast, the age of the participant did not predict, directly or indirectly, the level of intention-fulfillment.

The level of importance of the benefits of participating in the MOOC predicted both of the learner-centered outcome variables. This had a direct positive influence on both satisfaction and intention-fulfillment, as well as an indirect positive influence on the number of quizzes taken, which in turn influenced intention-fulfillment directly and satisfaction indirectly. The MOOC did not provide any credit beyond a certificate of completion, and we can thus see how lifelong learners who assign a higher value to the advantages provided by the MOOC, are more likely to invest more in the course and to achieve positive outcomes. An applied implication of this finding is the importance of clearly delineating the MOOC's benefits and contributions in a way that allows participants to evaluate the relevance of the MOOC for their personal goals.

An additional antecedent identified in this study was goal setting, which had a strong positive impact on course outcomes. As Zimmerman (2002) mentioned, the ability to set learning goals is an internal structure that is based on learner abilities and can be learned throughout one's life. Interestingly, our findings did not identify that the correlation between goal setting the learners-centered outcomes, learner satisfaction and intention fulfillment, were mediated by any of the behavioral variables.

Another thought-provoking finding of this study is the difference between the behavioral variables that influenced learner satisfaction and those that influenced intention-fulfillment. The number of video lectures accessed positively predicted learner satisfaction, while the level of intention-fulfillment was directly predicted by the number of weekly quizzes accessed, and by course duration. Accessing video lectures is a passive learning behavior while taking self-assessment quizzes, and to a lesser extent persisting in the course, are more active aspects of learner behavior. A possible insight is that more active course components, such as self-assessment quizzes that provide participants with feedback on their achievements and understanding, assist learners who are focused not only on enjoying the course (i.e. learner satisfaction) but also on using the course to fulfill the personal intentions they had when they set out to study the MOOC (intention-fulfillment).

The final predictor identified in the second study was perceived course usability, which was a strong predictor of both course outcomes. This finding reflects the fact that a course with poor usability will delay the learner's progress, and decrease the personal benefits from participating in it (Eom et al., 2006). The only direct predictor of perceived course usability was the number of quizzes taken, which, as discussed in the previous paragraph, is also an important predictor of the key outcome variables.

Identifying the role of those predictors on the barriers to satisfaction from the learning process was the focus of the third study (Ch. 4) "What are the barriers to learners' satisfaction in MOOCs and what predicts them? The role of age, intention, self-regulation, self-efficacy and motivation". This study identified the role of age, intention, self-regulation, self-efficacy, and motivation as barriers to satisfaction faced by many participants in MOOCs as well as their predictors. The study calculated a general barrier score and identified three kinds of barriers: (1) lack of interestingness/relevance, (2) lack of time/bad planning, and (3) lack of knowledge/technical problems. The barrier 'lack of interestingness/relevance' was predicted by the self-regulation indices of self-evaluation, study-strategy, and help-seeking. The second barrier, 'lack of time/bad planning', was predicted by the self-regulation indices of goal setting, time management, and study strategy as well as by the age of the respondent. The third barrier, 'lack of knowledge/technical problem' was predicted by the level of self-efficacy, extrinsic motivation, and the self-regulation index of time management, as well as by the behavioral intention to complete the course. The index averaging the extent of the barriers was predicted by the self-regulation indices of goal setting and study strategy, the level of self-efficacy, and the level of extrinsic motivation.

Both together and separately, the results of studies 2 and 3 (Ch. 3 & 4) illuminate the role of the self-regulation index – goal setting, on the level of learners' satisfaction and their barriers to satisfaction. As can be seen from study 2 (Ch. 3), the higher the level of goal setting, the higher the level of satisfaction from participation in the course, and there is a lower level of the general barriers to satisfaction. These results support previous studies that found a positive correlation between SRL and satisfaction in online courses (Artino, 2007; Kizilcec, Perez-Sanagustín, & Maldonado, 2017; Li, 2019; Puzziferro, 2008). It is interesting to note that the self-regulation index – environmental setting- did not enter the prediction model of the level of satisfaction nor did this affect the barriers to satisfaction. The effects of the other SRL indices – self-evaluation, help-seeking, study strategy, and time management were not under investigation in study 2 (Ch. 3) and therefore cannot be compared.

In study 2 (Ch. 3), age indirectly affected the level of satisfaction, through the number of videos accessed. In Study 3 (Ch. 4), age predicted only the second barrier, 'lack of time/bad planning'. Using the 'perspective of life' stages theory (Stoffelsen & Diehl, 2007), Henderikx, Kreijns, Muñoz, and Kalz (2019) showed that learners in their early adulthood (20-35 years) and mid-life (36-50 years) most often faced external barriers such as family and work issues. Those results add another layer to our understanding of the connection between age and learning outcomes in the context of open education.

Future studies should explore the effect of the 'importance of MOOC's benefits' and 'perceived course usability', two variables that had been identified as predictors of satisfaction in study 2 (Ch. 3), and their effects on the barriers to satisfaction. Since study 3 (Ch. 4) was based only on self-report questionnaires, future research should also look at the effect of learning behavior on the barriers to satisfaction.

Studies 4 and 5 (Ch. 5 & 6), which will be summarized in the next paragraphs, revealed novel methods to investigate learning behavior. The methods that were described in these two studies can be used to deepen our understanding of the nature of learning satisfaction and the barriers to reach it.

Earlier studies demonstrated that clustering participants according to their learning trajectories is more informative and has a higher potential for pedagogical improvements than clusters that are based on static counts behavioral data (Kizilcec et al., 2013). Similarly, in the fourth and fifth studies (Ch. 5 & 6) we have shown novel methods to cluster participants according to their learning trajectories.

The fourth study (Ch. 5) "User behavior pattern detection in unstructured processes – a learning management system case study" proposed and validated a user behavior pattern detection (UBPD) methodology which is based on detecting very short user activity sequences and clustering them based on shared variance to construct a more meaningful behavior pattern. The UBPD system identified personal learning chains (sequences) of activities in the unstructured processes of learning a MOOC. Borrowing a term from genetics research, where sequence mining is commonly used, the personal learning chains are named 'motifs' and define as "recurring patterns that appear in a network more frequently than expected in a random network" (Alon, 2007; Milo et al., 2002). In this fourth study, the use of UBPD had been demonstrated by using two datasets mined from learning management systems (LMS), the first dataset had been mined from a gamified course and the second dataset had been mined from a MOOC. Here we focus our discussion on the second case study since MOOCs are at the center of this dissertation.

In the investigated MOOC, nine behavioral patterns were identified. For example, one pattern represents users who view each week's lectures in sequential order as was planned by the course designers. Another pattern represents users who complete all the weekly quizzes, and other patterns represent users who watched the first videos of the course one at a time and not sequentially. The different behavioral patterns that had been identified suggest that participants choose unique learning paths through the course. The fourth study (Ch. 5) revealed the different behavioral learning patterns but did not aim to predict learner-centered outcomes. The fifth study (Ch. 6) used different sequential method to investigate the connection between behavioral learning patterns and the learner-centered outcomes, intention-fulfillment.

The fifth study (Ch. 6) "Identifying learning activity sequences that are associated with high intention-fulfillment in MOOCs" used a different analytic approach to cluster participants based on their learning trajectories. This study identified the influence of the activities and activity sequences of participants on their level of intention-fulfillment. Adapted from natural language processing (NLP) we used n-gram analysis to identify learning activity sequences and to perform keyness analysis to identify prominent learning activities. Using those techniques enabled us to identify the differences between participants who fulfilled the initial intentions (high-IF) and those who did not fulfill their initial intentions (low-IF). Differences were seen at the level of single activities, but major differences were found when longer n-grams were used. The high-IF group showed more consistency and less divergent learning behavior. Although on average the number of activities in the high-IF is higher compared to the low-IF group, the standard deviation of the number of activities and the maximum activities per participant are both higher in the low-IF group compared to the high-IF group. Among the high-IF group, video activity is more salient compared to the low-IF one. On the other hand, the track, lessons, quiz, and exam activities have higher occurrences in the low-IF group compared to the high-IF group.

It is interesting to note that although in the fifth study (Ch. 6), video watching was correlated with High-IF, in Study 2 (Ch. 3), the number of videos watched had correlated with the level of IF, but did not play a significant role in predicting the level of IF in the structural equation model. These results suggest that while looking at unique activities without taking into account the sequence of activities we should investigate other predictors such as the demographics and psycho-didactic characteristics of the participants. However, while investigating the learning sequences, High-IF was associated, among other things, with study patterns of sequentially watching video lectures. The longer the sequence of the activity the more accurate the IF prediction. Future studies should investigate whether these effects are valid while taking the participants' characteristics into account.

## Implications for researchers and practitioners

As stated at the beginning of the chapter, this dissertation aimed to answer the main question: How to evaluate learner-centered outcomes and their antecedents in open online education? While answering this question, several theoretical and practical issues emerged.

From the theoretical perspective, the first study (Ch. 2) offers a novel framework to analyze the different business models of organizations in the HE system. The model that was offered, which looks at the continuum between "cathedrals" and "bazaars" in the HE ecosystem, can help researchers and practitioners to develop organizational theories about the use of OER in the HE system, to understand the reciprocal relationship between the different actors in this sector and other sectors and to predict the future development of the use of OER in the mainstream educational systems. The use of this theoretical model enabled us to define and explain the need of the educational system and its stakeholders to use a different approach to learning outcomes. The approach looks at the predisposition of the participants and their intentions while participating in the course and focus on learning outcomes that had been defined by the participant.

In the second study (Ch. 3) we theoretically defined two learner-centered outcomes, namely, learner satisfaction and learner intention-fulfillments, and revealed practical ways to measure these variables. Practitioners can use the study's findings in three ways: Firstly, by helping participants to identify the importance of the MOOC's benefits and by encouraging participants to set their own goals for the learning period. Secondly, the results suggest that the level of usability of the course plays an important role in satisfying the learners' outcomes. Course designers should pay more intention to the usability and ease of use of the courses. Thirdly, using aggregate learning-behavior patterns, as was measured by the number of video lectures accessed, the number of quizzes accessed, and the duration of participation can help to develop real-time intervention using artificial intelligence systems that will be able to monitor and encourage participants to reach their initial intentions and to be satisfied from the participation at the course.

The third study (Ch. 4) focused on the barriers to satisfaction which participants in MOOCs are facing, and their predictors. Course designers should pay more attention to helping participants to develop their SRL, their feeling of self-efficacy, and to take into account the individual orientation in motivation (intrinsic versus extrinsic) and their behavioral intentions. In the third study (Ch. 4), as was demonstrated in the second study (Ch. 3), the importance of identifying learners' predisposition and paying closer intention to those characteristics can help participants to become more satisfied with their participation in the course and to encounter fewer barriers.

The fourth study (Ch. 5) showed that we can cluster participants according to short learning sequences. This study can help researches and course designers to classify different groups of participants, demonstrating that different participants in open distance education have different needs and different sequential learning patterns. Using the algorithm developed in this study helps recognize these different clusters and can help develop a personalized teaching approach, tailor-made for specific participant groups while developing different sets of learning tools for different individuals. In addition, the algorithm developed in this study helps to develop the theory regarding learning processes in online, distance, open education by showing to what extent valuable insights can be obtained from investigating short learning sequences in online courses.

The fifth study (Ch. 6) used novel methods from the NLP domain to predict the level of IF. The results of this study suggest that aggregating learning-behavior can predict the learner-centered outcome – IF. Also, the study showed that examining sequences of learning behaviors is beneficial in predicting learner-centered outcomes. From a practical perspective, researchers can benefit from deepening their use of tools from the NLP domain and other domains in an educational setting. From the theoretical perspective, this study shows the nature of the connection between behavioral process learning and learner-center outcomes.

## Limitations and suggestions for future research

From the above discussion, five limitations emerge, which may help to inform future research. As we revealed in the first study (Ch. 2), using instructor-focused learning outcomes in distance open education settings is based on a misunderstanding and misinterpretation of the different business models, the cathedral and the bazaar, which form two ends of the continuum discussed above. In open distance educational platforms, which are becoming more and more popular as a way of acquiring knowledge for lifelong learners, learning success can be measured more objectively by the learner rather than by the course instructor. However, the bazaar approach reveals several methodological challenges, such as how to measure learners' initial intentions and how to measure the level of intention fulfillment and the level of satisfaction. This model also raises questions about how to create tailor-made personalized interventions. Future research should delve deeper into the nature of learner-center outcomes, to develop more valid ways to measure these outcomes, and to develop interventions that will help people to define and implement their subjective outcomes. It is important to note that interventions should take in cautions since only small benefits had been found to scalable online interventions depending on individual and contextual characteristics (Kizilcec et al., 2020). As well, researches should take into consideration ethical issues, as will be discussed in the limitation section.

In the second and third studies (Ch. 3 & 4) many predisposition characteristics of the participants were taken into account (e.g. gender, age, SRL, motivation, initial intentions, and outcome beliefs). The results of these two studies suggest that these variables affected the learner-centered outcomes, namely IF and satisfaction, and the barriers to satisfaction in a range of different ways. However, it seems that additional predictors should also be examined: For example, factors such as information and communication technology (ICT) skills and educational background should also be examined as predictors of the learning behavior and learning outcomes. In our studies, these measures showed insufficient variability and could not be included in the analysis, but in future research, they should be taken into consideration since the ability to use ICT and the educational background plays a major role in the ability of the participants to use MOOCs (Hansen & Reich, 2015; Jisoo, Ahreum, & Junseok, 2018; Kizilcec et al., 2020).

The use of pre- and post- questionnaire in the second, third, and fifth studies (chapters 3, 4, & 6) raises the shortcomings of a non-response bias. The participants in our studies were a unique sub-group of participants who chose to answer the research questionnaires, and not a random sample of the MOOCs' participants. This limitation is typical for MOOC studies that use self-reported questionnaires (Breslow et al., 2013; Kizilcec & Halawa, 2015). To overcome this bias, we compared the sample's demographic characteristics, in the second study (Ch. 3), with the demographics of the course's population, and no significant differences were identified. Future research should encourage participants to answer the questionnaires by explaining the importance of the information for the course designers and researchers as well by providing incentives for the participants who answer the questionnaire. Additionally, research can use the data-mining approach and use data from all the course participants as was done in the fourth study (Ch. 5). Future research should develop non-responsive methods to investigate the antecedents of the two dependent variables - learner satisfaction and learner intention-fulfillment. Moreover, since the MOOCs that had been analyzed were in Hebrew, only Hebrew-speaking participants were able to participate in this study, and these limitations may reduce the external validity of the results. Future cross-culture and cross MOOCs platforms research should be done to improve the generalizability of the results.

The dependent variables learner satisfaction, learner intention-fulfillment, and the barriers for satisfaction were collected using self-reporting tools, which can be biased due to social desirability bias and prestige bias. Participants might bias their answers to be perceived by society as more successful than they are. Triangulation with additional sources such as interviews or behavioral indices could be utilized for future research. For example, a mixed-method research set-up would be appropriate to further explore and gain a deeper understanding of quantitative self-report results (Morse, 2016).

In the fourth and the fifth studies (Ch. 5 & 6), we focused our attention on the effect of sequential learning patterns in unstructured MOOCs on the prediction of learner-centered outcomes, IF. Up to now, most of the research that tried to identify the antecedents of MOOCs learners' IF and the level of satisfaction used aggregated data such as the number of video lectures that the participant accessed. However, detecting learning processes can be very useful: Learners can see their learning process compared to others, which can be further used to modify or enhance certain behaviors. Teachers can use these data to assist specific learners and to adapt their teaching styles. System designers can use these data to redesign or improve learning environments and to identify bottlenecks, and lastly, adaptive systems can automatically modify themselves based on actual usage data to encourage required changes in learning behaviors.

Although we found that focusing on short sequences of two to five steps is highly informative, more advanced process mining techniques may be used in the future to identify longer sequences that cover all the learning processes in a specific course up to learning process in several courses up to complete the learning program as defined by the learner.

In this dissertation, we used learning analytics (LA) as an analytical approach. MOOCs have specific characteristics that make them excellent candidates for LA: They typically include many participants, have detailed log files, a good diversity of participants, and a process that is loosely defined (Kizilcec et al., 2013). Nowadays, HEIs have started using LA to understand student behaviors and to improve instructional, curricular, and learner support resources, as well as improve learning environments (Rubel & Jones, 2016).

Despite these many advantages, the use of LA is morally complex and raises some ethical questions, especially regarding student privacy. Since LA often relies on aggregating significant amounts of sensitive and personal student data from a complex network of information flows, this may harm the right for privacy (Jones, 2019). Another concern is raised by Heath (2014) who claims that students' consent to provide their demographic data in the student application, admission, and administration context does not necessarily apply in any other context. Specifically, they do not necessarily agree to the flow of their private information to secondary uses of data for LA activities. As Vialardi, Bravo, Shafti, and Ortigosa (2009) cautioned predictive models that derive data stored within institutional information systems can be directly associated with an individual student identity. These ethical concerns should take into account in all the stage of the LA usage, but has to serve as landmarks and not as 'stop signs'.

Although all five studies we conducted were approved by the university's Ethics Committee and the participants agreed that the data that had been collected about them will be used for academic research, it is important to take into consideration that not all the participants are aware of the uses and possible misuses of their data. As predictive models, machine learning approaches, and artificial intelligence become more and more popular in creating a personalized learning experience, researchers and policymakers should exercise more caution in the use of the huge amounts of data that participants produce during their learning processes to avoid possible abuses of the information gathered from the data that were collected and analyzed.

## Conclusions

The Covid-19 pandemic accelerated digital changes around the world. Online learning becomes mainstream in higher educational institutions, the K-12 system, and in the organizational world. One of the popular formats for online learning is Massive Open Online Courses (MOOCs), which have taken higher education by storm and demonstrate the potential of scaling higher education with means of digital media and the internet.

The rapid changes in the educational system validate the cathedral and the bazaar model that had been suggested in this dissertation. The HE ecosystem showed synergistic relationships between organizations that represent many gradations on the continuum between "cathedral-type" and bazaar-type" organizations. More and more HEIs are using open materials and practices that have been developed in the bazaar while, on the other continuum, bazaars' type organizations showed highly dependency on the resources that originate from cathedrals' type organizations.

The studies presented in this dissertation have, individually and collectively, turned a spotlight on the importance of looking at learner-centered outcomes in open, distance, and online learning and suggest a novel perspective to analyze learners-centered outcomes and success in open distance education forms, such as MOOCs. Although MOOCs are niche activities in HEIs, the lessons that were learned can and should affect the educational system in the knowledge era, moreover so during the Covid-19 pandemic.

Learners are individuals who would like to acquire information, knowledge, and skills in a way that fits their needs, the expected benefits from participation, their socio-demographic attributes, and their psycho-pedagogical characteristics. The educational system, policymakers, and society as a whole should help lifelong learners to learn how to define their goals and regulate their learning process. Those efforts can and should be combined with the emerging support of personalized and artificial intelligent systems. Those steps will enable personal and social development by enabling people all over the world to increase their abilities and skills.

## References


Abelson, H. (2008). The creation of OpenCourseWare at MIT. *Journal of Science Education and Technology*, *17*(2), 164-174. https://doi.org/10.1007/s10956-00-9060-8

Agrawal, R., & Srikant, R. (1994). *Fast algorithms for mining association rules.* Proc. 20th int. conf. very large data bases, VLDB.

Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013). *Steering user behavior with badges*. 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil.

Anderson, C. (2009). *Free: the past and future of a radical price*. Hyperion.

Anthony, L. (2018a). *AntConc* (3.5.7). Waseda University. http://www.laurenceanthony.net/software

Anthony, L. (2018b). *AntConc Help (manual)*. http://www.laurenceanthony.net/software/antconc/releases/AntConc357/help.pdf

Antin, J., & Churchill, E. F. (2011). *Badges in social media: A social psychological perspective*. CHI 2011 Gamification Workshop, Vancouver, BC, Canada.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, *16*(1), 40. https://doi.org/10.1186/s41239-019-0172-z

Ali, A., & Ahmad, I. (2011). Key factors for determining students' satisfaction in distance learning courses: A study of Allama Iqbal Open University. *Contemporary Educational Technology*, *2*(2), 118-134. https://doi.org/10.17718/tojde.10766

Alqurashi, E. (2016). Self-efficacy in online learning environments: A literature review. *Contemporary Issues in Education Research-First Quarter*, *9*(1). https://doi.org/10.19030/cier.v9i1.9549

Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, *40*(1), 133-148. https://doi.org/10.1080/01587919.2018.1553562

Alraimi, K. M., Zo, H., & Ciganek, A. P. (2015). Understanding the MOOCs continuance: The role of openness and reputation. *Computers and Education*, *80*, 28-38. https://doi.org/10.1016/j.compedu.2014.08.006

Artino, A. R. (2007). Motivational beliefs and perceptions of instructional quality: Predicting satisfaction with online training. *Journal of Computer Assisted Learning, 24*(3), 260-270. https://doi.org/10.1111/j.1365-2729.2007.00258.x

Artino, A. R. (2007). Online military training: Using a social cognitive view of motivation and self-regulation to understand students` satisfection, perceived learning, and choice. *Quarterly Review of Distance Education*, *8*(3), 37-45. Retrieved from https://search.proquest.com/docview/231066149?accountid=6724

Ašeriškis, D., & Damaševičius, R. (2014). Gamification patterns for gamification applications. *Procedia Computer Science*, *39*, 83-90.

Aviv, R., Erlich, Z., & Ravid, G. (2005). Response neighborhoods in online learning networks: A quantitative analysis. *Journal of Educational Technology & Society*, *8*(4), 90-99.

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*.

Bandura, A. (1995). *Self-efficacy in changing societies*. New York, NY: Cambridge University Press.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning, 9*(2), 161-185. https://doi.org/10.1007/s11409-013-9107-6

Baraniuk, R. (2008). Challenges and opportunities for the open education movement: A Connexions case study. In *Opening up education*. Cambridge, MA: MIT Press.

Barata, G., Gama, S., Jorge, J., & Goncalves, D. (2013). *Engaging engineering students with gamification*. 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), Bournemouth, UK.

Barnard-Brak, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S.-L. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education*, *12*(1), 1-6. https://doi.org/10.1016/j.iheduc.2008.10.005.

Barnard-Brak, L., Paton, V., & Lan, W. (2010). Profiles in self-regulated learning in the online learning environment. *International Review of Research in Open and Distance Learning*. 11(1). DOI: https://doi.org/10.19173/irrodl.v11i1.769

Bates, T. (2013). Harvard's current thinking on MOOCs. Retrieved from https://www.tonybates.ca/2013/02/14/harvards-current-thinking-on-moocs/

Bates, T. (2015). *Teaching in a digital age: Guidelines for designing teaching and learning for a digital age*. Tony Bates Associates.

Belanger, Y., & Thornton, J. (2013). *Bioelectricity: A quantitative approach*. Retrieved from http://onlinecourses.duke.edu/

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78-117. https://doi.org/10.1177/0049124187016001004.

Bergek, A., Berggren, C., Magnusson, T., & Hobday, M. (2013). Technological discontinuities and the challenge for incumbent firms: Destruction, disruption or creative accumulation? *Research Policy*, *42*(6-7), 1210-1224. https://doi.org/10.1016/J.RESPOL.2013.02.009

Bezroukov, N. (1999a). A second look at the Cathedral and the Bazaar. *First Monday*, *4*(12).

Bezroukov, N. (1999b). Open source software development as a special type of academic research: Critique of vulgar Raymondism. *First Monday*, *4*(10). Retrieved from http://journals.uic.edu/ojs/index.php/fm/article/view/696

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamin.

Blessinger, P., & Bliss, T. J. (2016). *Open education: International perspectives in higher education*. Open Book Publishers.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction*, *7*(2), 161–186. https://doi.org/10.1016/S0959-4752(96)00015-1

Bonafini, F., Chae, C., Park, E., & Jablokow, K. (2017). How much does student engagement with videos and forums in a MOOC affect their achievements? *Online Learning Journal*, *21*(4), 223–240. https://doi.org/10.24059/olj.v21i4.1270

Bornschlegl, M., & Cashman, D. (2019). Considering the role of the distance student experience in student satisfaction and retention. *Open Learning: The Journal of Open, Distance and e-Learning, 34*(2), 139-155. https://doi.org/10.1080/02680513.2018.1509695

Boud, D. (2013). *Enhancing learning through self-assessment*. London, United Kingdom: Kogan Page.

Bozdoğan, D., & Özen, R. (2014). Use of ICT technologies and factors affecting pre-service ELT teachers' perceived ICT self-efficacy. *TOJET: The Turkish Online Journal of Educational Technology, 13*(2), 186-196. Retrieved from http://www.tojet.net/articles/v13i2/13219.pdf

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press. https://doi.org/https://doi.org/10.17226/9853

Brennan, K. (2013, July 24). In connectivism, no one can hear you scream: A guide to understanding the MOOC novice [Blog post]. *Hybrid Pedagogy*. Retrieved from http://www.rpajournal.com/studying-learning-in-the-worldwide-classroom-research-into-edxs-first-mooc/

Breslow, L., Pritchard, D., DeBoer, J., Stump, G., Ho, A., & Seaton, D. (2013). Studying learning in the worldwide classroom:research into edX's first MOOC. *Research & Practice in Assessment*, *8*, 13-25. Retrieved from http://www.rpajournal.com/studying-learning-in-the-worldwide-classroom-research-into-edxs-first-mooc/

Brooks, C., Thompson, C., & Teasley, S. (2015). A time series interaction analysis method for building predictive models of learners using log data. In *LAK '15 Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 126-135). Poughkeepsie, New York. https://doi.org/10.1145/2723576.2723581

Bryant, F. B., & Yarnold, P. R. (1995). *Principal-components analysis and exploratory and confirmatory factor analysis*. Washington, DC, US: American Psychological Association.

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, *24*(6), 1162-1175.

Butler, C. (2013). *The effect of leaderboard ranking on players' perception of gaming fun*. 5th International Online Communities and Social Computing Conference, Las Vegas, NV, US.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*, 245-281. https://doi.org/10.3102/00346543065003245

Byrne, M., & Flood, B. (2005). A study of accounting students' motives, expectations and preparedness for higher education. *Journal of Further and Higher Education*. https://doi.org/10.1080/03098770500103176

Calders, T., & Pechenizkiy, M. (2012). Introduction to the special section on educational data mining. *ACM SIGKDD Explorations Newsletter*, *13*(2), 3-6. https://doi.org/10.1145/2207243.2207245

Callan, R. C., Bauer, K. N., & Landers, R. N. (2015). How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. In *Gamification in education and business* (pp. 553–568). Cham, Switzerland: Springer.

Cattell, R. (2012). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Springer Science & Business Media.

Castaño-Muñoz, J., Kreijns, K., Kalz, M., & Punie, Y. (2017). Does digital competence and occupational setting influence MOOC participation? Evidence from a cross-course survey. *Journal of Computing in Higher Education*, *29*(1), 28-46. https://doi.org/10.1007/s12528-016-9123-z

Caswell, T., Henson, S., Jensen, M., & Wiley, D. (2008). Open Educational Resources: Enabling universal education. *The International Review of Research in Open and Distributed Learning*, *9*(1). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/469/1001

Celino, I., & Dell'Aglio, D. (2015). Capturing the semantics of simulation learning with linked data. In *Gamification: Concepts, methodologies, tools, and applications: Concepts, methodologies, tools, and applications* (pp. 273). Hershey, PA, USA: IGI Global.

Chang, S., & Smith, R. (2008). Effectiveness of personal interaction in a learner-centered paradigm distance education class based on student satisfaction. *Journal of Research on Technology in Education*, *40*(4), 407-426. https://doi.org/10.1080/15391523.2008.10782514

Chen, J., Feng, J., Sun, X., Wu, N., Yang, Z., & Chen, S. (2019). MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. *Mathematical Problems in Engineering, 2019*. https://doi.org/10.1155/2019/8404653

Chesbrough, H., Vanhaverbeke, W., & West, J. (2014). *New frontiers in open innovation*. Oxford, UK: Oxford University Press.

Chi, M, T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology*, *5*, 161-238.

Chin, W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research*, (pp. 295-336). Mahwah: Erlbaum.

Chinces, D., & Salomie, I. (2015). *Optimizing Spaghetti process models*. 2015 20th International Conference on Control Systems and Computer Science.

Christensen, C. M., Horn, M. B., Caldera, L., & Soares, L. (2011). Disrupting college: How disruptive innovation can deliver quality and affordability to postsecondary education. *Innosight Institute*.

Christensen, C. M., Horn, M. B., & Johnson, C. W. (2011). *Disrupting class: how disruptive innovation will change the way the world learns*. New York: McGraw-Hill.

Christensen, C. M., & Raynor, M. E. (2003). *The innovator's solution: Creating and sustaining successful growth*. Harvard Business Press.

Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., & Emanuel, E. J. (2013). The MOOC phenomenon: Who takes massive open online courses and why? https://doi.org/10.2139/ssrn.2350964

Chuang, I., & Ho, A. D. (2016). HarvardX and MITx: Four years of open online courses -- fall 2012-summer 2016. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2889436

Cisel, M. (2014). Analyzing completion rates in the first French xMOOC. In *The European MOOCs Stakeholders Summit* (pp. 26-32). Lausanne, Switzerland. Retrieved from https://www.openeducationeuropa.eu/sites/default/files/legacy_files/asset/From-field_37_6.pdf

Clark, L., Ting, I.-H., Kimble, C., Wright, P. C., & Kudenko, D. (2006). Combining ethnographic and clickstream data to identify user web browsing strategies. *Information Research: An International Electronic Journal*, *11*(2), 14.

Clow, D. (2013). *MOOCs and the funnel of participation*. Proceedings of the Third International Conference on Learning Analytics and Knowledge.

Codish, D., Rabin, E., & Ravid, G. (2019). User behavior pattern detection in unstructured processes - a learning management system case study. *Interactive Learning Environments*, *27*(5-6), 699-725. https://doi.org/10.1080/10494820.2019.1610456

Codish, D., & Ravid, G. (2014a). Academic course gamification: The art of perceived playfulness. *Interdisciplinary Journal of e-Skills and Lifelong Learning*, *10*, 131-151.

Codish, D., & Ravid, G. (2014b). *Personality based gamification: How different personalities perceive gamification*. Paper presented at the European Conference of Information Systems (ECIS) 2014, Tel-Aviv.

Codish, D., & Ravid, G. (2015). Detecting playfulness in educational gamification through behavior patterns. *IBM Journal of Research and Development*, *59*(6), 6:1–6:14.

Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). *Visualizing patterns of student engagement and performance in MOOCs*. Proceedings of the fourth international conference on learning analytics and knowledge.

Collini, S. (2012). *What are Universities For?* Penguin UK.

Colombo, M. G., Piva, E., & Rossi-Lamastra, C. (2014). Open innovation and within-industry diversification in small and medium enterprises: The case of open source software firms. *Research Policy*, *43*(5), 891-902. https://doi.org/10.1016/j.respol.2013.08.015

Commission European. (2020). 2020 European semester: Assessment of progress on structural reforms, prevention and correction of macroeconomic imbalances, and results of in-depth reviews under Regulation (EU) No. 1176(2011). Retrieved from https://ec.europa.eu/info/sites/info/files/2020-uropean_semester_communicationcountry_reports_en.pdf

Conole, G. (2012). Integrating OER into open educational practices. In J. Glennie, K. Harley, N. Butcher, & T. van Wyk (Eds.), *Open educational resources and change in higher education: Reflections from practice* (pp. 111-124).

Costa, C., Alvelos, H., & Teixeira, L. (2012). The use of Moodle e-learning platform: A study in a Portuguese University. *Procedia Technology*, *5*, 334-343.

Costa, J. P., Wehbe, R. R., Robb, J., & Nacke, L. E. (2013). Time's up: Studying leaderboards for engaging punctual behaviour. Gamification 2013 Conference, Stratford, ON, Canada.

Costello, E. (2014). *Participatory practices in open source educational software - The case of the Moodle bug tracker community* (phd). University of Dublin.

Craig, R. (2015). College disrupted: The Great Unbundling of Higher Education. St. Martin's Press.

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 6-14. https://doi.org/10.1145/2883851.2883931

Daily, J. (2014). HarvardX's and MITx's MOOC data visualized and mapped. *EdTech Magazine*. Retrieved from

http://www.edtechmagazine.com/higher/article/2014/02/harvardxs-and-mitxs-mooc-data-visualized-and-mapped

Davis, D., Chen, G., Hauff, C., & Houben, G. (2016). Gauging MOOC learners' adherence to the designed learning path. In T Barnes, M. Chi, M., & M. Feng (Eds.). *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 54-61). International Educational Data Mining Society. Retrieved from http://www.educationaldatamining.org/EDM2016/proceedings/edm2016_fullpapers.pdf

De Langen, F. H. T. (2013). Strategies for sustainable business models for open educational resources. *The International Review of Research in Open and Distributed Learning*, *14*(2), 53-66. https://doi.org/10.19173/irrodl.v14i2.1533

De Medeiros, A. A., & Weijters, A. (2005). *Genetic process mining*. Applications and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "course": reconceptualizing educational variables for massive open online courses. *Educational Researcher*, *43*(2), 74-84. https://doi.org/10.3102/0013189X14523038

DeCoster, J., Gallucci, M., & Iselin, A.-M. R. (2011). Best practices for using median splits, artificial categorization, and their continuous alternatives. *Journal of Experimental Psychopathology*, *2*(2), jep.008310. https://doi.org/10.5127/jep.008310

Despujol, I. M., Turro, C., Busqueis, J., & Canero, A. (2015). Analysis of demographics and results of student's opinion survey of a large scale MOOC deployment for the Spanish speaking community. In *Proceedings - Frontiers in Education Conference, FIE* (Vol. 2015-February). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/FIE.2014.7044102

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From game design elements to gamefulness: Defining gamification*. 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland.

Dev, P. C. (1997). Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? *Remedial and Special Education*, *18*(1), 12-19. https://doi.org/10.1177/074193259701800104

Dillahunt, T., Ng, S., Fiesta, M., & Wang, Z. (2016). Do massive open online course platforms support employability?. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing, CSCW* (pp. 233-244). San Francisco, United States: Association for Computing Machinery. https://doi.org/10.1145/2818048.2819924

Donald, J. G. (1999). Motivation for higher-order learning. *New Directions for Teaching and Learning*, *1999*(78), 27-35. https://doi.org/10.1002/tl.7803

Downes, S. (2012). Connectivism and connective knowledge: Essays on meaning and learning networks. *National Research Council Canada.* Retrieved from https://www.downes.ca/files/books/Connective_Knowledge-19May2012.pdf

Economist Intelligence Unit. (2015). *Connecting universities: Future models of higher education*. London, UK.

Eom, S. B., Wen, H. J., & Ashill, N. (2006). The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. *Decision Sciences Journal of Innovative Education*, *4*(2), 215-235.

Elgort, I., Lundqvist, K., McDonald, J., & Moskal, A. C. (2018). Analysis of student discussion posts in a MOOC: Proof of concept. In *Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18).* Sydney, NSW, Australia: SoLAR.

Elias, T. (2011). *Learning analytics: Definitions, processes and potential*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.7092&rep=rep1&type=pdf

Emanuel, E. J. (2013). Online education: MOOCs taken by educated few. *Nature*, *503*(7476), 342. https://doi.org/10.1038/503342a

Everitt, B. (1975). Multivariate analysis: The need for data, and other problems. *The British Journal of Psychiatry*, *126*(3), 237-240.

Farrow, R. (2017). Open education and critical pedagogy. *Learning, Media and Technology*, *42*(2), 130-146. https://doi.org/10.1080/17439884.2016.1113991

Faucon, L., Kidzinski, L., & Dillenbourg, P. (2016). *Semi-Markov model for simulating MOOC students.* Proceedings of the 9th International Conference on Educational Data Mining.

Ferreira, D., Zacarias, M., Malheiros, M., & Ferreira, P. (2007). *Approaching process mining with sequence clustering: Experiments and findings*. 5th International Conference, Business Process Management, Brisbane, Australia, September 24-28.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304-317. https://doi.org/10.1504/IJTEL.2012.051816

Field, A. P. (2005). *Discovering statistics using SPSS* (2nd ed.). London, United Kingdom: Sage.

Finley, K. (2015, September). Open sourcing is no longer optional, not even for Apple. *Wired*.

Fischer, G. (2000). Lifelong learning - more than training. *Journal of Interactive Learning Research, 11*(3), 265-294. Charlottesville, VA: Association for the Advancement of Computing in Education (AACE). Retrieved September 7, 2020 from https://www.learntechlib.org/primary/p/8380/.

Fitzgerald, B. (2006). The transformation of open source software. *MIS Quarterly*, 587-598.

Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

Fredin, E. (2018, April). First blockchain university promises to be the Uber for students and Airbnb for teachers.

Friedman, T. (2012). Come the revolution. *New York Times*. Retrieved from http://www.nytimes.com/2012/05/16/opinion/friedman-come-the-revolution.html?_r=0

Gabrielatos, C., & Marchi, A. (2012). Keyness: Appropriate metrics and practical issues. *CADS International Conference*.

Gannon-Cook, R. (2010). *What motivates faculty to teach in distance education? A case study and meta-literature review*. Lanham, MD: University Press of America.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127-203. https://doi.org/10.1007/s11257-018-9203-z

Garcia, T., & Pintrich, P. R. (1994). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of Learning and Performance: Issues and Educational Applications* (pp. 127-153). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127-203. https://doi.org/10.1007/s11257-018-9203-z

Garrido, M., Koepke, L., Anderson, S., & Mena, A. F. (2016). *The advancing MOOCs for development initiative: An examination of MOOC usage for professional workforce development outcomes in Colombia, the Philippines, & South Africa.* Seattle: Technology & Social Change Group, University of Washington Information School. http://hdl.handle.net/1773/35647

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, *7*(2), 95-105. https://doi.org/10.1016/j.iheduc.2004.02.001

Gee, J. P. (2005a). *Good video games and good learning*. New York, NY, USA: Phi Kappa Phi Forum.

Gee, J. P. (2005b). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, *2*(1), 5-16.

Geigle, C., & Zhai, C. (2017). *Modeling MOOC student behavior with two-layer hidden Markov models.* Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale.

Ghoneim, A., Abbass, H., & Barlow, M. (2008). Characterizing game dynamics in two-player strategy games using network motifs. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *38*(3), 682-690.

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education*, *94*, 241-251. https://doi.org/10.1016/j.compedu.2015.11.002

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*, 129-152.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*(3), 532-560.

Greco, A. N., & Wharton, R. M. (2008). Should university presses adopt an open sccess [electronic publishing] Business model for all of their scholarly books? In *Proceedings ELPUB 2008 Conference on Electronic Publishing*. Toronto, Canada. Retrieved from https://elpub.architexturez.net/system/files/pdf/149_elpub2008.content.pdf

Greco, G., Guzzo, A., & Pontieri, L. (2005). *Mining hierarchies of models: From abstract views to concrete specifications*. International conference on Business Process management.

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

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*. *15*(3), 42-57.

Guo, P., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@scale conference* (pp. 21-30). ACM New York, USA. https://doi.org/10.1145/2556325.2566247

Guri-Rosenblit, S. (2019). Open universities: Innovative past, challenging present, and prospective future. *International Review of Research in Open and Distance Learning*, *20*(4), 180-194. https://doi.org/10.19173/irrodl.v20i4.4034

Hakulinen, L., Auvinen, T., & Korhonen, A. (2013). *Empirical study on the effect of achievement badges in TRAKLA2 online learning environment conference*. Learning and Teaching in Computing and Engineering (LaTiCE), Macau, China.

Hamari, J., & Koivisto, J. (2013). *Social motivations to use gamification: An empirical study of gamifying exercise*. 21st European Conference on Information Systems, June 5-8, 2013, Utrecht, The Netherlands.

Hamari, J., & Koivisto, J. (2015). Why do people use gamification services? *International Journal of Information Management*, *35*(4), 419-431.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work? A literature review of empirical studies on gamification*. 47th Hawaii International Conference on System Sciences, Hawaii, USA.

Hammond, T., Danko, K., & Braswell, M. (2015). U.S. accounting professors' perspectives on textbook revisions. *Journal of Accounting Education*, *33*(3), 198-218. https://doi.org/10.1016/J.JACCEDU.2015.06.004

Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. (2001). *Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth*. Proceedings of the 17th international conference on data engineering.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152-161.

Hansen, J. D., & Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, *350*(6265), 1245-1248. https://doi.org/10.1126/science.aab3782

Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics*, *1*(1), 140-149. https://doi.org/10.18608/jla.2014.11.8

Henderikx, M., Kreijns, K., & Kalz, M. (2017). To change or not to change? That's the question, On MOOC-success, barriers and their implications. In C. Delgado Kloos, P. Jermann, M. Pérez-Sanagustin, D. T. Seaton, & S. White (Eds.), *Lecture notes in computer science: Vol. 10254. Digital education: Out to the world and back to the campus* (pp. 210-216). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-59044-8_25

Henderikx, M. A., Kreijns, K., & Kalz, M. (2017). Refining success and dropout in massive open online courses based on the intention–behavior gap. *Distance Education*,*38*(3), 353-368. https://doi.org/10.1080/01587919.2017.1369006

Henderikx, M., Kreijns, K., & Kalz, M. (2018). A Classification of barriers that influence intention achievement in MOOCs. In H. V. Pammer-Schindler, M. Pérez-Sanagustín & & M. S. Drachsler, R. Elferink (Eds.), *Lecture notes in computer science* (pp. 3-15). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-98572-5_1

Henderikx, M., Kreijns, K., Castaño Muñoz, J., & Kalz, M. (2019). Factors influencing the pursuit of personal learning goals in MOOCs. *Distance Education*, *40*(2), 187-204. https://doi.org/10.1080/01587919.2019.1600364

Henderikx, M., Kreijns, K., & Kalz, M. (2018). A classification of barriers that influence intention achievement in MOOCs. In H. V. Pammer-Schindler, M. Pérez-Sanagustín, M. S. Drachsler, & R. Elferink (Eds.), *Lecture Notes in Computer Science* (pp. 3-15). Springer, Cham. https://doi.org/10.1007/978-3-319-98572-5_1

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. https://doi.org/10.1016/j.edurev.2014.05.001

Ho, A. D., Chuang, I., Reich, J., Coleman, C. A., Whitehill, J., Northcutt, C. G., … Petersen, R. (2015). HarvardX and MITx: Two years of open online courses; Fall 2012-Summer 2014. *SSRN Electronic Journal, 10*, 1-37. https://doi.org/10.2139/ssrn.2586847

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The first year of open online courses, Fall 2012-Summer 2013. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2381263

Hong, B., Wei, Z., & Yang, Y. (2019). A two-layer cascading method for dropout prediction in MOOC. *Mechatronic Systems and Control*, *47*(2), 91-97. https://doi.org/10.2316/J.2019.201-2980

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.

Horvat, A., Dobrota, M., Krsmanovic, M., & Cudanov, M. (2015). Student perception of Moodle learning management system: A satisfaction and significance analysis. *Interactive Learning Environments*, *23*(4), 515-527. https://doi.org/10.1080/10494820.2013.788033

Hou, H. T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, *48*, 424-435.

Huang, T. C., Chen, M. Y., & Lin, C. Y. (2019). Exploring the behavioral patterns transformation of learners in different 3D modeling teaching strategies. *Computers in Human Behavior*, *92*, 670-678.

Ingvaldsen, J. E., & Gulla, J. A. (2008). *Preprocessing support for large scale process mining of SAP transactions*. Business Process Management Workshops.

ISO 9241-11. (1998). Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.

Jans, M., Van der Werf, J. M., Lybaert, N., & Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications*, *38*(10), 13351–13359.

Jansen, D., Schuwer, R., Teixeira, A., & Aydin, H. (2015). Comparing MOOC adoption strategies in Europe: Results from the HOME project survey. *International Review of Research in Open and Distributed Learning, 16*(6), 116-136. https://doi.org/10.19173/irrodl.v16i6.2154

Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially-shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*, *42*, 1-11. https://doi.org/10.1016/j.learninstruc.2015.10.006

Jisoo, L., Ahreum, H., & Junseok, H. (2018). A review of massive open online courses: MOOC's approach to bridge the digital divide. *22nd Biennial Conference of theInternational Telecommunications Society (ITS): Beyond the Boundaries: challenges for Business, Policy and Society.* https://www.econstor.eu/handle/10419/190394

Johnson, M. W., Christensen, C. M., & Kagermann, H. (2008). Reinventing your business model. *Harvard Business Review*, (December), 59-68.

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, *31*(6), 638-654. https://doi.org/10.1111/jcal.12107

Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education*, *16*(1), 24. https://doi.org/10.1186/s41239-019-0155-0

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning*, *15*(1). https://doi.org/http://dx.doi.org/10.19173/irrodl.v15i1.1651

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education*, *33*, 74-85. https://doi.org/10.1016/j.iheduc.2017.02.001

Kahan, T., Soffer, T., & Nachmias, R. (2017). Types of participant behavior in a massive open online course. *International Review of Research in Open and Distance Learning*, *18*(6), 1-18. https://doi.org/10.19173/irrodl.v18i6.3087

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151.

Kalman, Y. M. (2014). A race to the bottom: MOOCs and higher education business models. *Open Learning: The Journal of Open, Distance and e-Learning*, *29*(1), 5-14. https://doi.org/10.1080/02680513.2014.922410

Kalz, M. (2015). Lifelong learning and its support with new technologies. In J. D. Wright(Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 93-99). Oxford: Elsevier.

Kalz, M., Kreijns, K., Walhout, J., Castaño-Munoz, J., Espasa, A., & Tovar, E. (2015). Setting-up a European cross-provider data collection on open online courses. *The International Review of Research in Open and Distributed Learning*, *16*(6). https://doi.org/10.19173/irrodl.v16i6.2150

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y. Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research*, *3*(3), 109-136.

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, *72*(7), 14. doi:10.1016/j.chb.2016.09.062.

Kankanhalli, A., Taher, M., Cavusoglu, H., & Kim, S. H. (2012). *Gamification: A new paradigm for online user engagement*. 33rd International Conference on Information Systems, Orlando, US.

Kaplan, A. M., & Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. *Business Horizons*, *59*(4), 441-450. https://doi.org/10.1016/j.bushor.2016.03.008

Keller, J. (1983). Motivational design of instruction. In C. Reigeluth (Ed.), *Instructional design theories and models: An overview* (pp. 386-434). Hillsdale, NJ: Erlbaum.

Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention-A literature review. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 1236-1244). Chesapeak, VA: AACE.

Kim, J., Singh C, D., & Rhim, E. (2015). Predicting student attrition in MOOCs using sentiment analysis and neural networks. *AIED 2015 Workshop Proceedings*.

Kitsantas, A., & Zimmerman, B. J. (2002). Comparing self-regulatory processes among novice, non-expert, and expert volleyball players: A microanalytic study. *Journal of Applied Sport Psychology, 14*(2), 91-105. https://doi.org/10.1080/10413200252907761

Kizilcec, R, F, & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the Second (2015) ACM Conference on Learning@Scale* (pp. 57-66). New York, USA: ACM Press. https://doi.org/10.1145/2724660.2724680

Kizilcec, R, F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 170-179. http://dl.acm.org/citation.cfm?id=2460330

Kizilcec, R, Perez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, *104*, 18-33. https://doi.org/https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R., 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. https://doi.org/10.1073/pnas.1921417117

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 60-65).

Koller, D., Ng, A., Chuong, D., & Zhenghao, C. (2013). Retention and intention in massive open online courses. *Educause Review*, *48*(3), 62-63. Retrieved June 11, 2015, from http://www.educause.edu/ero/article/retention-and-intention-massive-open-online-courses-depth-0

Konstan, J. A., Walker, J. D., Brooks, D. C., Brown, E. K., & Ekstrand, M. D. (2015). Teaching recommender systems at large scale: Evaluation and lessons learned from a hybrid MOOC. *ACM Transactions on Computer-Human Interaction*, *22*(2), 1-23. https://doi.org/10.1145/2728171

Kovanović, V., Joksimović, S., Gašević, D., Siemens, G., & Hatala, M. (2015). What public media reveals about MOOCs: A systematic analysis of news reports. *British Journal of Educational Technology*, *46*(3), 510-527. https://doi.org/10.1111/bjet.12277

Kuo, Y. C., Walker, A. E., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *Internet and Higher Education*, *20*, 35-50. https://doi.org/10.1016/j.iheduc.2013.10.001

Lackner, E., Kopp, M., & Ebner, M. (2014). How to MOOC? A pedagogical guideline for practitioners. *Proceedings of the 10th International Scientific Conference" ELearning and Software for Education*.

Lambiotte, R., & Kosinski, M. (2014). Tracking the digital footprints of personality. *Proceedings of the IEEE*, *102*(12), 1934–1939.

Landers, R. N., & Landers, A. K. (2015). An empirical test of the theory of gamified learning the effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*, *45*(6), 17.

Lau, H. C., Ho, G. T., Chu, K., Ho, W., & Lee, C. K. (2009). Development of an intelligent quality management system using fuzzy association rules. *Expert Systems with Applications*, *36*(2), 1801–1815.

Lau, H. C., Ho, G. T., Zhao, Y., & Chung, N. (2009). Development of a process mining system for supporting knowledge discovery in a supply chain network. *International Journal of Production Economics*, *122*(1), 176-187.

Laurillard, D. (2016). The educational problem that MOOCs could solve: Professional development for teachers of disadvantaged students. *Research in Learning Technology*, *24*(2016). https://doi.org/10.3402/rlt.v24.29369

Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 803–822). Mahwah, NJ: Lawrence Erlbaum Associates.

Lee, Y., & Choi, J. (2013). A structural equation model of predictors of online learning retention. *The Internet and Higher Education, 16*, 36-42. https://doi.org/10.1016/J.IHEDUC.2012.01.005

Leemans, M., & Van der Aalst, W. M. (2014). *Discovery of frequent episodes in event logs*. International Symposium on Data- Driven Process Discovery and Analysis.

Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & Education, 48*(2), 185-204. https://doi.org/10.1016/J.COMPEDU.2004.12.004

Li, J., Bose, R. J. C., & Van der Aalst, W. M. (2010). *Mining context-dependent and interactive business process maps using execution patterns*. International Conference on Business Process Management.

Li, K. (2019). MOOC learners' demographics, self-regulated learning strategy, perceived learning and satisfaction: A structural equation modeling approach. *Computers & Education*, *132*, 16-30. https://doi.org/10.1016/J.COMPEDU.2019.01.003

Li, W., Grossman, T., & Fitzmaurice, G. (2012). *Gamicad: A gamified tutorial system for first time autocad users*. 25th annual ACM symposium on User interface software and technology, Cambridge, Massachesetts.

Li, X., Wang, T., & Wang, H. (2017). *Exploring N-gram features in clickstream data for MOOC learning achievement prediction* (pp. 328-339). Springer, Cham. https://doi.org/10.1007/978-3-319-55705-2_26

Liaw, S., & Huang, H. (2011). Exploring learners' acceptance toward mobile learning. In T. Teo (Ed.), *Technology acceptance in education* (pp. 145-157). Rotterdam, The Netherlands: Sense Publishers.

Lieberoth, A. (2015). Shallow gamification testing psychological effects of framing an activity as a game. *Games and Culture*, *10*(3), 229-248.

Lim, S., Coetzee, D., Hartmann, B., Fox, A., & Hearst, M. A. (2014). Initial experiences with small group discussions in MOOCs. In *L@S 2014 - Proceedings of the 1st ACM Conference on Learning at Scale* (pp. 151-152). New York, New York, USA: Association for Computing Machinery. https://doi.org/10.1145/2556325.2567854

Little, J. (2016). *The MOOC survey: Drivers, implementation and impact*. Retrieved from http://www.learningtechnologist.co.uk/2015/10/15/mooc-survey/

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. https://doi.org/10.1016/j.iheduc.2015.12.003

Liu, M., Kang, J., & McKelroy, E. (2015). Examining learners' perspective of taking a MOOC: reasons, excitement, and perception of usefulness. *Educational Media International*, *52*(2), 129-146. https://doi.org/10.1080/09523987.2015.1053289

Liyanagunawardena, T., Parslow, P., & Williams, S. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distance Learning*, *14*(3), 202-227. Retrieved from http://centaur.reading.ac.uk/36002/

Loeckx, J. (2016). Blurring boundaries in education: Context and impact of MOOCs. *The International Review of Research in Open and Distributed Learning*, *17*(3). https://doi.org/10.19173/irrodl.v17i3.2395

Lohr, S. (2018). IBM to buy Red Hat, the top Linux distributor, for $34 Billion. *The New York Times*. Retrieved from https://www.nytimes.com/2018/10/28/business/ibm-red-hat-cloud-computing.html

Loizzo, J., & Ertmer, P. A. (2016). MOOCocracy: The learning culture of massive open online courses. *Educational Technology Research and Development*, *64*(6), 1013-1032. https://doi.org/10.1007/s11423-016-9444-7

Lowry, P. B., Gaskin, J. E., Twyman, N. W., Hammer, B., & Roberts, T. L. (2013). Taking "fun and games" seriously: Proposing the hedonic-motivation system adoption model (HMSAM). *Journal of the Association for Information Systems*, *14*(11), 617-671.

Luengo, D., & Sepúlveda, M. (2012). *Applying clustering in process mining to find different versions of a business process that changes over time*. Business Process Management Workshops.

Ma, R., & Mendez, M. C. (2020). *Free online learning due to Coronavirus (Updated Continuously) - Class Central*. Class Central.https://www.classcentral.com/report/free-online-learning-coronavirus/

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99.

Macfarlane, B. (2011). The morphing of academic practice: Unbundling and the rise of the para-academic. *Higher Education Quarterly*, *65*(1), 59-73. https://doi.org/10.1111/j.1468-2273.2010.00467.x

Macleod, H., Haywood, J., Woodgate, A., & Alkhatnai, M. (2015). Emerging patterns in MOOCs: Learners, course designs and directions. *TechTrends: Linking Research & Practice to Improve Learning*, *59*(1), 56-63. https://doi.org/10.1007/s11528-014-0821-y

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big Data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179-196. https://doi.org/10.1016/J.CHB.2017.11.011

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Delgado-Kloos, C. (2018). Predicting learners' success in a self-paced MOOC through sequence patterns of self-regulated learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11082 LNCS, pp. 355-369). Springer Verlag. https://doi.org/10.1007/978-3-319-98572-5_27

Mannila, H., Toivonen, H., & Verkamo, A. I. (1997). Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, *1*(3), 259-289.

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers and Education, 80*, 77-83. https://doi.org/10.1016/j.compedu.2014.08.005

Marszal, A. (2012, December). UK universities to launch free degree-style online courses. *The Telegraph.* London. Retrieved 26.11.2020 https://www.telegraph.co.uk/education/educationnews/9743703/UK-universities-to-launch-free-degree-style-online-courses.html

Martin, F. G. (2012). Will massive open online courses change how we teach. *Communications of the ACM*, *55*(8), 26-28. https://doi.org/10.1145/2240236.2240246

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). *Do points, levels and leaderboards harm intrinsic motivation?: An empirical analysis of common gamification elements*. Gamification 2013, Stratford, Ontario, Canada.

Miles, R. E., & Snow, C. C. (1978). *Organizational Strategy, Structure, and Process*. McGraw-Hill.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824-827.

Mobasher, B., Cooley, R., & Srivastava, J. (2000). Automatic personalization based on web usage mining. *Communications of the ACM*, *43*(8), 142-151.

Moreno-Marcos, P. M., Muñoz-Merino, P. J., Maldonado-Mahauad, J., Pérez-Sanagustín, M., Alario-Hoyos, C., & Kloos, D. C. (2020). Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers and Education*, *145*, (103728). https://doi.org/10.1016/j.compedu.2019.103728

Morris, L. V. (2013). MOOCs, emerging technologies, and quality. *Innovative Higher Education*, *38*(4), 251-252. https://doi.org/10.1007/s10755-013-9263-2

Morris, N., Hotchkiss, S., & Swinnerton, B. (2015). Can demographic information predict MOOC learner outcomes. In *Proceedings of the European Stakeholders Summit on Experience and Best Practices in and Around MOOCs (EMOOCs2015)* (pp. 199-207). Université Catholique de Louvain, Mons, Belgium. Retrieved from http://www.academia.edu/download/37666738/Papers.pdf#page=199

Morse, J. (2016). *Mixed method design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.

Muilenburg, L. Y., & Berge, Z. L. (2007). Student barriers to online learning: A factor analytic study. *Distance Education*, *26*(1), 29-48. https://doi.org/10.1080/01587910500081269

Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, *77*(4), 541-580.

Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: an overview, II. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(6), e1219.

Mukala, P., Buijs, J. C. A. M., & Van Der Aalst, W. M. P. (2015). Exploring students' learning behaviour in moocs using process mining techniques. In *BPM Center Report BPM-15-10. BPMCenter.org*. https://pdfs.semanticscholar.org/f90c/c6dfdfbd14ba86bd6dc381a0985067329866.pdf

Mulder, F. (2015). Open(ing up) education for all boosted by MOOCs? In Bonk, C. J., Lee, M. M., Reeves, T.C., & Reynolds, T.H., (Eds.), *MOOCs and Open Education Around the World*. Taylor and Francis Inc. https://doi.org/10.4324/9781315751108

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824-827. https://doi.org/10.1126/science.298.5594.824

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management: Reinventing innovation management research in a digital age. *MIS Quarterly*, *41*(1).

Naveh, G., Tubin, D., & Pliskin, N. (2012). Student satisfaction with learning management systems: A lens of critical success factors. *Technology, Pedagogy and Education*, *21*(3), 337-350. https://doi.org/10.1080/1475939X.2012.720413

Noam, E. M. (1996). Electronics and the dim future of the university. *Bulletin of the American Society for Information Science and Technology*, *22*(5), 6-9. https://doi.org/10.1002/bult.24

Nussbaumer, A., & Hillemann, E. (2015). A competence-based service for supporting self-regulated learning in virtual environments. *Journal of Learning Analytics 2*(1), 101-133. Retrieved from http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4130

OECD. (2007). *Qualifications systems – bridges to lifelong learning*. Retrieved from http://www.oecd.org/education/skills-beyond-school/38465471.Pdf

Onah, D., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioural patterns. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *Proceedings of EDULEARN14: The 6th International Conference on Education and New Learning Technologies* (pp. 5825–5834). Valencia, Spain: IATED. Retrieved from https://library.iated.org/publications/EDULEARN14

Orr, D., Weller, M., & Farrow, R. (2018). *Models for online, open, flexible and technology enhanced higher education across the globe-a comparative analysis Final Report*. Retrieved from https://icde.memberclicks.net/assets/RESOURCES/Models-report-April-2018_final.pdf

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. Wiley.

Pardo, A., Han, F., & Ellis, R. A. (2016). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2016.2639508

Park, J., & Society, H. C. (2009). Factors influencing adult learners' decision to drop out or persist in online learning. *Journal of Educational Technology & Society, 12*(4), 207-217. Retrieved from https://www.jstor.org/stable/10.2307/jeductechsoci.12.4.207

Patel, P., & Parmar, M. (2014). Improve heuristics for user session identification through web server log in web usage mining. *International Journal of Computer Science and Information Technologies*, *5*(3), 3562–3565.

Paulsen, M., & Gentry, J. (1995). Motivation, learning strategies, and academic performance: A study of the college finance classroom. *Financial Practice and Education*, *5*(1), 78-89. Retrieved from https://elibrary.ru/item.asp?id=2261362

Pennacchiotti, M., & Popescu, A. M. (2011). A machine learning approach to twitter user classification. *ICWSM*, *11*(1), 281-288.

Pintrich, P. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor: The University of Michigan. Retrieved from ERIC database. (ED338122)

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801-813. https://doi.org/10.1177/0013164493053003024

Pozón-López, I., Higueras-Castillo, E., Muñoz-Leiva, F., & Liébana-Cabanillas, F. J. (2020). Perceived user satisfaction and intention to use massive open online courses (MOOCs). *Journal of Computing in Higher Education*, 1-36. https://doi.org/10.1007/s12528-020-09257-9

Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted Learning*, *32*(3), 202-217. https://doi.org/10.1111/jcal.12131

Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning aspredictors of final grade and satisfaction in college-level online courses. *American Journal of Distance Education*, *22*(2), 72-89. https://doi.org/10.1080/08923640802039024

Rabin, E., Silber-Varod, V., & Kalman, Y. M. (2019). Using natural language processing techniques to predict perceived achievements in Massive Online Open Courses. *KM Conference*. University of Life Sciences (SGGW), Warsaw, Poland. June, 2019.

Rabin, E, Kalman, Y. M., & Kalz, M. (2019). Predicting learner-centered MOOC outcomes: Satisfaction and intention-fulfillment. *International Journal of Educational Technology in Higher Education*, *16*(14). https://doi.org/10.1186/s41239-019-0144-3

Rabin, E., Kalman, Y. M., & Kalz, M. (2019b). The cathedral's ivory tower and the open education bazaar: Catalyzing innovation in the higher education sector. *Open Learning: The Journal of Open, Distance and e-Learning, 35*(1), 82-99. https://doi.org/10.1080/02680513.2019.1662285

Raymond, E. (1999). The Cathedral and the Bazaar: Musings on Linux and open source by an accidental revolutionary. *Sebastopol, CA.: O'Reilly*. http://www.springerlink.com/index/pdf/10.1007/s12130-999-1026-0

Raymond, E. S. (2001). *The Cathedral and the Bazaar: Musings on Linux and open source by an accidental revolutionary*. O'Reilly.

Read, M. (2011). Cultural and organizational drivers of open educational content. In R. Katz (Ed.), *The Tower and the Cloud. Higher education in the age of cloud computing* (pp. 140-149). EDUCAUSE. Retrieved from https://lcc.lipscomb.edu/uploads/24663.pdf#page=162

Rebuge, Á, & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, *37*(2), 99-116.

Reich, J. (2014). MOOC completion and retention in the context of student intent. *Educause Review Online*. Retrieved from http://www.educause.edu/ero/article/mooc-completion-andretention-context-tudent-intent

Reich, J. (2015). Rebooting MOOC research. *Science*, *347*(6217), 34-35. https://doi.org/10.1126/science.1261627

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, *363*(6423), 130-131. https://doi.org/10.1126/science.aav7958

Reimann, P., Markauskaite, L., & Bannert. M. (2014). E-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology 45*(3), 528-540. https://doi.org/10.1111/bjet.12146

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333-341. https://doi.org/10.1016/J.CHB.2016.02.074.

Rifkin, J. (2014). *The zero marginal cost society: The internet of things, the collaborative commons, and the eclipse of capitalism*.

Robinson, C., Yeomans, M., Reich, J., Hulleman, C., & Gehlbach, H. (2016). Forecasting student achievement in MOOCs with natural language processing. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 383-387. https://doi.org/10.1145/2883851.2883932

Roca, J. C., Chiu, C. M., & Martínez, F. J. (2006). Understanding e-learning continuance intention: An extension of the Technology Acceptance Model. *International Journal of Human-Computer Studies*, *64*(8), 683-696. https://doi.org/10.1016/j.ijhcs.2006.01.003

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014). Social factors that contribute to attrition in MOOCs. In *Proceedings of the first ACM conference on Learning@scale conference*. Atlanta, Georgia, USA. https://doi.org/10.1145/2556325.2567879

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, *11*(10), 1-13. https://doi.org/10.7275/9wph-vv65

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *Information Society*, *32*(2), 143-159. https://doi.org/10.1080/01972243.2016.1130502

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*. *55*(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Sabadie, J., Muñoz, J., Punie, Y., Redecker, C., & Vuorikari, R. (2014). OER: A European policy perspective. *Journal of Interactive Media in Education*, 2014, 1-12. doi:https://doi.org/10.5334/2014-05

Samuelson, P. (2006). IBM's pragmatic embrace of open source. *Commun. ACM*, *49*(10), 21-25. https://doi.org/10.1145/1164394.1164412

Santos, T., Costa, C., & Aparicio, M. (2014). Do we need a shared European MOOC platform? In *Position papers for European cooperation on MOOCs*, (pp. 99-112).

Schmid, L., Manturuk, K., Simpkins, I., Goldwasser, M., & Whitfield, K. E. (2015). Fulfilling the promise: Do MOOCs reach the educationally underserved? *Educational Media International*, *52*(2), 116-128. https://doi.org/10.1080/09523987.2015.1053288

Schreiber, J. (2008). Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy*, *4*, 83-97. https://doi.org/10.1016/j.sapharm.2007.04.003.

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist, 40*(2), 85-94. https://doi.org/10.1207/s15326985ep4002_3

Scott, M. (2011). *WordSmith tools manual* (No. 6). Lexical Analysis Software Ltd. https://lexically.net/LexicalAnalysisSoftware/

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Benjamins.

Semenova, T. V., & Rudakova, L. M. (2016). Barriers to taking massive open online courses (MOOCs). *Russian Education & Society*, *58*(3), 228-245. https://doi.org/10.1080/10609393.2016.1242992

Shah, D. (2018). *By the numbers: MOOCs in 2018 - ClassCentral*. Retrieved February 16, 2019, from https://www.class-central.com/report/mooc-stats-2018/

Shah, D. (2019). *By the numbers: MOOCs in 2019 - ClassCentral*. Retrieved February 25, 2020, from https://www.classcentral.com/report/mooc-stats-2019/

Shah, D. (2020). How different MOOC providers are responding to the pandemic (Updated) – Class Central. Retrieved July 23, 2020, from https://www.classcentral.com/report/mooc-providers-response-to-the-pandemic/

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and Personality Psychology Compass*, *10*(9), 503-518. https://doi.org/10.1111/spc3.12265

Shen, D., Cho, M.-H., Tsai, C.-L., & Marra, R. (2013). Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. *The Internet and Higher Education, 19*, 10-17. https://doi.org/10.1016/J.IHEDUC.2013.04.001

Siemens, G. (2004). Connectivism: A learning theory for the digital age. *Elearnspace*. Retrieved from http://www.elearnspace.org/Articles/connectivism.htm

Siemens, G, & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30-40. Retrieved from http://eric.ed.gov/?id=EJ950794

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. (2011). *Open Learning Analytics: An integrated modularized platform*. https://research.monash.edu/en/publications/open-learning-analytics-an-integrated-modularized-platform

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). *Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions*. http://arxiv.org/abs/1407.7131

Sisodia, D. S., & Verma, S. (2012). *Web usage pattern analysis through web logs: A review*. Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on Computer Science and Software Engineering (JCSSE), Bangkok, 2012, pp. 49-53, doi: 10.1109/JCSSE.2012.6261924.

Spiliopoulou, M. (2000). Web usage mining for web site evaluation. *Communications of the ACM*, *43*(8), 127-134.

Srikant, R., & Agrawal, R. (1996). *Mining sequential patterns: Generalizations and performance improvements*. 5th International Conference on Extending Database Technology, Avignon, France, March 25-29.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, *1*(2), 12-23.

Stoffelsen, J., & Diehl, P. (2007). *Handboek levensfasebewust personeelsbeleid* [Handbook for life-stage conscious personnel management]. Alphen aan den Rijn, The Netherlands: Kluwer.

Streiner, D. L. (1994). Figuring out factors: the use and misuse of factor analysis. *The Canadian Journal of Psychiatry*, *39*(3), 135-140.

Sun, J. C. Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, *89*, 53-74. https://doi.org/10.1016/j.compedu.2015.08.004

Tait, A. (2008). What are open universities for? *Open Learning*, *23*(2), 85-93. https://doi.org/10.1080/02680510802051871

Townsend, C., & Liu, W. (2012). Is planning good for you? The differential impact of planning on self-regulation. *Journal of Consumer Research, 39*(4), 688-703. https://doi.org/10.1086/665053

Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318-327.

Tseng, V. S., & Lin, K. W. (2006). Efficient mining and prediction of user behavior patterns in mobile web systems. *Information and Software Technology*, *48*(6), 357-369.

UNESCO. (2002). *Forum on the impact of open courseware for higher education in developing countries. Final report.* Retrieved from http://unesdoc.unesco.org/images/0012/001285/128515e.pdf

UNESCO. (2020). Education: From disruption to recovery. Retrieved July 23, 2020, from https://en.unesco.org/covid19/educationresponse

Van den Beemt, A., Buijs, J., & Van der Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. *The International Review of Research in Open and Distributed Learning*, *19*(5). https://doi.org/10.19173/irrodl.v19i5.3748

Van der Aalst, W. M., & Weijters, A. (2004). Process mining: A research agenda. *Computers in Industry*, *53*(3), 231-244.

Van der Aalst, W. M. (2011a). *Process mining: Discovering and improving Spaghetti and Lasagna processes*. Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on.

Van der Aalst, W. M. (2011b). *Process mining: Discovery, conformance and enhancement of business processes*. Berlin, Germany: Springer Science & Business Media.

Van der Aalst, W. M., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., … Buijs, J. (2012). *Process mining manifesto*. Business Process Management Workshops.

Van der Aalst, W. M., & Günth, C. (2007). *Finding structure in unstructured processes: The case for process mining*. 7th International conference on Application of Concurrency to System Design, Bratislava, Slovakia. 10-13 July.

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, *28*(4), 695–704.

Van Helden, J. (2003). Regulatory sequence analysis tools. *Nucleic Acids Research*, *31*(13), 3593–3596.

Van Rosmalen, P., Sloep, P. B., Brouns, F., Kester, L., Berlanga, A., Bitter, M., & Koper, R. (2008). A model for online learner support based on selecting appropriate peer tutors. *Journal of Computer Assisted Learning, 24*(6), 483-493. https://doi.org/10.1111/j.1365-2729.2008.00283.x

Vaughan-Nichols, S. J. (2016, June). Why Microsoft is turning into an open-source company. *ZDNet*.

Vaughan-Nichols, S. J. (2018). Why IBM bought Red Hat: It's all open source cloud, all the time. *ZDnet*. Retrieved from https://www.zdnet.com/article/why-ibm-bought-red-hat-its-all-open-source-cloud-all-the-time/

Vialardi, C., Bravo, J., Shafti, L., & Ortigosa, A. (2009). Recommendation in higher education using data mining techniques url, bibtex presentation. *Proceedings ofSecond Educational Data Mining Conference*, 190-199.

Vladoiu, M. (2011). *State-of-the-art in open courseware initiatives worldwide*. *Informatics in Education*, *10*(2), 271-294.

Wang, Y., & Baker, R. (2018). Grit and intention: Why do learners complete MOOCs? *The International Review of Research in Open and Distributed Learning*, *19*(3). https://doi.org/10.19173/irrodl.v19i3.3393

Weller, M. (2015). MOOCs and the Silicon valley narrative. *Journal of Interactive Media in Education*, *2015*(1). https://doi.org/10.5334/jime.am

Werbach, K. (2014). *(Re) defining gamification: A process approach*. 9th International Conference PERSUASIVE 2014, Padua, Italy, May 21-23.

West, J., & Bogers, M. (2014). Leveraging external sources of innovation: A review of research on open innovation. *Journal of Product Innovation Management*, *31*(4), 814-831. https://doi.org/10.1111/jpim.12125

West, J., & Gallagher, S. (2006). Challenges of open innovation: The paradox of firm investment in open-source software. *R&D Management*, *36*(3), 319-331. https://doi.org/10.1111/j.1467-9310.2006.00436.x

Whipp, J. L., & Chiarelli, S. (2004). Self-regulation in a web-based course: A case study. *Educational Technology Research and Development*, *52*(4), 5-21. https://doi.org/10.1007/BF02504714

Wilby, P. (2018, January). A visionary to save the Open University – or the man who will run it into the ground? *The Guardian*.

Williams, L., & Pennington, D. (2018). *An authentic self: Big Data and passive digital footprints*. International Symposium on Human Aspects of Information Security & Assurance (HAISA 2018).

Winer, A., & Geri, N. (2019). Learning analytics performance improvement design (LAPID) in higher education: Framework and concerns. *Online Journal of Applied Knowledge Management*, *7*(2), 41-55.

Xie, Z. (2020). Modelling the dropout patterns of MOOC learners. *Tsinghua Science and Technology*, *25*(3), 313-324. https://doi.org/10.26599/TST.2019.9010011

Xing, W., Kim, S., & Goggins, S. (2015). Modeling performance in asynchronous CSCL: An exploration of social ability, collective efficacy and social interaction. In *Exploring the material conditions of learning: Proceedings of the computer supported collaborative learning (CSCL 2015), International Society of the Learning Sciences* (pp. 276-283). Gothenburg, Sweden. Retrieved from https://www.isls.org/cscl2015/papers/MC-0268-FullPaper-Goggins.pdf

Yen, C. J., Tu, C.-H., Sujo-Montes, L., & Sealander, K. (2016). A predictor for PLE management: Impacts of self-regulated online learning on students' learning skills. *Journal of Educational Technology Development & Exchange, 9*(1), 29-48. https://doi.org/10.18785/jetde.0901.03

Yoon, S., Kim, S., & Kang, M. (2018). Predictive power of grit, professor support for autonomy and learning engagement on perceived achievement within the context of a flipped classroom. *Active Learning in Higher Education*. https://doi.org/10.1177/1469787418762463

Young, J. (2020). Will COVID-19 lead to another MOOC moment? Retrieved June 24, 2020, from https://www.edsurge.com/news/2020-03-25-will-covid-19-lead-to-another-mooc-moment

Yuan, L., & Powell, S. (2013). MOOCs and disruptive innovation: Implications for higher education. *ELearning Papers*, *33*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.422.5536&rep=rep1&type=pdf

Yukselturk, E., & Yildirim, Z. (2008). Investigation of interaction, online support, course structure and flexibility as the contributing factors to students' satisfaction in an online certificate program. *Journal of Educational Technology & Society. 11*(4), 51-65. https://doi.org/10.2307/jeductechsoci.11.4.51

Zalli, M., Nordin, H., & Hashim, R. (2019). The role of self-regulated learning strategies on learners` satisfaction in massive open online courses (MOOCs). In *ICOEL2019 –International conference on e-learning* (pp. 1-9). Retrieved from https://global.uthm.edu.my/icoel2019/proceeding_icoel2019.pdf#page=4

Zhang, Q., Bonafini, F. C., Lockee, B. B., Jablokow, K. W., & Hu, X. (2019). Exploring demographics and students' motivation as predictors of completion of a massive open online course. *International Review of Research in Open and Distributed Learning*, *20*(2). https://doi.org/10.19173/irrodl.v20i2.3730

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation*, (pp. 13-39). San Diego: Academic Press.

Zimmerman, B.J. (2002). Becoming a self-regulated learner: an overview. *Theory Into Practice*, *41*(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2.

Zimmerman, B., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*(1), 51-59. https://doi.org/http://dx.doi.org/10.1037/0022-0663.82.1.51

## Acknowledgments

This dissertation is dedicated to my late mother, Zippora, who wisely taught me the value of education, and to my father, Shlomo, for his unconditional support.

First and foremost, I would like to thank my academic advisors Prof. Marco Kalz and Prof. Yoram Kalman for their courageous jump into the deep waters of supervising a distance learner and for always being available to guide and support me throughout this long journey.

Next, I would like to thank my wife Noa and my children, Dror, Gili, and Uri - you are great! Without their support and encouragement, it would not have been possible to work on this dissertation.

I would also like to thank the faculty members of the Open University of the Netherlands (OUN) as well as Open University of Israel (OUI). I am especially grateful to the members of OUI's Shoham, Center for Technology in Distance Education for their help in providing the data for the studies as well as to OUI's Research Authority for their financial support.

As well, I would like to thank the members of the GO-GN network for their support and promotion of open education.

Lastly, I would like to thank all the friends, who are too numerous to mention, for their support – without their sound advice, this study would not have been possible.

## Summary

***Through the Lens of the Learner: Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses***

In the digital era, technology is leading to massive changes in the way we learn. Full participation in today's knowledge-based society requires people to become lifelong learners, by upgrading their skills throughout their adult lives to cope with challenges of contemporary societies, both in their work and in their private lives. The ability to learn and adapt to the changes in the ecosystem around us is an increasingly important basic skill in ever-changing technological universe (OECD, 2007). The changes in the knowledge society are fueled by digital innovation. One of the markers for these changes was the rapid rise of use of open educational resources (OER) and mainly massive, online, open courses (MOOCs), which are online courses aimed at unlimited participation and open access via the web. MOOCs enable learners with different academic backgrounds to learn at any place and any time, almost free of charge.

OERs and MOOCs are different from formal educational courses in the sense that participants may have diverse goals and expect a variety of different learning outcomes that can be defined by the participants themselves rather than by the course instructors. As a result, the focus of this dissertation in learner-centered outcomes and in their antecedes.

The aim of this dissertation was to answer the central research question: How to evaluate learner-centered outcomes and their antecedents in open online education? To address this question, two learner-centered outcomes, namely, learner satisfaction and learner intention-fulfillments were identified as alternative course outcome measures.

To guide the research project, five studies were conducted. These five studies defined the theoretical problem and empirically revealed some of the answers using several learning analytics techniques. The first study, in *Chapter 2,* presents a comparative analysis between the business models of traditional HEI and open education. The analysis investigates the impact of digital innovation on the business models of higher education institutions using Raymond's (1999) well-known "Cathedral and Bazaar" metaphor on software engineering methods. The changes promoted by the "bazaar" facilitate the adoption of MOOCs by the mainstream "cathedral", but require, at the same time, the development of new learner-centered outcome measures, which will be appropriate for the emerging educational ecosystem. This chapter contributes to the evolving literature on the strategic impact of open online education on the HEI landscape and emphasizes the need to define and explore more appropriate learning outcomes.

The second study, in *Chapter 3*, introduces two learner-centered outcomes for non-formal lifelong learning frameworks such as MOOCs, namely: learner satisfaction and learner intention-fulfillment. The study empirically defines them and reveals their predictors in a MOOC. The research results clarify the complex nature of the relationship between learner socio-demographic characteristics and psycho-pedagogical characteristics when entering the course, learner behavior, and learner-centered outcomes. The effects of

socio-demographic characteristics and psycho-pedagogical characteristics on the barriers to satisfaction among MOOC participants are discussed in the third study, in *Chapter 4*. Identifying these barriers to satisfaction and predicting them provides additional insight into the nature of learner satisfaction as a learning outcome.

The fourth and the fifth studies, which are presented in *Chapter 5* and *Chapter* 6, extend previous studies that have shown that clustering participants based on their learning trajectories is more informative and has a higher potential for pedagogical improvement, compared to clustering participants based on static-counting of behavioral data (Kizilcec et al., 2013). The fourth study in *Chapter 5* seeks to explore a novel approach to detect user behavior patterns by spotting very short user activity sequences and clustering them based on shared variance. This will allow us to construct meaningful behavior patterns in unstructured processes in MOOCs and in other forms of online learning.

The fifth study in *Chapter 6* identifies the effect of the learning activity sequences of the participants as a predictor of the level of participant intention-fulfillment. In the study, a novel approach borrowed from the natural language process (NLP) domain had been used to identify different learning activity sequences. The association between the identified learning sequences and the level of IF has been significant and meaningful.

The last chapter (*Chapter 7)* provides an overview of the findings in each chapter and gathers insights from the five studies that have been presented. The chapter concludes with a general discussion and conclusions. Implications, limitations, and future research suggestions are offered.

The studies presented in this dissertation have, individually and all together, turned a spotlight on the importance of looking at learner-centered outcomes and suggest a novel perspective to analyze them learners-centered outcomes and success in open distance education forms, such as MOOCs. The educational system, policymakers, and society as a whole should help lifelong learners to learn how to define their goals and regulate their learning process. Those efforts can and should be combined with the emerging support of personalized and artificial intelligent systems.

# Samenvatting

***Door de ogen van de deelnemer: het gebruik van Learning Analytics om leerresultaten te voorspellen in Massive Open Online Courses.***

De technologische ontwikkelingen in het huidige digitale tijdperk leiden tot enorme veranderingen in de manier waarop wij leren. Om deel te kunnen nemen aan de kennismaatschappij is leven lang leren heel belangrijk. Hierdoor kunnen mensen gedurende hun volwassen leven continue vaardigheden ontwikkelen en verbeteren die nodig zijn om met de uitdagingen van de hedendaagse samenleving om te kunnen gaan, zowel in hun werk als in hun privéleven. Het vermogen om te leren en ons aan te passen aan de veranderingen in het ecosysteem om ons heen is een steeds belangrijkere basisvaardigheid in het steeds veranderende technologische universum (OESO, 2007). De veranderingen in de kennismaatschappij worden gevoed door digitale innovatie. Een van de grote veranderingen was de snelle opkomst van Open Educational Resources (OER), en voornamelijk Massive, Open, Online Courses (MOOC's). Deze online cursussen zijn open toegankelijk voor iedereen, onafhankelijk van academische achtergrond, plaats of tijd.

Open Educational Resources en MOOC's verschillen van formele onderwijscursussen doordat deelnemers uiteenlopende (leer)intenties kunnen hebben en verschillende leerdoelen die door henzelf gedefinieerd kunnen worden in plaats van door cursusdocenten. De focus van dit proefschrift ligt op de door de deelnemer zelf bepaalde leerdoelen en hun antecedenten.

Het doel van dit proefschrift was het beantwoorden van de centrale onderzoeksvraag: Hoe kun je in open online onderwijs de door deelnemers zelf bepaalde leerdoelen en bijbehorende antecedenten evalueren? Om deze vraag te beantwoorden werden twee deelnemer gerelateerde beoordelingsmaatstaven geïdentificeerd als leerresultaat, namelijk tevredenheid en mate van intentievervulling.

Om de centrale onderzoeksvraag te kunnen beantwoorden zijn er vijf onderzoeken uitgevoerd. In deze vijf onderzoeken werd het theoretische probleem gedefinieerd waarop met behulp van verschillende learning analytics technieken een antwoord gezocht werd. Het eerste onderzoek, in *hoofdstuk 2*, presenteert een vergelijkende analyse tussen de businessmodellen van het traditionele hoger onderwijs en het open onderwijs. De analyse onderzocht de impact van digitale innovatie op de businessmodellen van hoger onderwijs instellingen met behulp van Raymond's (1999) bekende "Cathedral and Bazaar" metafoor over software engineering methodes. De veranderingen die door de "Bazaar" worden gepromoot, vergemakkelijken de adoptie van MOOC's door de mainstream "Cathedral". Tegelijkertijd vereist dit de ontwikkeling van deelnemer gerelateerde beoordelingsmaatstaven die passen bij het opkomende onderwijsecosysteem. Dit hoofdstuk draagt bij aan de evoluerende literatuur over de strategische impact van open online onderwijs op het hoger onderwijslandschap en benadrukt de noodzaak om meer geschikte beoordelingsmaatstaven te definiëren en te verkennen.

Het tweede onderzoek, in *hoofdstuk 3*, introduceert twee deelnemer gerelateerde beoordelingsmaatstaven voor levenslange niet-formele onderwijskaders, zoals MOOC's, namelijk: tevredenheid van de deelnemer en intentievervulling van de deelnemer. De onderzoeksresultaten laten het complexe karakter zien van de relatie tussen sociaal-demografische kenmerken en psycho-pedagogische kenmerken van de deelnemers, het leergedrag van de deelnemer en de resultaten van de deelnemer. De effecten van de sociaal-demografische kenmerken en psycho-pedagogische kenmerken op de tevredenheid van de MOOC-deelnemers worden besproken in het derde onderzoek, in *hoofdstuk 4*. Het identificeren van tevredenheidsbarrières en het voorspellen ervan geeft extra inzicht in de tevredenheid van de deelnemer als leerresultaat.

Het vierde en vijfde onderzoek, die in hoofdstuk 5 en hoofdstuk 6 worden gepresenteerd, vormen een uitbreiding op eerdere onderzoeken die aangetoond hebben dat het clusteren van deelnemers op basis van hun leertraject informatiever is en een hoger potentieel heeft voor pedagogische verbetering, in vergelijking met het clusteren van deelnemers puur op basis van een optelsom van gedragsgegevens (Kizilcec et al., 2013). Het vierde onderzoek in *hoofdstuk 5*, verkent een nieuwe aanpak om gedragspatronen van deelnemers te detecteren door het identificeren van zeer korte leeractiviteit sequenties en deze te clusteren op basis van gedeelde variantie. Hierdoor kunnen zinvolle gedragspatronen geconstrueerd worden uit ongestructureerde processen in MOOC's alsook in andere vormen van online leren.

Het vijfde onderzoek, in *hoofdstuk 6*, laat zien in hoeverre leeractiviteit sequenties van deelnemers voorspeller zijn van de mate van intentie-vervulling van deelnemers. In het onderzoek is een methode gebruikt uit het domein van natuurlijke taalprocessen (NLP), om verschillende leeractiviteit sequenties te identificeren. Het verband tussen de geïdentificeerde leersequenties en het niveau van intentievervulling was statistisch significant en betekenisvol.

Het laatste hoofdstuk, *hoofdstuk 7*, geeft een overzicht van de bevindingen en inzichten van de vijf gepresenteerde onderzoeken. Het hoofdstuk wordt afgesloten met een algemene discussie en conclusie. implicaties, limitaties, waarna suggesties gedaan worden voor toekomstige onderzoek. De onderzoeken die in dit proefschrift beschreven zijn brengen individueel en samen het belang van deelnemer gerelateerde beoordelingsmaatstaven als leerresultaat onder de aandacht en bieden nieuwe mogelijkheden om deze deelnemer gerelateerde beoordelingsmaatstaven en prestaties in open afstandsonderwijsvormen, zoals MOOC's, te analyseren. Het onderwijssysteem, de beleidsmakers en de maatschappij als geheel moeten deelnemers tijdens hun hele leven helpen om te leren hoe ze hun (leer)doelen kunnen definiëren en hun leerproces kunnen reguleren. Deze inspanningen kunnen en moeten gecombineerd worden met de ondersteuning van gepersonaliseerde en kunstmatige intelligente systemen.

## תמצית

### בעיני הלומד: שימוש באנליטיקות למידה לניבוי מדדי הצלחה ממוקדי לומד בקורסים מקוונים, פתוחים ומרובי משתתפים

בעידן הדיגיטלי, הטכנולוגיה מובילה לשינויים מרחיקי לכת באופן בו אנו לומדים. כיום, השתתפות מלאה בחברת המידע מחייבת למידה לאורך החיים ופיתוח מתמשך של מיומנויות, הן מקצועיות והן אישיות. היכולת ללמוד ולהתאים את היכולות האישיות לצרכי החברה והטכנולוגיה המשתנים ללא הרף הן מיומנויות בסיסיות שחשיבותן הולכת וגדלה (OECD, 2007). השינויים בחברת המידע מונעים על ידי חדשנות דיגיטלית. אחד הסממנים לשינויים אלה הוא העלייה המהירה בשימוש בחומרי למידה פתוחים ובעיקר בקורסים מקוונים, פתוחים ורבי משתתפים המכונים קורסי מוק. קורסי מוק הם קורסים מקוונים המיועדים לשימוש של מספר בלתי מוגבל של משתתפים ומאפשרים גישה חופשית באמצעות רשת האינטרנט. קורסי המוק מאפשרים ללומדים בעלי רקעים אקדמיים מגוונים ללמוד בכל מקום ובכל זמן, תוך שיפור חווית הלמידה, כמעט ללא עלות.

חומרי למידה פתוחים וקורסי מוק שונים מלמידה בקורסים אקדמיים פורמליים בכך שהם מאפשרים ללומדים שונים להגדיר בעצמם את מדדי ההצלחה שלהם וזאת בניגוד לקורסים אקדמיים פורמליים בהם מדדי ההצלחה של הלמידה מוגדרים על ידי מפתחי הקורסים. לכן, המיקוד של מסה זו הינה במדדי הצלחה ממוקדי לומד בקורסים מקוונים, פתוחים ורבי משתתפים ובמנבאים שלהם.

שאלת המחקר המרכזית במסה זו הינה: כיצד ניתן להעריך את מדדי ההצלחה ממוקדי הלומד וכיצד ניתן לזהות את המנבאים שלהם בלמידה פתוחה ומקוונת? על מנת לענות על שאלת המחקר המרכזית, הוגדרו שני מדדי הצלחה ממוקדי לומד: שביעות הרצון מההשתתפות בקורס ומימוש ציפיות הלומד מהקורס. מדדים אלה זוהו כמדדי הצלחה אלטרנטיביים למדדי ההצלחה שנקבעו על ידי מפתחי הקורס.

בפרויקט מחקר זה בוצעו חמישה מחקרים. חמשת המחקרים, כולם ביחד וכל אחד לחוד, מגדירים את הבעיה התיאורטית ובוחנים בצורה אמפירית את שאלת המחקר תוך שימוש במגוון טכניקות של אנליטיקות למידה. המחקר הראשון, המוצג *בפרק 2*, מציג ניתוח השוואתי בין המודלים העסקיים של ההשכלה הגבוהה המסורתית ולמידה פתוחה. הניתוח בוחן את ההשפעה של חדשנות דיגיטלית על המודלים העסקיים של המוסדות להשכלה גבוהה בהסתמך על המטפורה המפורסמת של Raymond (1999), "הקתדרלה והבזאר". מטפורה זו בחנה מודלים עסקים שונים בתחום פיתוח התוכנה. בהשאלה מהמטפורה, השינויים אשר קודמו על ידי "הבזאר" סייעו לאימוץ קורסי המוק על ידי מוסדות ההשכלה הגבוהה המסורתיים המכונים "קתדרלות". אך בו זמנית, דרשו פיתוח של מדדי הצלחה חדשים הממוקדים בלומד אשר יתאימו למערכת האקולוגית החינוכית המתהווה. פרק זה תורם לספרות המתפתחת על ההשפעה האסטרטגית של חינוך פתוח ומקוון על המוסדות להשכלה גבוהה ומדגיש את הצורך להגדיר ולחקור מדדי הצלחה חדשים ללמידה.

המחקר השני, המוצג *בפרק 3*, מציג שני מדדי הצלחה הממוקדים בלומד: שביעות הרצון מההשתתפות בקורס ומימוש ציפיות הלומד מהקורס. מדדים אלה מתאימים ללמידה לא פורמלית לאורך החיים כפי שמציעים קורסי המוק. המחקר מגדיר באופן אמפירי מדדים אלה וחושף את

המנבאים שלהם בקורסי מוק. תוצאות המחקר מבהירות את מערכת היחסים המורכבת בין מאפייניהם הסוציו-דמוגראפיים והפסיכו-פדגוגיים של המשתתפים בעת הכניסה לקורס, התנהגות הלומדים במהלך הקורס ומדדי ההצלחה ממוקדי הלומד. במחקר השלישי, המוצג *בפרק 4*, נבחנת השפעתם של המאפיינים הסוציו-דמוגראפיים והפסיכו-פדגוגיים על המחסומים להשגת שביעות רצון מתהליך הלמידה בקרב משתתפי קורסי המוק. המחקר בוחן מה הם המחסומים להשגת שביעות רצון מתהליך הלמידה ומה מנבא את אותם מחסומים ובכך מספק תובנות נוספות לגבי טיבעה של שביעות הרצון כמדד להצלחת הלמידה.

המחקר הרביעי והחמישי, המוצגים *בפרק 5* ו*בפרק 6*, מרחיבים מחקרים קודמים שהראו כי קיבוץ משתתפים באשכולות על פי מסלולי הלמידה שלהם, מספק יותר מידע ובעל יותר פוטנציאל לשיפור פדגוגי, יחסית לקיבוץ משתתפים באשכולות על בסיס ספירה סטטית של נתוני התנהגות (Kizilcec et al., 2013).

המחקר הרביעי, *בפרק 5*, מבקש לחקור גישה חדשנית לאיתור דפוסי התנהגות המשתמשים על ידי איתור רצף קצר של פעילויות ומקבץ אותן על בסיס שונות משותפת. ניתוח זה מאפשר לבנות דפוסי התנהגות משמעותיים בתהליכים לא מובנים כמו בקורסי מוק וצורות אחרות של למידה מקוונת.

המחקר החמישי, המוצג *בפרק 6*, זיהה את ההשפעה של רצפי פעילות הלמידה של המשתתפים כמנבא לרמת מימוש כוונת המשתתף. במחקר זה, נעשה שימוש בגישה חדשה שהושאלה מתחום עיבוד שפה טבעית (NLP) לזיהוי רצפי פעילויות למידה שונים. הקשר בין רצפי הלמידה שזוהו לרמת מימוש כוונת המשתתף היה מובהק ומשמעותי.

הפרק האחרון (*פרק 7*) מספק סקירה כללית של הממצאים בכל פרק ואסופת תובנות מחמשת המחקרים שהוצגו. הפרק מסתיים בדיון כללי ובמסקנות. השלכות, מגבלות והצעות מחקר עתידיות מוצעות.

המחקרים שהוצגו בעבודה זו שופכים אור על חשיבות ההתבוננות במדדי הצלחה הממוקדים בלומדים. תוצאות המחקרים מציעות נקודת מבט חדשה לניתוח מדדי ההצלחה הממוקדים בלומדים בתצורות שונות של למידה פתוחה ומקוונת כדוגמת קורסי מוק. המערכת החינוכית, קובעי המדיניות והחברה בכלל אמורים לסייע ללומדים לאורך החיים ללמוד כיצד להגדיר את יעדיהם וכיצד לווסת את תהליך הלמידה שלהם. ניתן וצריך לשלב מאמצים אלה עם התמיכה המתפתחת של מערכות המציעות למידה מותאמת אישית, העושות שימוש באינטליגנציה מלאכותית.

## Declarations:

1. Learning analytics can be used for theory development and to address concrete problems in educational practice.

2. Educational research should take into account the learners' predispositions.

3. Participants' intentions and goals should be taken into consideration when measuring the level of success in (open online) course.

4. The changes promoted by the "bazaar" facilitate the adoption of MOOCs by the mainstream "cathedral", but require, at the same time, the development of new learner-centered outcomes measures, which will be appropriate for the emerging educational ecosystem.

5. Schools and universities should assist students in developing their self-regulation skills.

6. Education can be available freely and without cost via digital means, but learners need basic skills and knowledge in order to effectively use the educational resources.

7. The Covid-19 pandemic accelerated digital changes around the world. Those changes should be embraced by educational institutions that aim to prepare their graduates for the post Corona world.

**Through the Lens of the Learner:**
*Using Learning Analytics to Predict Learner-Centered Outcomes in Massive Open Online Courses*

In the digital era, technology is leading to massive changes in the way we learn. The changes in the knowledge society are fueled by digital innovation. One marker for these changes is the rapid growth of open educational resources (OERs) and mainly massive, online, open courses (MOOCs). MOOCs are online courses with unlimited participant capacity, offered via the web. They enable learners with different academic backgrounds to learn at any place and any time, almost free of charge.

MOOCs are different from formal educational courses in the sense that participants may have diverse goals and expect a variety of different learning outcomes that can be defined by the participants themselves rather than by the course instructors. This dissertation focuses on learner-centered outcomes, namely, learner satisfaction and learner intention-fulfillments, as alternative course outcome measures, and in the antecedes of these outcomes.

The dissertation describes five studies. These five studies defined the theoretical problem and empirically revealed some of the answers using several learning analytics techniques. Individually and all together, the studies turned the spotlight on the importance of using learner-centered outcomes, and suggest a novel perspective to analyze these outcomes.

Although MOOC based learning is a niche activity in higher education institutions, the lessons that were learned can and should affect the educational system in the knowledge era, and moreover so during the COVID-19 pandemic. The results of the studies suggest that the educational system, policymakers, and society as a whole should help lifelong learners to learn how to define their goals and regulate their learning processes, using sophisticated learning analytics to collect and analyze learners' online behavior.

Open Universiteit
www.ou.nl